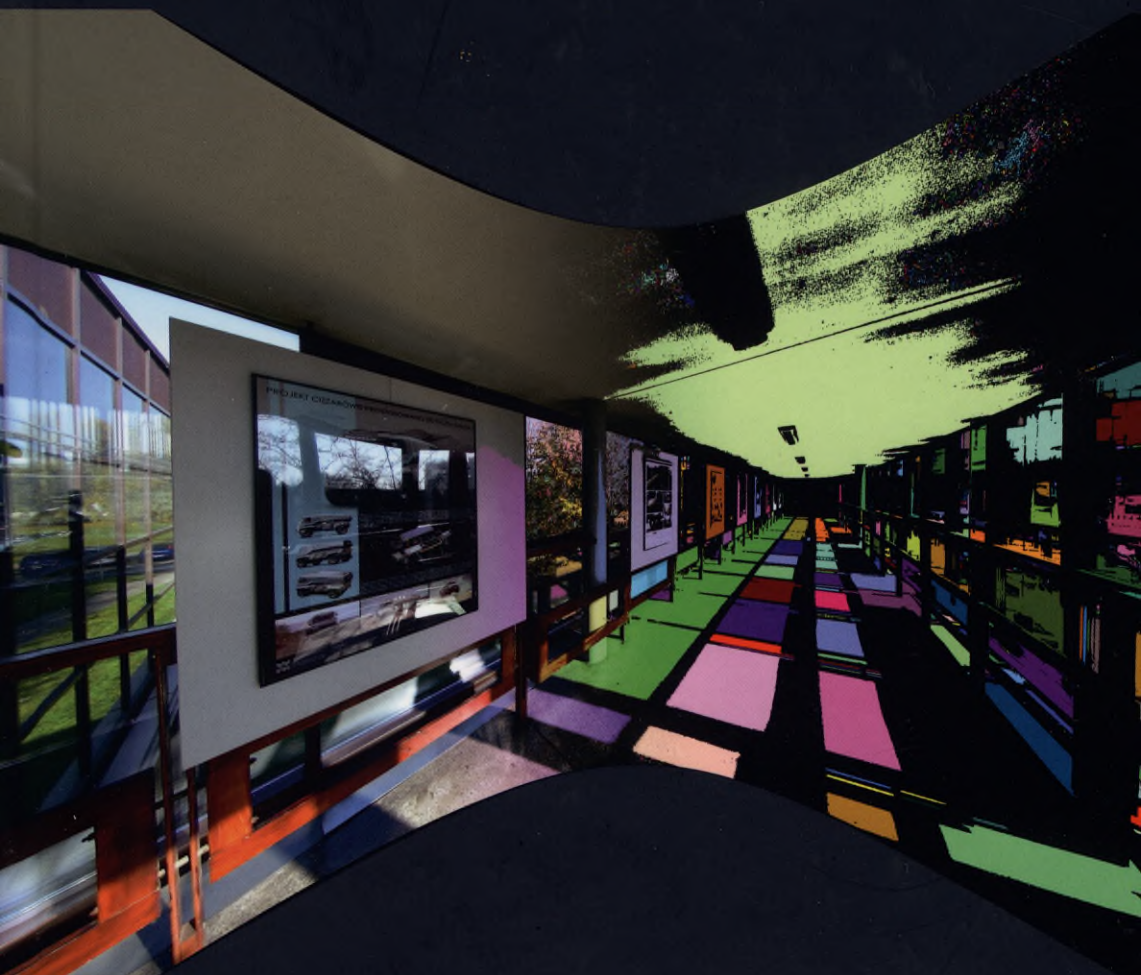


Leszek Wojnar

Analiza obrazu Jak to działa?



Kraków 2020

Biblioteka Politechniki Krakowskiej



100000337767

Leszek Wojnar

Analiza obrazu

Jak to działa?



Politechnika Krakowska
im. Tadeusza Kościuszki

Kraków 2020

PRZEWODNICZĄCY KOLEGIUM REDAKCYJNEGO
WYDAWNICTWA POLITECHNIKI KRAKOWSKIEJ

Tadeusz Kapecki

PRZEWODNICZĄCY KOLEGIUM REDAKCYJNEGO WYDAWNICTW DYDAKTYCZNYCH

Agata Zachariasz

REDAKTOR SERII

Artur Ganczarski

RECENZENCI

Krzysztof J. Kurzydłowski

Janusz Szala

SEKRETARZ SEKCJI

Agnieszka Filosek

OPRACOWANIE REDAKCYJNE

Michał Stachowski

OPRACOWANIE GRAFICZNE, PROJEKT OKŁADKI

Leszek Wojnar

SKŁAD, ŁAMANIE

Małgorzata Murat-Drożyńska

© Copyright by Politechnika Krakowska

ISBN 978-83-66531-52-9

Wydawnictwo PK, ul. Skarżyńskiego 1, 31-866 Kraków; tel. 12 628 37 25, fax 12 628 37 60

e-mail: wydawnictwo@pk.edu.pl □ www.wydawnictwo.pk.edu.pl

Adres do korespondencji: ul. Warszawska 24, 31-155 Kraków

Druk i oprawę wykonano w Dziale Poligrafii Politechniki Krakowskiej.

Ark. wyd. 17,5

Zam. 6/2021

Nakład 150 egz.

BPK-N-9/2022



Spis treści

1. Wstęp, który warto przeczytać	5
2. Cyfrowe obrazy	13
3. Akwizycja obrazów	27
4. Operacje geometryczne i punktowe	45
5. Filtry	87
6. Korekcja cienia	131
7. Binarizacja	147
8. Operacje morfologiczne	173
9. Analiza obrazów kolorowych	217
10. Zaawansowane metody analizy	253
11. Elementy stereologii	269
12. Pomiary	291
13. Analiza obrazów 3D	333
14. Budowa algorytmów	345
15. Podsumowanie	365
16. Literatura	367

1. Wstęp, który warto przeczytać

Czasami odnosimy wrażenie, że czytanie wstępu to strata czasu, bo zawiera tylko jakieś ogólniki i ewentualnie podziękowania. Sporo w tym racji, ale tylko we wstępie mogłem zawrzeć odpowiedzi na kilka fundamentalnych pytań:

- Dla kogo jest przeznaczona ta książka?
- Dlaczego w ogóle powstała?
- Co można w niej znaleźć?
- Jak z niej korzystać?

Myślę, że przed pojęciem lektury warto znać odpowiedzi na te pytania i dlatego zachęcam gorąco do przeczytania wstępu.

Komputerowa analiza obrazu coraz częściej jest obecna w naszym życiu. Poniższe przykłady nie stanowią wyczerpującej listy, ale pokazują bogactwo zastosowań:

- OCR (*Optical Character Recognition*) pozwala na zamianę zeskanowanego tekstu na formę edytowalną,
- niemal w każdym sklepie czytniki analizują kody kreskowe i na ich podstawie kasa wylicza należność za zakupy,
- automatyczne odczytywanie tablic rejestracyjnych jest wykorzystywane przez systemy odcinkowego pomiaru prędkości, a także do kontroli wjazdu i wyjazdu na niektórych parkingach oraz autostradach,
- kontrola jakości, polegająca na sprawdzaniu kompletności zespołów elektrycznych lub wypełnienia blistrów tabletkami, odbywa się za pomocą systemów wizyjnych,
- ocena struktury materiałów metodami analizy obrazu doczekała się już nawet formalnych norm,
- obrazowa diagnostyka medyczna coraz częściej korzysta ze wspomagania komputerowego przy interpretacji uzyskanych obrazów.

Zastosowania analizy obrazu wciąż się rozszerzają. Obecnie za pomocą smartfona możemy odczytywać kody QR i dzięki temu np. płacić za paliwo bez konieczności odchodzenia od samochodu. Niektóre smartfony są odblokowywane w wyniku analizy obrazu twarzy właściciela zamiast kodu wprowadzanego z klawiatury. Zatem poznanie podstaw komputerowej analizy obrazu niewątpliwie ma sens.

Dla kogo jest przeznaczona ta książka?

Zabrzmiało to być może w sposób zarozumiały, ale przeznaczona jest dla wszystkich. Oczywiście, przede wszystkim jest adresowana do tych, którzy nie mają doświadczenia z analizą obrazu lub jest ono niewielkie. Dlatego pisana jest w sposób maksymalnie prosty i przystępny.

Niektóre osoby mogą mieć wątpliwość, że treść tej książki nie jest już potrzebna, gdyż obecnie normą stało się tworzenie oprogramowania przez składanie znalezionych przede wszystkim w Internecie narzędzi, tak jak składamy działające zabawki w LEGO Technic. Poza tym, coraz powszechniej korzystamy ze sztucznej inteligencji.

Wszystko to prawda, ale warto pamiętać o kilku sprawach:

- Często potrzebne jest szybkie rozwiązanie jakiegoś dość prostego zagadnienia i wtedy znajomość opisanych w tej książce narzędzi będzie bardzo użyteczna.
- Najbardziej efektywne jest zastosowanie metod sztucznej inteligencji nie do obrazu wyjściowego, tylko do obrazu już przetworzonego albo przetwarzania danych pozyskanych w wyniku analizy obrazu. A do tego niezbędna jest wiedza na temat tych metod.
- Sztuczna inteligencja jest dobrym narzędziem do automatyzacji procesów decyzyjnych lub klasyfikacji, pod warunkiem że mamy odpowiednie wzorce do nauki. Jeżeli jednak chcemy zbadać pewne zjawiska, których jeszcze nie znamy, to sztuczna inteligencja zwykle niewiele nam pomoże.
- Jeżeli korzystamy z gotowych rozwiązań, zwłaszcza typu *public domain* (domena publiczna), to może się okazać, że chcielibyśmy je ulepszyć w wyniku naszych doświadczeń. I w tym momencie znajomość podstaw analizy obrazu okaże się bezcenna.

Pomimo rozwoju sztucznej inteligencji jeszcze daleko jej do pełnej sprawności. Za dowód niech posłużą dwa przykłady z zakresu motoryzacji:

- Rozpoznawanie znaków drogowych działa jedynie w odniesieniu do ograniczeń prędkości i zakazu wyprzedzania, choć wydaje się, że grafika znaków drogowych jest mocno uproszczona, a zbiór znaków ograniczony. W dodatku system, nawet wspomagany nawigacją satelitarną, nie radzi sobie z faktem, że bardzo często skrzyżowanie odwołuje obowiązujący zakaz.
- Utrzymywanie samochodu na pasie ruchu działa dobrze, jeśli linie są świeżo namalowane. Gdy stają się mniej wyraźne, system się gubi. A kierowca musi sobie poradzić nawet we mgle i przy padającym deszczu.

Pamiętajmy, że nikt z nas nie zna się na wszystkim. Dlatego nawet osoby posiadające duże doświadczenie w analizie obrazu mogą zapewne znaleźć w tej książce nowe, dotąd nieznane elementy, albo zobaczyć jeszcze niewykorzystywane potencjalne możliwości niektórych narzędzi. Również wyjaśnienia, jak działają poszczególne przekształcenia, mogą być inspirujące.

Dlaczego ta książka w ogóle powstała?

Pytanie jak najbardziej zasadne, bo napisano już co najmniej kilkadziesiąt książek podejmujących kwestię zasad komputerowej analizy obrazu. Odpowiedź jest prosta: dlatego, że nie znalazłem wśród nich odpowiedniego podręcznika.

Liczne podręczniki zawierają wartościowe treści, ale podane są one zwykle w sposób dość trudny do zrozumienia, bo odwołują się często do skomplikowanego aparatu matematycznego lub opisują poszczególne narzędzia w sposób bardzo ogólnikowy. Zgodnie z tytułem, w tej książce skupiłem się na wyjaśnianiu, jak to działa, bez uciekania się do złożonych formuł matematycznych. Zrozumienie sposobu działania poszczególnych narzędzi powinno być zachętą do ich efektywnego wykorzystania.

W 1999 r. nakładem wydawnictw CRC Press ukazała się moja książka *Image analysis. Applications in materials engineering*. Była to nowatorska w skali całego świata pozycja, stanowiąca niejako rodzaj „książki kucharskiej” analizy obrazu. Zawiera bowiem opisany szczegółowo szereg autorskich, sprawdzonych algorytmów rozwiązujących konkretne problemy. Można je było wykorzystać wprost lub zmodyfikować, jeżeli napotykało się zbliżony problem.

Od czasu wydania tej pozycji minęło jednak sporo czasu i metody analizy obrazu znacznie się rozwinęły. Stąd pojawił się pomysł napisania książki ulokowanej gdzieś pośrodku pomiędzy dziełami teoretycznymi i czystą praktyką. Zawarłem w niej zatem opisy sposobu działania różnych narzędzi, pokazałem wyniki ich stosowania oraz wskazałem potencjalne możliwości praktycznego użycia. Można się z niej również dowiedzieć, jakie są niedostatki i ograniczenia omawianych narzędzi.

Przedstawiona powyżej koncepcja wynika również z faktu, że książka miała w początkowym zamyśle stanowić pomoc do wykładów z podstaw analizy obrazu. Konsekwencją tego założenia jest rezygnacja z opisywania konkretnych problemów lub służących ich rozwiązaniu algorytmów. Tym zagadnieniom poświęcona jest druga książka, autorstwa Adama Piwowarczyka, *Analiza obrazu. Jak to się robi?* Podobieństwo tytułów nie jest przypadkowe, gdyż obydwie pozycje uzupełniają się, choć są w pełni autonomiczne.

Co można znaleźć w tej książce?

Znane chińskie przysłowie mówi, że jeden obraz jest wart więcej niż tysiąc słów. Trudno o trafniejsze motto do książki na temat analizy obrazów. Ponieważ starałem się pokazać, jak działają poszczególne narzędzia, to ilustracje stały się bardzo ważnym składnikiem całej pracy. Liczba ilustracji, czasami łączonych po kilka w ramach ilustracji o jednym numerze, jest większa niż liczba stron książki. Ponad 90% ilustracji jest zupełnie nowych i opracowano je w ciągu ostatnich kilku miesięcy na potrzeby zdalnych wykładów, prowadzonych podczas pandemii koronawirusa lub wyłącznie dla tej książki. Ilustracje powstawały przed pisanem tekstu i tym samym narzucały układ treści. Dlatego stanowią integralną część całości, a nie jedynie dekorację. Ich usunięcie sprawiłoby, że książka stałaby się niezrozumiała.

Oprócz ilustracji książka zawiera wspomniane już wcześniej opisy poszczególnych narzędzi. W ograniczonej liczbie zamieszczono też wzory, które w ścisły sposób opisują pewne działania i które można wykorzystać w pracy nad algorytmami analizy obrazu. Ich liczba oraz wpływ na treść zostały tak dobrane, że osoby czujące głęboką niechęć do analizy wzorów matematycznych powinny zrozumieć całą treść nawet w przypadku ich pominięcia.

Jak korzystać z tej książki?

Można byłoby zacytować Jerzego Owsiaka i powiedzieć: Róbta, co chceta. Oczywiście, nie chciałbym, aby ta książka była używana wyłącznie jako przycisk zabezpieczający stertę papierów przed porwaniem przez przeciąg lub podkładka pod zbyt nisko stojący monitor. Mówiąc jednak serio, nie jest konieczne czytanie jej od deski do deski. Można wybrać dowolny rozdział, który nas w danej chwili interesuje. Czytanie poprzednich rozdziałów nie jest bezwzględnie konieczne, zwłaszcza dla osób, które mają już doświadczenie z analizą obrazu. Jeżeli jakiś fragment okaże się niezrozumiały, to trzeba będzie cofnąć się do wcześniejszych rozdziałów, aby uzupełnić niezbędną wiedzę. W szybkim wyszukiwaniu potrzebnej wtedy informacji pomocne może być skorzystanie z wersji elektronicznej.

Być może część osób stwierdzi, że lepszym rozwiązaniem byłoby zamieszczenie na YouTube serii odpowiednich filmów instruktażowych. Jestem przeciwnego zdania. Niewątpliwie czytanie oznacza większy wysiłek niż oglądanie filmu. Warto jednak pamiętać o wynikach badania zachowania się mózgu podczas oglądania filmu. Nawet, jeżeli ma to być film instruktażowy, to mózg emituje fale alfa, typowe dla stanu relaksu. Czytając, można nauczyć się więcej i wbrew pozorom szybciej. Poza tym wyszukiwanie

konkretnej informacji w książce traktowanej jako poradnik jest wygodniejsze niż przewijanie materiałów wideo. Oczywiście, można użyć pliku .pdf, ale to może wymagać instalacji kolejnego monitora, aby zapewnić sobie odpowiedni komfort pracy. A książka jest jednak tańsza od monitora. Dodatkowo dla mnie przyjemność stanowi już samo przeglądanie ładnie wydanej książki. Może pod tym względem przypominam dinozaura, ale zapewne obcowanie z ładnymi przedmiotami dla każdego ma znaczenie.



Il. 1.1. Porównanie cyfrowego i analogowego zegara. Opis w tekście

Na zakończenie wstępu próbka problemów związanych z analizą obrazu, wykonywaną przez nasz zmysł wzroku. Wyobraźmy sobie, że jeden z naszych znajomych umówił się z nami na ważne spotkanie o godzinie 19.00. Dysponuje nowoczesnym, cyfrowym zegarkiem i patrzy na wyświetlacz. Widzi (Il. 1.1a) cztery cyferki, rozdzielone w środku dwoma kropkami. Odczytuje godzinę: 16.42 i zaczyna liczyć: do pełnej godziny brakuje $60 - 42 = 18$ minut. Będzie wtedy godziną 17.00. Tak więc do spotkania mam jeszcze 2 godziny i 18 minut, zatem nie muszę się bardzo spieszyć.

Nasz drugi znajomy, umówiony na to samo spotkanie, ma również bardzo nowoczesny zegarek, ale jego ekran wyświetla obraz tradycyjnego zegara ze wskazówkami. Spojrzał na niego i widzi układ wskazówek pokazany na Il. 1.1b. Nie zastanawia się, która jest godzina! Gdybyśmy go o nią zapytali, to najpewniej powiedziałby, że nie wie i musi spojrzeć na zegarek. W jego mózgu jest już zakodowana informacja, że spotkanie rozpocznie się wtedy, kiedy wskazówki ustawią się w pozycji pokazanej szarym kolorem na Il. 1.1c. Szybko oceniając kąt między małymi wskazówkami na dole tarczy, nasz znajomy stwierdza: Do spotkania mam jeszcze ponad dwie godziny, zatem nie muszę się bardzo spieszyć.

Podsumowując działania naszych dwóch znajomych, można stwierdzić, że:

- Niezależnie od rodzaju posiadanego zegarka obaj doszli do tego samego wniosku, że mają jeszcze dość dużo czasu do spotkania.

- Posiadacz zegarka z cyfrowym wyświetlaczem poznał czas pozostały do spotkania z dokładnością do jednej minuty. Jest to jednak pewien szum informacyjny, bo tak dokładna informacja nie jest mu do niczego potrzebna.
- Drugi znajomy nie wie dokładnie, ile ma czasu do spotkania, ale jego wiedza na ten temat jest w zupełności wystarczająca do podjęcia odpowiedniej decyzji, czy już wychodzić z domu, czy jeszcze nie.
- A teraz najważniejsza różnica: na pozyskanie tej samej w gruncie rzeczy informacji drugi znajomy potrzebował znacznie mniej czasu.

Oczywiście, niektóre osoby mają problem z nauczeniem się odczytywania czasu na zegarze z klasycznym układem wskazówek. Warto jednak zwrócić uwagę, że tam, gdzie musimy jednocześnie kontrolować wskazania wielu urządzeń, a czas odczytu ma decydujące znaczenie dla bezpieczeństwa, używa się nadal wskaźników ze wskazówkami. Tak jest np. w kokpitach najnowocześniejszych samolotów. Proste zagadnienie interpretacji obrazu wyświetlacza zegarka, a tyle problemów! Chyba trzeba poczytać, co znajduje się na kartach tej książki.

Tradycyjnie we wstępie zamieszczamy podziękowania. Nie jest to obowiązkiem, ale daje szansę zwrócenia uwagi, że ktoś nam pomógł. W szczególności chciałbym podziękować:

- Nieżyjącemu już prof. Jerzemu Rysiowi za przekazanie podstaw wiedzy stereologicznej i zainteresowanie tą dziedziną, które później przeniosło się na metody automatyzacji pomiarów i zagadnienia analizy obrazu. Jestem mu bardzo wdzięczny za wsparcie i pomoc, zwłaszcza podczas mojego postępowania habilitacyjnego.
- Dr. Jean-Louis Chermant dziękuję za wieloletnią przyjaźń, dzielenie się doświadczeniem i umożliwienie udziału w kilku ważnych konferencjach, jak również za pierwsze kontakty z Bruno Lay.
- Dr. Bruno Lay, założycielowi i prezesowi firmy ADCIS S.A., za wieloletnią współpracę i zapewnianie dostępu do najnowszych wersji systemu analizy obrazu Aphelion.
- Dr. Gervais Gauthier z firmy ADCIS S.A. za wyjaśnienia, makropolecenia oraz rozwiązywanie problemów pojawiających się podczas pracy z rozbudowanym systemem Aphelion, który był podstawowym narzędziem analizy obrazu podczas pracy nad tą książką.
- Baśce, za to, że ułatwiła mi przetrwanie ponadpółrocznego okresu bardzo intensywnej pracy nad książką.

- Redaktorowi naukowemu oraz recenzentom za cenne uwagi, które pozwoliły na usunięcie nieuniknionych w tak obszernej pracy usterek oraz istotne ulepszenie opisu niektórych zagadnień.

Na tym kończę podziękowania, aby ich lista nie okazała się zbyt długa, bo swoją cegiełkę do wydania tej książki dodało wiele osób. Często odbywało się to w ramach obowiązków służbowych, co nie zmienia faktu, że ich pomoc była bardzo wartościowa, za co serdecznie dziękuję. A wiemy, że mogli przeszkadzać.

Trudno sobie wyobrazić, abym dziękował koronawirusowi, ale prawdą jest, że konieczność pracy zdalnej była impulsem do przygotowania odpowiednich materiałów dydaktycznych, które stały się załącznikiem od dawna planowanej książki, na której opracowanie wciąż brakowało czasu.

Życzę miłej lektury!

A handwritten signature in blue ink, reading "Łukasz Wojnar". The signature is fluid and cursive, with the first name "Łukasz" and the last name "Wojnar" clearly distinguishable.

Kraków, listopad 2020 r.

2. Cyfrowe obrazy

Obraz jest pojęciem wieloznacznym. Może oznaczać na przykład:

- dzieło sztuki, najczęściej zawieszane na ścianie, zwykle wykonane techniką malarską; malowidło,
- dzieło sztuki filmowej (najlepszy obraz Spielberga),
- zbiór wszystkich wartości przyjmowanych przez funkcję dla każdego elementu danego podzbioru jej dziedziny,
- obserwowaną lub zapamiętaną sytuację, wydarzenie (ciągle mam w głowie obraz tego wypadku),
- całościowy kształt informacji dotyczący jakiegoś problemu lub zdarzenia (obraz sytuacji),
- całościowy kształt wiedzy oraz poglądów dotyczących pewnego zagadnienia (fałszywy obraz rzeczywistości).

W naszym przypadku będziemy zajmowali się obrazami zdefiniowanymi przez normę ASTM E7-99 w następujący sposób: **obraz jest odwzorowaniem obiektu za pomocą promieniowania, zwykle przez system soczewek lub zwierciadeł**. Będziemy analizowali obiekty będące cyfrową reprezentacją zdefiniowanych powyżej obrazów, czyli po prostu zbiorami danych, przede wszystkim liczb. Można pójść dalej z uogólnieniem i stwierdzić, że będziemy zajmowali się pewnymi formami informacji.

Jeżeli dysponujemy odpowiednim oprogramowaniem, to posiadając odpowiedni zbiór danych, czyli cyfrowy obraz, możemy wymusić na komputerze wyświetlenie jego zawartości nie w postaci ciągu zer i jedynek, ale w postaci pokazanej na



a



b

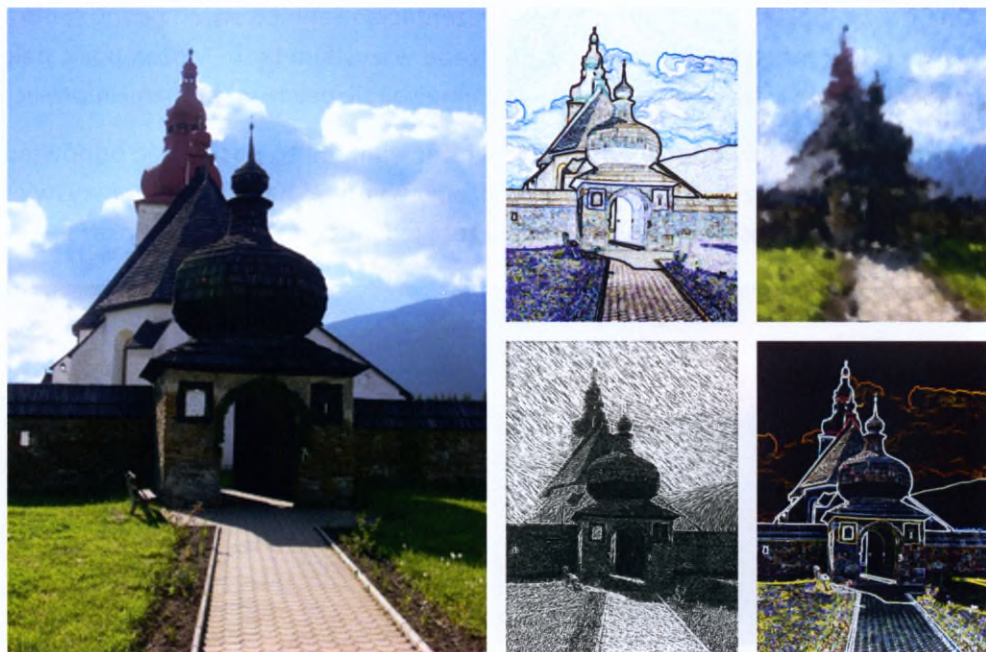
Il. 2.1. Przetwarzanie obrazu: a) obraz początkowy, b) obraz przetworzony

il. 2.1a. Dalej, wykorzystując nasze oprogramowanie, możemy tak zmienić obraz z il. 2.1a, że przyjmie postać pokazaną na il. 2.1b. Taką metamorfozę obrazu nazywamy przetwarzaniem.

Na potrzeby naszej książki przyjmujemy następującą definicję: **przetwarzanie obrazu jest procesem obróbki danych, w którym zarówno wejściowy, jak i końcowy zbiór danych jest obrazem lub zestawem obrazów**. Z zestawem obrazów mamy do czynienia wtedy, gdy kilka lub pewna ich liczba jest ze sobą ściśle powiązana. Mogą to być np. sekwencje wideo, obrazy stereoskopowe lub złożone, w których poszczególne składowe przekazują różne informacje o tej samej scenie.

Początkowy obraz, pokazany w większym rozmiarze, oraz cztery wersje jego przetworzenia pokazuje il. 2.2. Tego typu przetwarzanie wykonuje się zwykle w celach artystycznych, podczas przygotowania obrazów do prezentacji w druku lub Internecie itp. Będziemy też przetwarzali obrazy w ramach niniejszej książki, ale celem tego przetwarzania nie jest uzyskanie jakiegoś efektu artystycznego, tylko wydobycie z nich pewnej informacji.

Spróbujemy teraz zdefiniować pojęcie analizy obrazu, czyli to, czym mamy się zajmować. Zaczniemy od dwóch prostych przykładów.



Il. 2.2. Obraz początkowy i cztery wersje jego przetworzenia

Przykład pierwszy: analiza obrazu z il. 2.1b pozwala stwierdzić, że wygląda jak zdjęcie zachodu słońca, choć jest to fotografia wykonana o wschodzie.

Przeanalizujmy jeszcze drugi przykład analizy obrazu. Zastanówmy się nad odpowiedzią na proste pytanie: co widzimy na początkowym obrazie na il. 2.2?

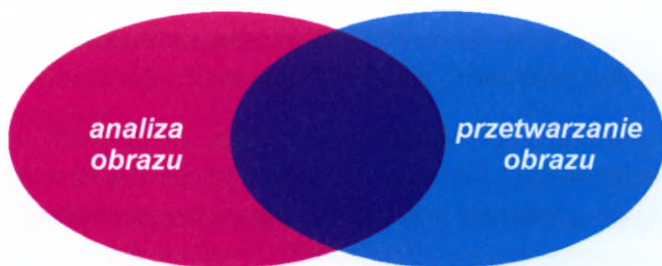
- Zapewne wszyscy powiedzą, że kościół. Odpowiedź jest prawidłowa, ale chciałbym dowiedzieć się czegoś więcej.
- Dowiem się, że kościół jest otoczony kamiennym murem z bramą z kopułą, że jest kryty gontem i ma wieżę pokrytą blachą. A ja nalegam na coś jeszcze.
- W końcu ktoś powie, że kościół stoi w górskim terenie, mamy ładną pogodę i na niebie trochę chmur.
- Najprawdopodobniej nikt nie powie, że jest to kościół w miejscowości Oravske Matiasovce.

Wszystko to, co zostało napisane powyżej o tym obrazie, jest możliwe do ustalenia bez jakichkolwiek dodatkowych narzędzi. To w obrazie tkwi informacja, że widzimy na nim kościół, że ma białe ściany itd. Jest jednak pewna wątpliwość – dlaczego prawie nikt nie powie, że jest to kościół w wiosce Oravske Matiasovce? Czyżby brakowało w tym obrazie informacji? Nie, to nam brakuje wiedzy. Bo znawca architektury sakralnej Spisza i Orawy od razu rozpozna ten kościółek. Dochodzimy zatem do bardzo ważnego twierdzenia: **wynik analizy obrazu zależy od użytych narzędzi, w tym przetwarzania obrazu, ale przede wszystkim od poziomu wiedzy osoby analizującej obraz lub poziomu informacji zapisanych w analizującym obrazu systemie i sposobu zorganizowania tej wiedzy.**

Widzimy więc, że problem nie jest prosty i banalny, co w jakimś stopniu tłumaczy objętość tej książki. Czym zatem jest analiza obrazu?

Analiza obrazu jest procesem przetwarzania (obróbki) danych, w którym dane wejściowe są obrazem lub zestawem obrazów, a wynik całego procesu nie jest obrazem. W szczególności wynikiem analizy obrazu może być liczba, zestaw liczb, tekst lub wartość zmiennej logicznej opisującej jakąś decyzję, np. o wykonaniu pewnego ruchu.

Relacje między przetwarzaniem i analizą obrazu przedstawia il. 2.3. Obydwie dziedziny mają dużą część wspólną. Ponieważ il. 2.3 ma jedynie charakter poglądowy, to wielkości poszczególnych obszarów nie mają odzwierciedlenia w rzeczywistych zakresach analizy i przetwarzania obrazu. Różnice polegają na tym, że w analizie obrazu nie korzystamy z przekształceń typowo artystycznych, np. zamieniających fotografię na obraz malowany farbą olejną lub rysowany pastelami. Nie korzystamy też z narzędzi do retuszu, które mają np. usunąć czerwone oczy na fotografii, czy poprawić odrapany fragment muru.



Il. 2.3. Relacja pomiędzy analizą i przetwarzaniem obrazu

Z kolei przetwarzanie obrazu nie korzysta z narzędzi służących do zliczania obiektów i wyznaczania ich charakterystyk geometrycznych, ani z narzędzi wspomagających analizę statystyczną lub wyciąganie wniosków.

W systemach komputerowych korzystamy z dwóch odmian grafiki: wektorowej i rastrowej. W grafice wektorowej definiuje się obiekty, które są dla komputera podstawą generowania obrazu. Można powiedzieć, że grafika wektorowa to jak gdyby przepis dla komputera, jak zbudować obraz. Możemy np. zdefiniować współrzędne wierzchołków trójkąta i polecić komputerowi określony sposób jego wypełnienia jakimś kolorem albo przejściem tonalnym, z uwzględnieniem przezroczystości itp. Ogromną zaletą grafiki wektorowej jest jej skalowalność. Obraz wektorowy zaprojektowany na ekranie komputera możemy powiększyć do rozmiarów ściany budynku bez obawy o utratę jakości.

Drugą odmianę stanowią grafiki lub obrazy rastrowe, które są jak gdyby mozaiką drobnych punktów, nazywanych pikselami. Zawsze, nawet w największych obrazach, mamy skończoną liczbę pikseli i po odpowiednio dużym powiększeniu obrazu zobaczymy pojedyncze piksele. Zatem obrazu rastrowego nie możemy bezkarnie powiększać w nieskończoność. Różnicę pomiędzy tymi dwoma rodzajami obrazów pokazuje il. 2.4.



Il. 2.4. Porównanie obrazu wektorowego (z lewej) i rastrowego (z prawej). Różnice stają się widoczne dopiero po odpowiednim powiększeniu.

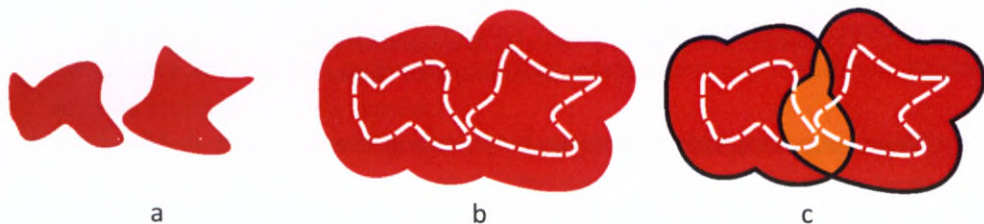
W tym przypadku powiększenie wynosi 5x

Grafika wektorowa jest dobrym narzędziem do projektowania ilustracji, pisanie tekstu (krój czcionki zdefiniowany w postaci wektorowej pozwala na dowolne kształtowanie wielkości tekstu), a nawet wykonywania złożonych modeli 3D. Staje się jednak niewygodna przy próbie rejestracji wielu drobnych szczegółów.

Rejestracja szczegółów to domena grafiki rastrowej. Jest ona w prostej linii spadkobierczynią klasycznej fotografii. Różnica polega na zastąpieniu kryształów halogenków srebra przez miniaturowe światłoczułe układy elektroniczne oraz zapis efektów naświetlania w pamięci komputera zamiast na błonie fotograficznej.

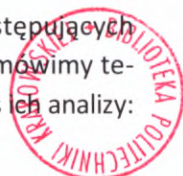
W komputerowej analizie obrazu będziemy zajmowali się obrazami rastrowymi. Warto przy tym pamiętać, że obraz wektorowy zawsze możemy zapisać w postaci rastrowej. Niestety, przekształcanie w drugą stronę jest możliwe jedynie w ograniczonym zakresie.

Niekiedy podczas analizy obrazów będziemy korzystali z wirtualnych obiektów, czyli rozwiązania znanego z grafiki wektorowej. Czasami potrzebujemy pewne elementy obrazu (Il. 2.5a) powiększyć, dodając warstwy pikseli do ich brzegu. Jeżeli będą położone blisko siebie, to taka operacja doprowadzi do ich połączenia (Il. 2.5b). Może to być niekorzystne dla wyników operacji przetwarzania obrazu. W takiej sytuacji na podstawie początkowego obrazu generuje się odpowiednie obiekty, które można powiększać bez obawy ich połączenia w momencie, gdy zaczną zachodzić na siebie (Il. 2.5c). Oczywiście, w dowolnym momencie można na podstawie tych obiektów zmodyfikować lub utworzyć nowy obraz rastrowy.



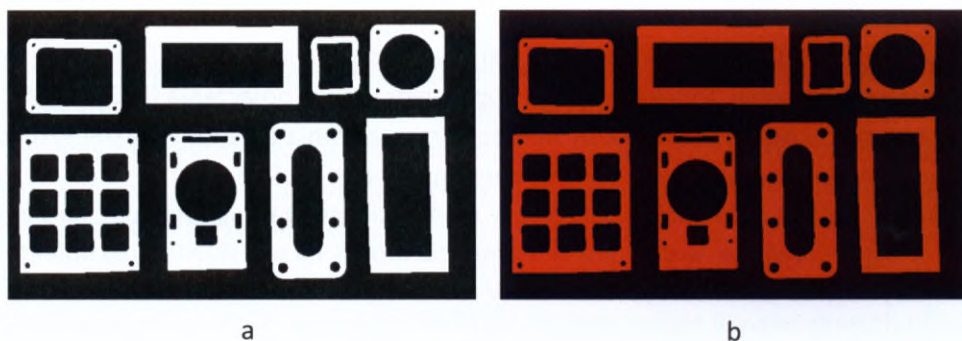
Il. 2.5. Różnice pomiędzy obrazami rastrowymi i wektorowymi: a) początkowe obiekty przed przekształceniem, b) rozrost obiektów w obrazie rastrowym prowadzi do ich połączenia, c) rozrost obiektów w obrazie wektorowym zachowuje ich odrębność

W analizie obrazu używa się różnych odmian obrazów, czasami niewystępujących w systemach do tworzenia i obróbki grafiki do celów artystycznych. Omówimy teraz podstawowe rodzaje obrazów, z którymi możemy zetknąć się podczas ich analizy:





II. 2.6. Podstawowe rodzaje obrazów: a) obraz binarny, b) obraz szary, c) obraz kolorowy



II. 2.7. Wyświetlanie obrazu binarnego: a) w systemach do grafiki komputerowej, b) w systemach analizy obrazu

- Najprostszym rodzajem są obrazy binarne (ang. *binary images*, il. 2.6a). Ich nazwa pochodzi od faktu, że zawierają tylko dwie wartości: 0 oraz 1. Dzięki temu zajmują bardzo mało pamięci i są łatwe do obróbki. Były pierwszym rodzajem obrazów, których przetwarzanie było komercyjnie dostępne już w latach 60. ubiegłego wieku. Przetwarzanie innych obrazów nie było możliwe ze względu na ograniczenia pamięci i mocy obliczeniowej. Obecnie obrazy binarne wykorzystuje się do zapisywania tekstu dla systemów OCR, czyli rozpoznawania pisma (*Optical Character Recognition*), a czasami także do zapisu ilustracji wykonanych jednobarwną kreską. Obrazy binarne są jednak bardzo ważne w analizie obrazów z dwóch powodów. Po pierwsze stanowią podstawę do wykonywania pomiarów wielkości geometrycznych, o czym więcej wiadomości znajduje się w rozdziale dotyczącym

pomiarów. Po drugie, stanowią ważny element pośredni lub pomocniczy podczas złożonych operacji przetwarzania obrazu. Więcej na ten temat przeczytamy w rozdziale poświęconym operacjom morfologicznym. Na zakończenie ważna uwaga, dotycząca wyświetlania obrazów binarnych. Zazwyczaj systemy graficzne wyświetlają je tak, że wartościom zerowym odpowiada czerni, a piksele o wartości 1 są wyświetlane jako białe. Może to być mylące, gdyż zdarzają się obrazy niebędące binarnymi, w których zarejestrowano tylko czerni i biel. Aby ograniczyć możliwość pomyłki, ze względu na istotną rolę obrazów binarnych, w profesjonalnych systemach komputerowej analizy obrazu piksele o wartości 1 standardowo wyświetla się jako czerwone (Il. 2.7).

- Najczęściej wykorzystywane w komputerowej analizie są obrazy szare, nazywane też wieloodcieniowymi (ang. *grey* lub *greyscale*, il. 2.6b). Stanowią sposób prezentowania i zapisywania informacji gromadzonej przez ogromną liczbę współczesnych narzędzi badawczych i kontrolnych, wśród których można wymienić m.in.: radar i sonar, aparat rentgenowski, ultrasonograf (USG), mikroskop elektronowy transmisyjny (TEM) i skaningowy (SEM), tomograf komputerowy (TK), rezonans magnetyczny (MRI), pozytonową tomografię emisyjną (PET), jak również bogate archiwa czarno-białej fotografii oraz filmów. Jest to zatem ogromne bogactwo obrazów. Poza tym, podobnie jak w przypadku obrazów binarnych, obrazy szare stanowią ważny element złożonych algorytmów przetwarzania obrazu.

W przypadku obrazów szarych bardzo ważna jest ich głębia bitowa, czyli po prostu liczba odcieni szarości, którą można zapisać w pliku. Najczęściej obrazy szare są zapisywane jako 8-bitowe (Il. 2.6b). Taka forma zapisu pozwala na odwzorowanie 256 odcieni szarości, co daje spory zapas na różnego rodzaju modyfikacje związane z utratą części poziomów szarości, gdyż ludzkie oko rozpoznaje zwykle jedynie około 50 odcieni szarości.

Z tego względu nie odczuwamy braku szczegółów w obrazie 6-bitowym, który zapisuje 64 odcienie szarości (Il. 2.8a). W obrazie 4-bitowym zwykle wyraźnie widać już skokowe zmiany (zobacz chmury na il. 2.8b), a obraz 2-bitowy to już drastyczna utrata szczegółów (niebo pozbawione chmur na il. 2.8c).

W praktyce analizy obrazu spotkamy obrazy szare o różnej głębii bitowej. Jeżeli chcemy dodać lub pomnożyć dwa obrazy 8-bitowe, to wynik musi być obrazem 16-bitowym. Jeżeli potrzebujemy dodatkowo wartości ujemnych, np. podczas odejmowania obrazów, to trzeba zarezerwować 32 bity (Signed 16-bit). Niektóre złożone operacje, jak np. transformata Fouriera, a nawet zwykłe dzielenie, wymagają możliwości zapisu wartości pikseli jako liczb rzeczywistych. Zwykle poświęca się w tym celu 32 lub 64 bity.

Duża głębokość bitowa wymaga odpowiedniego sposobu wyświetlania. Często największą wartość wyświetla się jako białą, a najniższą jako czarną. W przypadku obrazów z pikselami o wartościach ujemnych zero wyświetla się zwykle jako 50% szarości, czyli 128 w przypadku 8-bitowej głębokości obrazu.



Il. 2.8. Zmiana głębokości bitowej obrazu: a) 6 bitów – 64 odcienie szarości, b) 4 bity – 16 odcieni szarości, c) 2 bity – 4 odcienie szarości

- Cyfrowa fotografia pozwoliła zapomnieć o problemach z kolorami, które były typowe dla klasycznej, analogowej fotografii. Obecnie standardem stały się fotografie kolorowe, a najwięcej takich obrazów produkują smartfony. Obrazy kolorowe (ang. *color images*, il. 2.6c) są obrazami wielopasmowymi. Składają się z kilku, najczęściej trzech, obrazów składowych, odpowiedzialnych za różne elementy, niezbędne do wyświetlenia kolorowego obrazu. Najpopularniejsze są obrazy RGB (*Red Green Blue*), które wykorzystują nie tylko smartfony, ale również kamery i monitory komputerów. Zwykle na każdy kanał przypada 8 bitów, zatem obrazy kolorowe są 24-bitowe, co pozwala na zapisanie 17 777 216 kolorów. Czasami przeprowadza się kompresję liczby kolorów i zapisuje obrazy kolorowe np. w 8 bitach, ale oczywiście dzieje się to za cenę dużej straty liczby kolorów. Istnieją też obrazy kolorowe zapisywane za pomocą większej liczby kanałów, jak np. CMYK, używany w poligrafii, lub RGBA, w którym dodano kanał alfa, definiujący przezroczystość. Niekiedy używa się też obrazów pseudokolorowych, w których na potrzeby wizualizacji poszczególnym poziomom szarości przypisuje się różne kolory. Więcej na temat obrazów kolorowych w rozdziale dotyczącym ich analizy.



Il. 2.9. Przykład obrazu etykietowanego, utworzonego na podstawie 3-bitowej wersji obrazu z il. 2.6b

- Obrazy etykietowane (ang. *label images*) stanowią odmianę obrazów binarnych. Poszczególne obiekty w obrazie binarnym to spójne obszary zajmowane przez piksele o wartości 1. Czasami trudno je wyróżnić. Rozwiązano to w ten sposób, że kolejnym obszarom obrazu binarnego przypisuje się różne wartości pikseli. W ten sposób pojedynczym obiektem przestaje być spójny obszar pikseli o wartości 1, a staje się nim obszar pikseli o jednakowej wartości. W celu łatwiejszego rozróżniania obiektów poszczególnym wartościom pikseli przypisuje się różne kolory. Na il. 2.9 pokazano przykład obrazu etykietowanego, powstałego na podstawie 3-bitowej wersji obrazu z il. 2.6b. Więcej na ten temat w rozdziale poświęconym operacjom morfologicznym oraz pomiarom.
- Spotykamy też obrazy złożone. Są to obrazy wielokanałowe, w których każdy kanał odpowiada za inną informację dotyczącą danego piksela. Przykładem mogą być obrazy krawędzi (ang. *edge images*), w których jeden kanał odpowiada za wartość krawędzi, a drugi informuje o ich lokalnej orientacji. Złożonymi obrazami są też, oczywiście, obrazy kolorowe.
- Możemy też spotkać obrazy trójwymiarowe, rejestrujące wyniki np. tomografii komputerowej. Zazwyczaj są to zbiory obrazów dwuwymiarowych zapisane w różnych formatach. W przypadku diagnostyki medycznej obrazy trójwymiarowe są zwykle zapisywane w standardzie DICOM (*Digital Imaging and Communications in Medicine*). Jego omawianie wykracza poza ramy niniejszego podręcznika.

Ponieważ obraz kolorowy (Il. 2.6c) może być zapisany również jako szary (Il. 2.6b), należy przeanalizować zagadnienie konwersji, czyli zamiany jednego rodzaju obrazu w drugi. Jeżeli dysponujemy tylko obrazem w skali szarości, to jednoznaczna zamiana tego obrazu na kolorowy jest niemożliwa. Można to wytłumaczyć chociażby tym, że jeden piksel jest opisywany zwykle w obrazie szarym przez 8 bitów, a w obrazie kolorowym przez 24 bity. Zatem musielibyśmy wygenerować brakującą informację. Wykorzystując dodatkowe dane, wykraczające poza zawartość obrazu, można kolorować m.in. stare, czarno-białe zdjęcia. Trzeba jednak pamiętać, że np. kolor zielony wygląda na obrazie szarym tak samo jak czerwony. Jeżeli dotyczy to trawy, to z wyborem nie ma problemu. Ale w przypadku ubiorów lub sprzętu domowego trzeba czasami po prostu zgadywać i wybierać kolor w sposób mniej lub bardziej przypadkowy. Warto zapamiętać, że obraz szary nie zawiera żadnej informacji o kolorach.



a



b



c



d

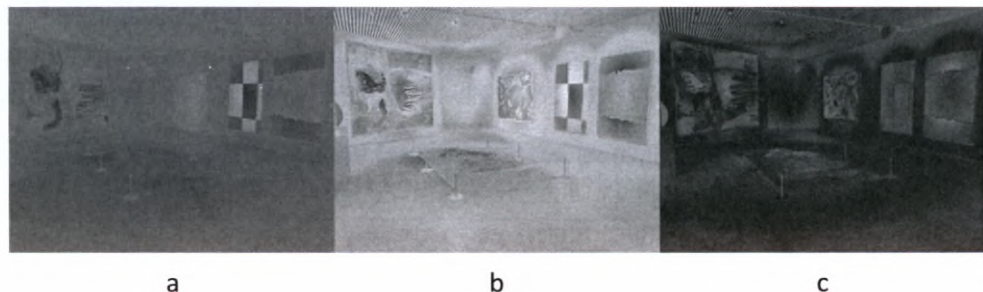
Il. 2.10. Konwersja obrazu kolorowego na szary: a) początkowy obraz kolorowy, b) kanał Intensytności z modelu HSI, c) kanał Y z modelu YUV, d) domyślne parametry konwersji w pakiecie CorelDraw

Można natomiast zamieniać obrazy kolorowe na obrazy szare. Więcej na ten temat znajduje się w rozdziale o analizie obrazów kolorowych. W tym miejscu ograniczymy się do zwrócenia uwagi na istnienie pewnego problemu. Najprostszą metodą zamiany obrazu kolorowego RGB na szary jest przypisanie każdemu pikselowi średniej arytmetycznej z wartości kanałów RGB (Il. 2.10b). Taki sposób wyznaczania obrazu szarego reprezentuje kanał intensywności w modelu HSI. Nieco inne proporcje składowych preferuje model YUV, związany historycznie z telewizją. Kanał Y z tego modelu daje bardzo naturalnie wyglądający obraz szary (Il. 2.10c). Jeszcze inny sposób konwersji proponuje pakiet graficzny CorelDraw. Można w nim w pewnych granicach zmieniać sposób generowania obrazu szarego. Obraz pokazany na il. 2.10d powstał przy domyślnych parametrach.

Wszystkie wyniki zamiany obrazu kolorowego na obraz szary, pokazane na il. 2.10, wyglądają poprawnie, a różnice między nimi są naprawdę mało widoczne. Bardziej obiektywną informację na temat tych różnic daje analiza il. 2.11, gdzie pokazano wyniki odejmowania poszczególnych obrazów z il. 2.10. Ponieważ w różnicach występują wartości ujemne, to na il. 2.11 wartości zerowej przypisano 50% szarości. Różnice w poziomach szarości, w celu poprawy czytelności il. 2.11, pomnożono przez dwa.

Nie będziemy głębiej zajmowali się problemem konwersji z jednego rodzaju obrazu na drugi. Ograniczymy się do zwrócenia uwagi, że różne sposoby generowania obrazu szarego mogą wprowadzić pewne różnice, które będą miały wpływ na ostateczny wynik analizy. Nie przesądzając, która metoda jest najlepsza, trzeba pamiętać, aby podczas analizy dużej serii obrazów używać zawsze tej samej metody zamiany obrazów kolorowych na szare.

Ostatnim zagadnieniem, którym zajmiemy się w ramach tego rozdziału, będzie wybór sposobu rejestracji danych obrazowych. Opracowano kilkadziesiąt formatów zapisywania obrazów, ale większą popularność uzyskało zaledwie kilka z nich. Do najczęściej



Il. 2.11. Różnice pomiędzy obrazami w skali szarości: a) HSI-YUV, b) HSI-Corel, c) Corel-YUV. Do celów wizualizacji dwukrotnie zwiększono rzeczywiste różnice i przyjęto 50% szarości dla braku różnicy. Różnice ujemne wyświetlane są jako ciemniejsze, a dodatnie jako jaśniejsze niż 50%

używanych, częściowo również w analizie obrazu, należą: TIFF, BMP, JPEG, RAW oraz PNG.

- TIFF (*Tagged Image File Format*) powstał w 1986 r. Można uznać ten format jako *de facto* standard w zapisywaniu obrazów rastrowych. Umożliwia zapisywanie dodatkowych informacji i komentarzy, zapewnia też obsługę przezroczystości. Jego największą wadą jest duża objętość plików, które zwykle słabo poddają się kompresji. Jak zwykle, coś tracimy, jeśli chcemy zyskać coś innego. Za cenę dużej objętości plików otrzymujemy najwyższą jakość.
- Format BMP (*Windows Bitmap*) pozwala na zapisywanie obrazów szarych i kolorowych. Podobnie jak TIFF zajmuje dużo miejsca, ale jest dość popularny, gdyż żaden z jego elementów nie został opatentowany. Zapewnia bezstratny zapis informacji obrazowej.
- JPEG (*Joint Photographic Experts Group*) dotyczy bardzo popularnego, zwłaszcza w Internecie, sposobu kompresji i zapisywania obrazów. Swoją popularność zawdzięcza niewielkiej objętości plików, co ułatwia ich magazynowanie, a przede wszystkim przesyłanie, oraz pewnym dodatkowym możliwościami, np. szybkiego podglądu miniatur obrazu. Pewną ciekawostką stanowi fakt, że używane powszechnie pliki .jpg to jeden z wielu, a nie jedyny format zapisu danych obrazowych w tym standardzie. Największym problemem plików .jpg jest wykorzystanie stratnej kompresji. Z jednej strony bardzo zyskujemy na objętości pliku, ale z drugiej – tracimy część informacji. Standard został opracowany bardzo sprytnie i nasz wzrok długo nie dostrzega utraty informacji. Dzieje się tak dzięki kompresowaniu przede wszystkim kolorów wypełniających większe powierzchnie, przy możliwie długim zachowaniu ostrości krawędzi. Trzeba dość dokładnie wpatrywać się w obrazy na il. 2.12, aby dostrzec w nich utratę jakości pomimo drastycznego zmniejszenia wielkości pliku (por. tab. 2.1). Różnice są mało widoczne, ale dane w tabeli 2.1 wskazują, że mogą one istotnie wpływać na wynik analizy. Praktyczny wniosek jest następujący: na potrzeby analizy obrazu można wykorzystywać pliki *.jpg, ale należy używać możliwie jak najwyższej jakości zapisu plików.

Tabela 2.1. Porównanie wybranych obrazów JPG z obrazem TIFF. Podane liczby dotyczą obrazów, których fragmenty pokazuje il. 2.12

Rozmiar pliku lub odchylenie wartości piksela w obrazie szarym	jakość obrazu JPG		
	najwyższa	średnia	najniższa
rozmiar pliku (dla TIFF 547 kB)	254 kB	47 kB	23 kB
średnie odchylenie względne	2,63%	11,82%	21,02%
największe odchylenie w dół	-16	-51	-88
największe odchylenie w górę	+11	+55	+96



a



b



c



d

Il. 2.12. Porównanie różnych formatów zapisu obrazu: a) TIFF, b) JPG najwyższej jakości, c) JPG średniej jakości, d) JPG niskiej jakości

- RAW to ogólne określenie formatów zapisu danych obrazowych bez wstępnej obróbki (ang. *raw* – ‘surowy’). Plik RAW zapisują dane w sposób bezstratny, kopiując po prostu dane uzyskane przez światłoczułą matrycę aparatu. Ponadto zamiast 8 bitów na piksel zapisują zwykle 12 lub 14 bitów, co pozwala na daleko idącą korektę obrazów bez widocznej utraty jakości. RAW nie jest formatem zapisu danych obrazowych, tylko ogólną nazwą rodziny tych formatów. Praktycznie każdy producent sprzętu fotograficznego opracował własny format. Do jego obróbki często nie wystarczy ogólnie dostępne oprogramowanie. Na szczęście,

kupując sprzęt danej firmy, zwykle mamy możliwość darmowego dostępu do profesjonalnego oprogramowania obsługującego format RAW danej firmy. Tego typu plików raczej nie będziemy używali bezpośrednio w analizie obrazu.

- PNG (*Portable Network Graphics*) to jeden z najpóźniej opracowanych formatów plików graficznych. Umożliwia zapis zarówno obrazów szarych, jak i kolorowych w dużym zakresie głębi bitowych, aż do 48 bitów na piksel. Pozwala na bezstratny zapis, gwarantując tym samym bardzo dobrą jakość obrazu. Zyskuje coraz większą popularność, gdyż dodatkowo nie jest obciążony ograniczeniami patentowymi.

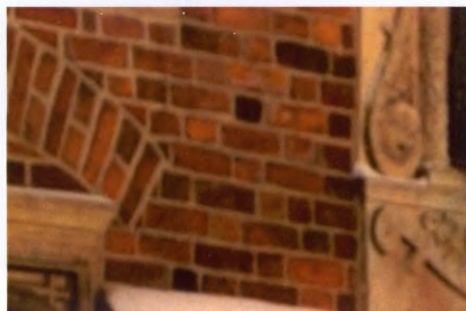
Przedstawiony w tym rozdziale pobieżny przegląd różnych zagadnień związanych z cyfrowymi obrazami daje wystarczające podstawy do zrozumienia wielu zagadnień omawianych w następnych rozdziałach.

3. Akwizycja obrazów

Akwizycja obrazów to po prostu ich rejestrowanie lub pozyskiwanie obrazów archiwalnych (co nie oznacza, że bardzo starych, tylko otrzymanych wcześniej) na potrzeby analizy obrazu. Wymagania dotyczące obrazów, podobnie jak sposoby ich pozyskiwania, są bardzo zróżnicowane. To, czego oczekuje się od obrazów na potrzeby diagnostyki medycznej, może być zupełnie inne niż wymagania na potrzeby materiałoznawstwa, teledetekcji lub kontroli jakości. Również narzędzia, które służą do pozyskiwania obrazów, niejednokrotnie drastycznie różnią się od siebie. Podsumowując ten krótki wstęp, na temat akwizycji obrazów można napisać bardzo obszerną książkę. Dlatego ograniczymy nasze rozważania dotyczące tego zagadnienia do kilku najważniejszych i uniwersalnych problemów.

Bardzo często można spotkać się ze stwierdzeniem, że na potrzeby komputerowej analizy należy używać obrazów o jak najlepszej jakości i rozdzielczości. Trudno tę opinię podważyć, ale trzeba się zastanowić, co oznacza ona w praktyce.

Na il. 3.1 widać ten sam fragment dwóch zdjęć. W istocie jest to to samo zdjęcie, tylko zarejestrowane przez aparat fotograficzny w dwóch formatach, wspomnianych w poprzednim rozdziale. Zdjęcie zapisane w formacie RAW (il. 3.1a) zawiera nieco więcej szczegółów niż zdjęcie zarejestrowane w formacie JPG (il. 3.1b), na którym widać, zwłaszcza z prawej strony, efekty kompresji obrazu. Trzeba jednak przyznać, że różnice są niewielkie, natomiast zostały okupione dużą różnicą w wielkości plików. Pełne obrazy zostały zarejestrowane w nietypowej rozdzielczości 4928x3264 piksele.



a



b

Il. 3.1. Dwie wersje tego samego zdjęcia: a) obraz zapisany w formacie RAW, b) obraz zapisany w formacie JPG

Obraz RAW zajął 16 MB, natomiast obraz JPG o wysokiej jakości zajął 5,72 MB. Wizualna ocena il. 3.1 potwierdza, że pliki JPG o najwyższej jakości mogą być używane do analizy obrazu.

Il. 3.2a-c pokazuje trzy zdjęcia fragmentów budynku. Fragmenty są różne, ale mają te same materiały i kolory elewacji, a ponadto są podobnie oświetlone. Istotne jest to, że zdjęcia zostały zrobione przy różnych ustawieniach aparatu. Fotografię na Il. 3.2a wykonano przy czasie otwarcia migawki $1/50$ s, co nie tylko pozwala zarejestrować wolno poruszające się osoby, ale przede wszystkim umożliwia wykonanie zdjęcia „z ręki”, czyli bez wsparcia statywu. Niestety, z powodu sztucznego oświetlenia, które jest słabsze w porównaniu ze światłem dziennym, zostało to okupione wysokim poziomem szumu i spowodowaną przez szum utratą ostrości. Drugie zdjęcie (il. 3.2b) zawiera już znacznie mniej szumów, ale ustawienie czułości na ISO 800 spowodowało wydłużenie czasu naświetlania do $1/6$ s, co wymaga już stosowania statywu. Niestety, jakość zdjęcia dalej nie jest zbyt dobra. Aby uzyskać taką jakość zdjęć jak na il. 3.1, konieczne jest ograniczenie czułości do ISO 200 (lub nawet do ISO 100) oraz



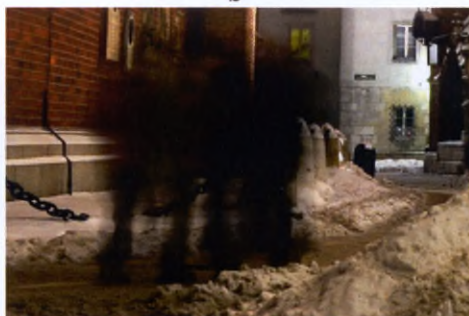
a



b



c



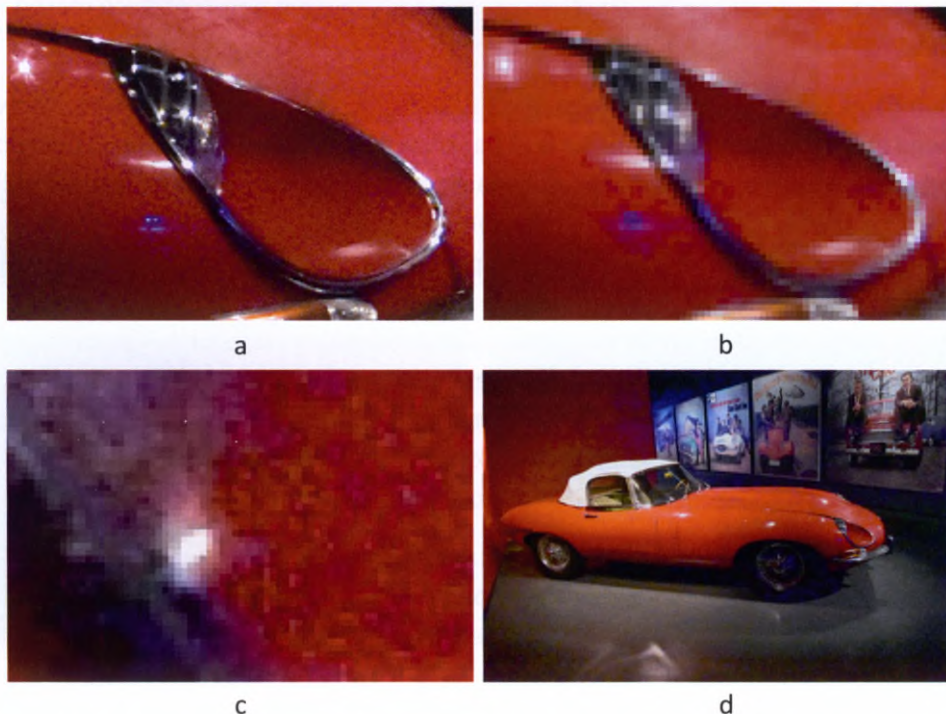
d

Il. 3.2. Detale architektoniczne tego samego budynku, sfotografowane przy różnych ustawieniach aparatu: a) czułość ISO 6400, czas $1/50$ s, b) czułość ISO 800, czas $1/6$ s, c) czułość ISO 200, czas 2 s. Parametry rejestracji obrazu (c) dają dobrą jakość fotografii budynku, ale sylwetki ludzi są wyraźnie poruszone (d)

wydłużenie czasu naświetlania do ok. 2 sekund (il. 3.2c). Jakość zdjęcia będzie dobra, ale nie możemy wtedy rejestrować ruchu i nawet osoby stojące w miejscu mogą zostać zarejestrowane w postaci rozmytych sylwetek, nazywanych przez fotografów duchami (il. 3.2d).

Jak wynika z dotychczasowych rozważań, uzyskanie odpowiedniego obrazu może wymagać wielu kompromisów i znalezienie optymalnego rozwiązania często nie jest łatwe.

Nie zawsze mocno zaszumiony obraz nadaje się tylko do wyrzucenia, zwłaszcza, jeżeli został zarejestrowany w dużej rozdzielczości. Kultowy Jaguar E-type (il. 3.3) został sfotografowany w muzeum, gdzie panowało słabe oświetlenie i nie było możliwości skorzystania ze statywu. W okolicach reflektora można dostrzec silny szum (il. 3.3a). Ponieważ aparat rejestrował obraz w rozdzielczości 6000x4000 pikseli, to można było go najpierw wygładzić, a potem zmniejszyć jego rozdzielczość 10-krotnie. Wspomniane operacje są opisane w kolejnych dwóch rozdziałach. W wyniku tych przekształceń okolice reflektora zostały pozbawione szumu, choć jednocześnie



Il. 3.3. Fotografia samochodu wykonana w trudnych warunkach: a) okolice reflektora z widocznym silnym sumem, b) ten sam element po 10-krotnym zmniejszeniu rozdzielczości, c) fragment obrazu z Il. (a) powiększony tak, aby uzyskać rozdzielczość obrazu (b), d) docelowe zdjęcie

zgubiona została część szczegółów (il. 3.3b). Nie ma to jednak znaczenia, jeżeli potrzebny jest wydruk lub analiza obrazu samochodu w takiej wielkości jak na il. 3.3d. il. 3.3c pokazuje fragment il. 3.3a powiększony tak, aby piksele były odwzorowywane w takiej wielkości jak na il. 3.3b. Opisana powyżej operacja wykorzystuje losowy charakter szumu i dlatego pomniejszony obraz, zawierający uśrednione wartości, jest pozbawiony istotnych wad. Trzeba jednak pamiętać o jednym silnym ograniczeniu, czyli wielkości obrazu. Rzadko mamy tak komfortową sytuację, że potrzebny nam końcowy obraz stanowi zaledwie 1% obrazu początkowego, jeżeli bierzemy pod uwagę liczbę pikseli.

Obraz samochodu na il. 3.3 wygląda dobrze, ale jest rzeczywiście niewielki, zatem można mieć obawy, czy takie obrazy nadają się do czegokolwiek w analizie obrazu. Zastanowimy się teraz, czy zawsze potrzebne są obrazy o gigantycznej rozdzielczości? Wyobraźmy sobie, że ktoś zaprosił nas do udziału w badaniach nad wpływem zanieczyszczenia środowiska na aktywność pająków, a jednym z elementów programu ma być ocena sieci pajęczej (il. 3.4a). Jeżeli trzeba przeanalizować setki lub tysiące obrazów, to oczywiście uzasadnione jest opracowanie odpowiednich metod ich automatycznej lub przynajmniej półautomatycznej analizy.

Pomijamy teraz problem kalibracji zdjęć. Zawsze można przyłożyć obok fotografowanego obiektu odpowiednią miarkę (il. 3.5a) lub w inny sposób zadbać o stałą odległość od fotografowanego obiektu. Jeżeli badacze będą chcieli analizować długości i odległości pojedynczych nitek, to musimy wybrać fragment obrazu o odpowiednio dużej rozdzielczości (il. 3.4b). Widoczne na nim purpurowe linie to wynik detekcji z wykorzystaniem działów wodnych. Metoda ta jest opisana w rozdziale poświęconym operacjom morfologicznym. Może się jednak okazać, że analiza szczegółów geometrii pajęczyny niewiele wnosi i jest trudna do przeprowadzenia, bo tylko część zdjęć rejestruje cienkie nitki odpowiednio wyraźnie. Być może wystarczająca byłaby ocena rozmiarów pajęczyny, a do tego wystarczy obraz pokazany na il. 3.4c. Odpowiednie narzędzia są opisane w rozdziałach poświęconym filtrom i operacjom morfologicznym.

Wielkość obrazu istotnie wpływa na czas jego analizy. O ile w przypadku pojedynczego zdjęcia nie ma to większego znaczenia, o tyle podczas analizy setek lub tysięcy obrazów każde wydłużenie czasu analizy pojedynczego obrazu powoduje znaczące opóźnienie w uzyskaniu końcowych wyników. Obraz potrzebny do detekcji nitek pajęczyny (il. 3.4b) ma rozdzielczość 1200x1200 pikseli, natomiast w przypadku oceny jej powierzchni i gabarytów (il. 3.4c) wystarczyło 300x300 pikseli. Drugi obraz ma więc 16-krotnie mniej pikseli. W dodatku, w przypadku przetwarzania tego obrazu można będzie używać do procesów filtracji (patrz rozdział

poświęcony filtrom) również 16-krotnie mniejszych liczebności elementów w niektórych filtrach. Zatem w niektórych momentach obraz o większej rozdzielczości będzie analizowany aż 256 razy wolniej. Dlatego pytanie o zasadność bardzo dużej rozdzielczości badanych obrazów ma pełne uzasadnienie.

Jeżeli potrzebujemy ilościowej charakterystyki obrazu, wykraczającej poza udział powierzchniowy obiektów lub ocenę ich kształtu, to w celu zapewnienia późniejszej



a

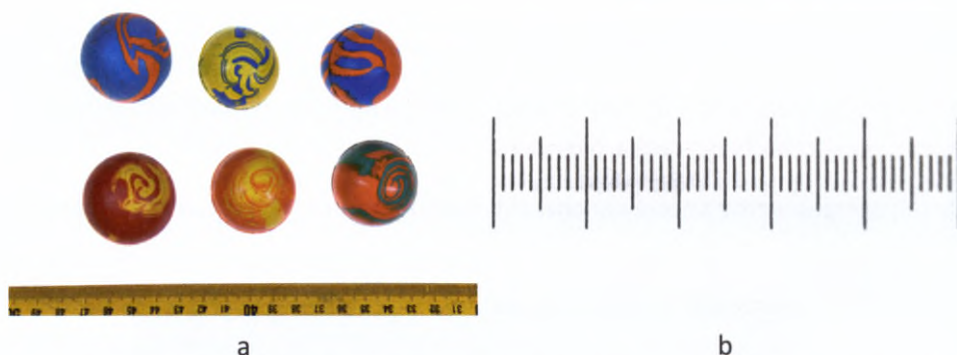


b



c

II. 3.4. Analiza pajęczyny: a) początkowa fotografia, 6000x4000 pikseli, b) fragment do detekcji nitek, 1200x1200 pikseli, wraz z wynikiem zaznaczonym purpurową linią, c) fragment do pomiaru powierzchni, 300x300 pikseli, wykryte granice pajęczyny zaznaczono czerwoną linią

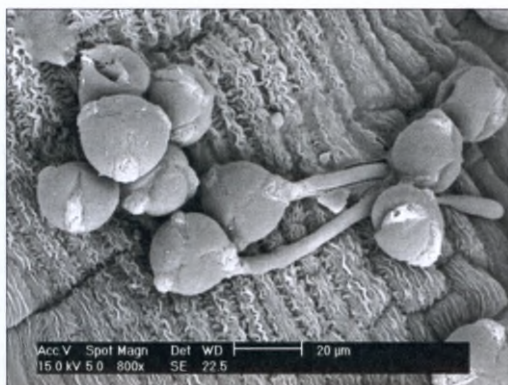


Il. 3.5. Proste narzędzia wspomagające kalibrację obrazu: a) linia z podziałką sfotografowana jednocześnie z rejestrowanym obiektem, b) wzorec odległości zarejestrowany za pomocą mikroskopu świetlnego

możliwości pomiarów obraz powinien być skalibrowany, czyli musimy wiedzieć, jakiej fizycznie odległości odpowiada przejście pomiędzy sąsiednimi pikselami. Odpowiednie informacje można zarejestrować, dokładając do wykonywanego zdjęcia skalibrowaną miarkę (il. 3.5a). W przypadku braku takiej miarki można doraźnie postąpić się jakimkolwiek przedmiotem, którego wymiary są nam znane. Często wykorzystuje się w tym celu pudełko zapalek, a dobrym narzędziem jest też kartka papieru, która w przypadku najpopularniejszego formatu A4 ma wymiary 210x297 mm.

W przypadku rejestracji obrazów za pomocą mikroskopów świetlnych fotografuje się specjalny preparat z wzorcem odległości (il. 3.5b). Obecnie coraz więcej mikroskopów, zwłaszcza elektronowych, od razu dodaje do obrazu informację o jego rozdzielczości oraz parametrach pracy mikroskopu (il. 3.6). Jest to wygodne do celów publikacyjnych, ale przeszkadza w analizie obrazu, gdyż albo musimy zmniejszyć analizowaną powierzchnię, odcinając fragment ze skalą, albo w inny sposób wyłączyć ten fragment z analizy. Technicznie możliwe jest wprowadzenie w miejscu znacznika brakującej części obrazu, ale zawsze jest to ingerencja w dane, czego zdecydowanie należy unikać. Niektóre systemy akwizycji danych obrazowych zapisują informacje o rozdzielczości w nagłówku pliku rejestrującego obraz. W tym ostatnim przypadku nie jest konieczna kalibracja, gdyż system analizy obrazu odczytuje potrzebne informacje bezpośrednio z pliku.

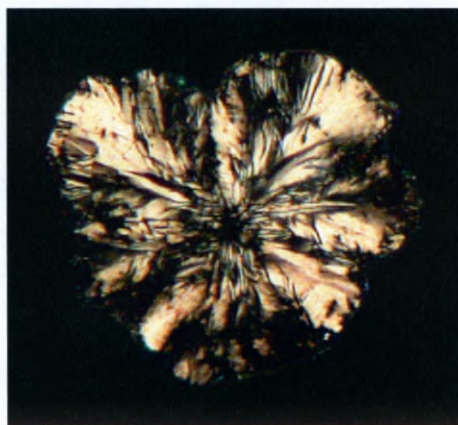
Liczne narzędzia badawcze umożliwiają rejestrację obrazów w różnych trybach pracy. W przypadku mikroskopii świetlnej może to być obraz w polu jasnym (il. 3.7a) lub ciemnym, z wykorzystaniem kontrastu fazowego, polaryzacji (il. 3.7b), różnicowego kontrastu fazowego (DIC, ang. *Differential Interference Contrast*), znanego również jako kontrast Nomarskiego (il. 3.7c), lub z wykorzystaniem fluorescencji (il. 3.7d).



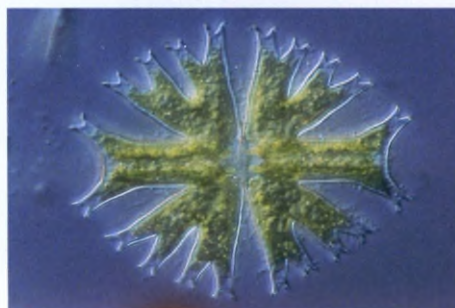
II. 3.6. Kiełkujący pyłek fiołka *Viola banksii*, obserwowany za pomocą mikroskopu skaningowego (fot. N. Wiśniewska, źródło: <http://biomist.pl/biologia/elektronowa-mikroskopia-skaningowa-sem/5369#prettyPhoto>, dostęp: 3.11.2020)



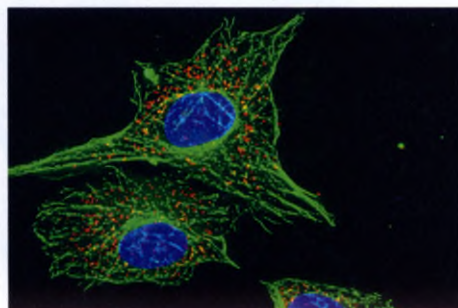
a



b



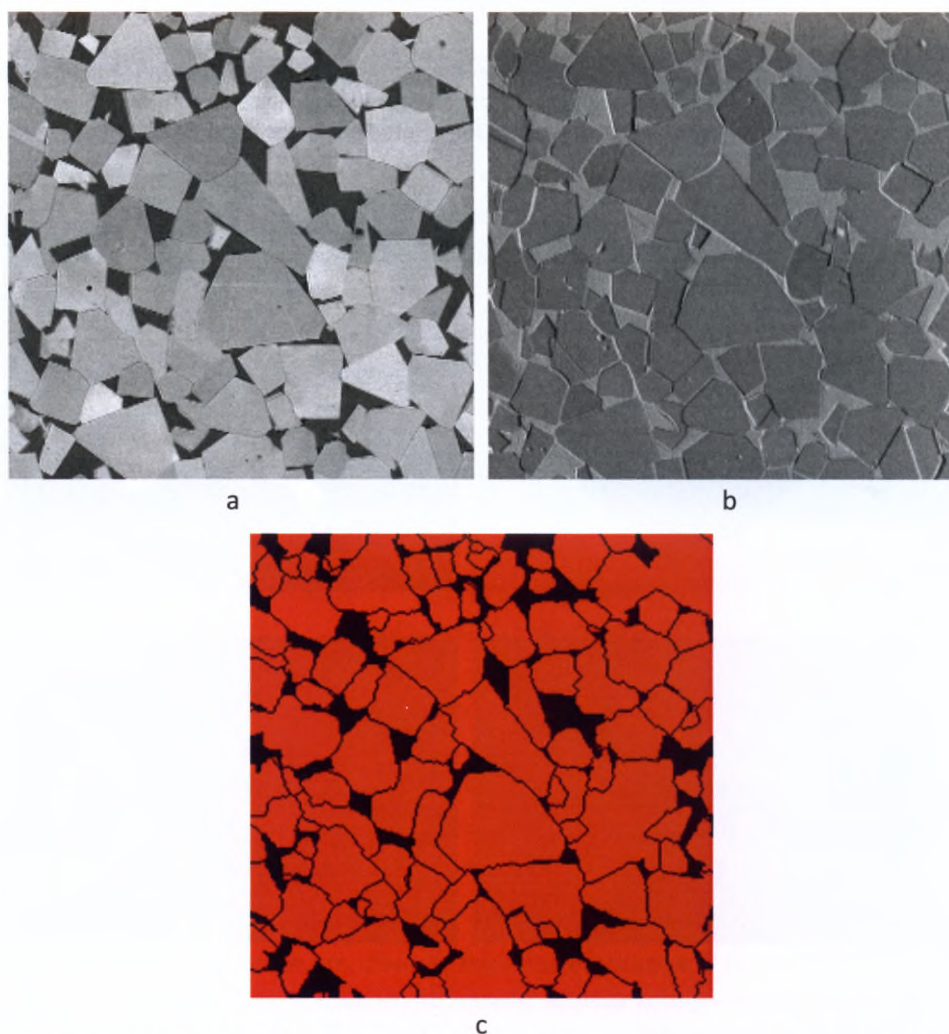
c



d

II. 3.7. Mikroskopia świetlna. Grafit w żeliwie, obserwowany przy jasnym polu widzenia (a) i w świetle spolaryzowanym (b), c) komórka algi *Micrasterias radiata*, kontrast DIC (źródło: https://en.wikipedia.org/wiki/File:Micrasterias_radiata.jpg, dostęp: 10.09.2020), d) pobudzone komórki, obraz fluorescencyjny (źródło: <https://www.leica-microsystems.com/science-lab/the-fundamentals-and-history-of-fluorescence-and-quantum-dots/>, dostęp: 10.09.2020)

Każda metoda ma swoje zalety oraz ograniczenia, dlatego należy ją dobrać w zależności od badanego zjawiska i materiału. Należy też wziąć pod uwagę, że niektóre techniki badawcze wymagają dodatkowego, niejednokrotnie drogiego wyposażenia i nie każdy mikroskop można odpowiednio rozbudować. Na przykład rzadkością są mikroskopy pozwalające na pracę zarówno w świetle przechodzącym, jak i w świetle odbitym. Szczegółowe informacje na temat technik mikroskopii świetlnej można znaleźć zarówno w specjalistycznych podręcznikach, jak i w licznych materiałach szkoleniowych, opracowywanych przez producentów mikroskopów.

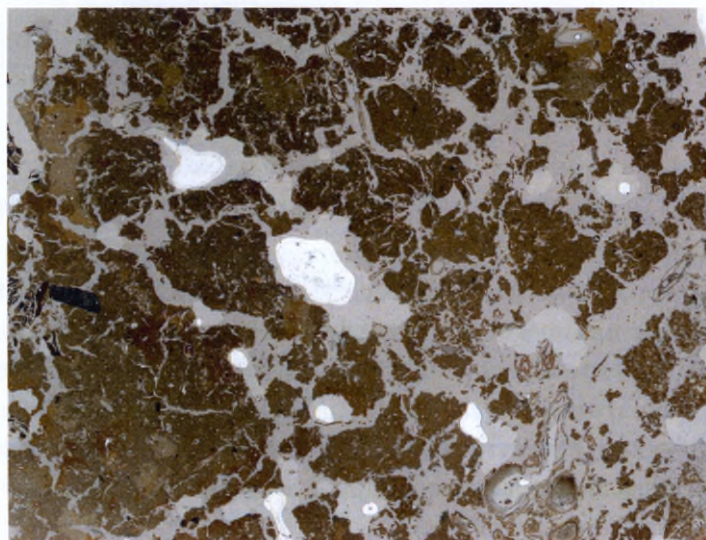


Il. 3.8. Mikroskopia skaningowa: a-b) spiekany węgiel wolframu WC w osnowie kobaltowej, elektrony BSE i SE (źródło: pliki demonstracyjne systemu Aphelion firmy ADCIS S.A.), c) wynik detekcji ziaren WC na podstawie obrazów (a) i (b)

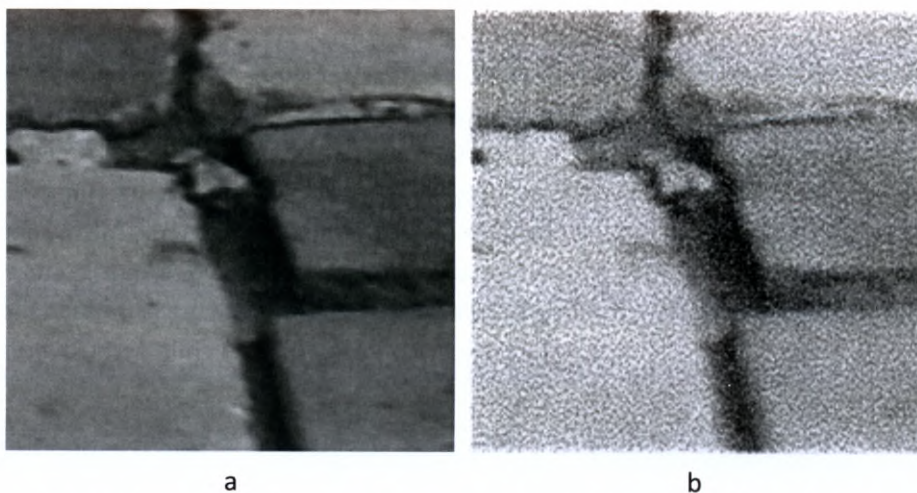
Podobnie bardzo dużo możliwości daje skaningowa mikroskopia elektronowa (SEM). Nawet najprostsze modele mogą generować obrazy pochodzące od elektronów wtórnych (SE, ang. *secondary electrons*) oraz odbitych lub, mówiąc poprawnie, sprężyscie wstecznie rozproszonych (BSE, ang. *Back-Scattered Electrons*). Liczba elektronów rejestrowanych przez detektory zależy w przypadku elektronów BSE od masy atomowej próbki (il. 3.8a), zaś w przypadku elektronów SE decydujące znaczenie ma topografia powierzchni (il. 3.8b). Jeżeli mamy wyszukać obiekty o zbliżonym składzie chemicznym, to zwykle należy wybrać tryb BSE. Jeżeli mikroskop skaningowy posiada moduł umożliwiający analizę składu chemicznego, to obraz gęstości rozmieszczenia atomów poszczególnych pierwiastków też będzie w tym pomocny. Oczywiście, jeżeli mamy dwa obrazy tego samego elementu, to można je wykorzystać np. do detekcji granic obiektów, gdyż każdy obraz da nam część potrzebnych informacji (il. 3.8c).

Jako źródło obrazów do analizy może też służyć skaner (il. 3.9). Jego wielką zaletą jest to, że zapewnia równomierne oświetlenie, co jest poważnym problemem podczas wykonywania zdjęć, a także obserwacji pod mikroskopem świetlnym.

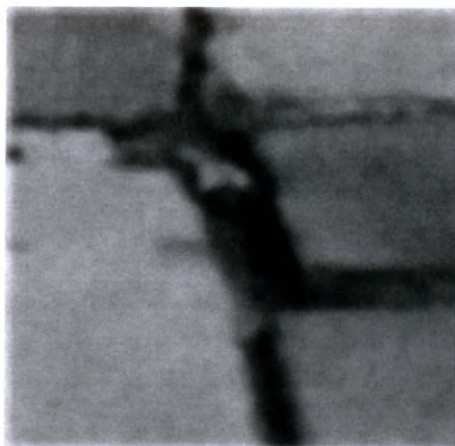
Jeżeli używamy skanera wysokiej jakości, to takie źródło obrazów do analizy jest w pełni godne polecenia, choć skaner kojarzy nam się zwykle z urządzeniem do wczytywania dokumentów. I właśnie to typowe zastosowanie może okazać się cennym źródłem obrazów, gdyż w przypadku potrzeby analizy archiwalnych danych możemy zostać poproszeni o analizę obrazu, który został wcześniej wydrukowany.



Il. 3.9. Obraz przekroju przez glebę, uzyskany za pomocą skanera (obraz dzięki uprzejmości prof. Krystyny Ciarkowskiej, Uniwersytet Rolniczy w Krakowie)

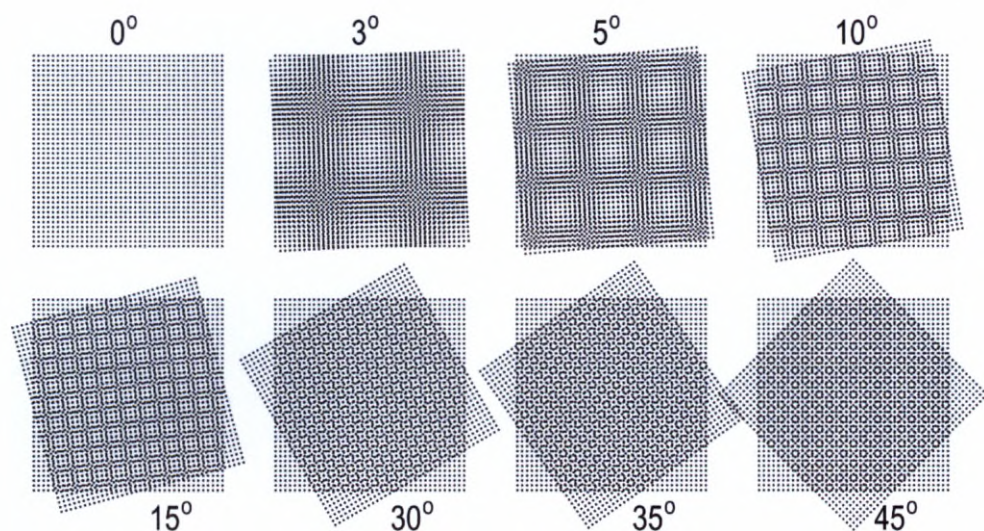


Il. 3.10. Drukowanie obrazu: a) obraz przeznaczony do druku, b) wydruk z drukarki atramentowej z rastrem stochastycznym, widoczny układ plamek przypominający szum



Il. 3.11. Próba odtworzenia obrazu początkowego na podstawie wydruku z il. 3.10b. Opis w tekście

Nie wchodząc w szczegóły techniki drukarskiej, należy pamiętać, że żaden sprzęt nie drukuje odcieni szarości, tylko symuluje je za pomocą układu drobnych czarnych plamek, nazywanego rastrem, który częściowo pochłania światło i dzięki temu daje wrażenie istnienia różnych odcieni szarości. Tak zwany raster stochastyczny, używany np. w drukarkach atramentowych, składa się z kropek o tej samej wielkości, ale o różnej gęstości (czyli rozmieszczonych bliżej lub dalej od siebie) i pozornie przypadkowym rozmieszczeniu. Wskanowany obraz takiego wydruku nie pokazuje ostrych granic poszczególnych plamek, co jest konsekwencją nierównej powierzchni kartki papieru i zjawiska wsiąkania tuszu w papier (il. 3.10b).



Il. 3.12. Zjawisko mory. Opis w tekście

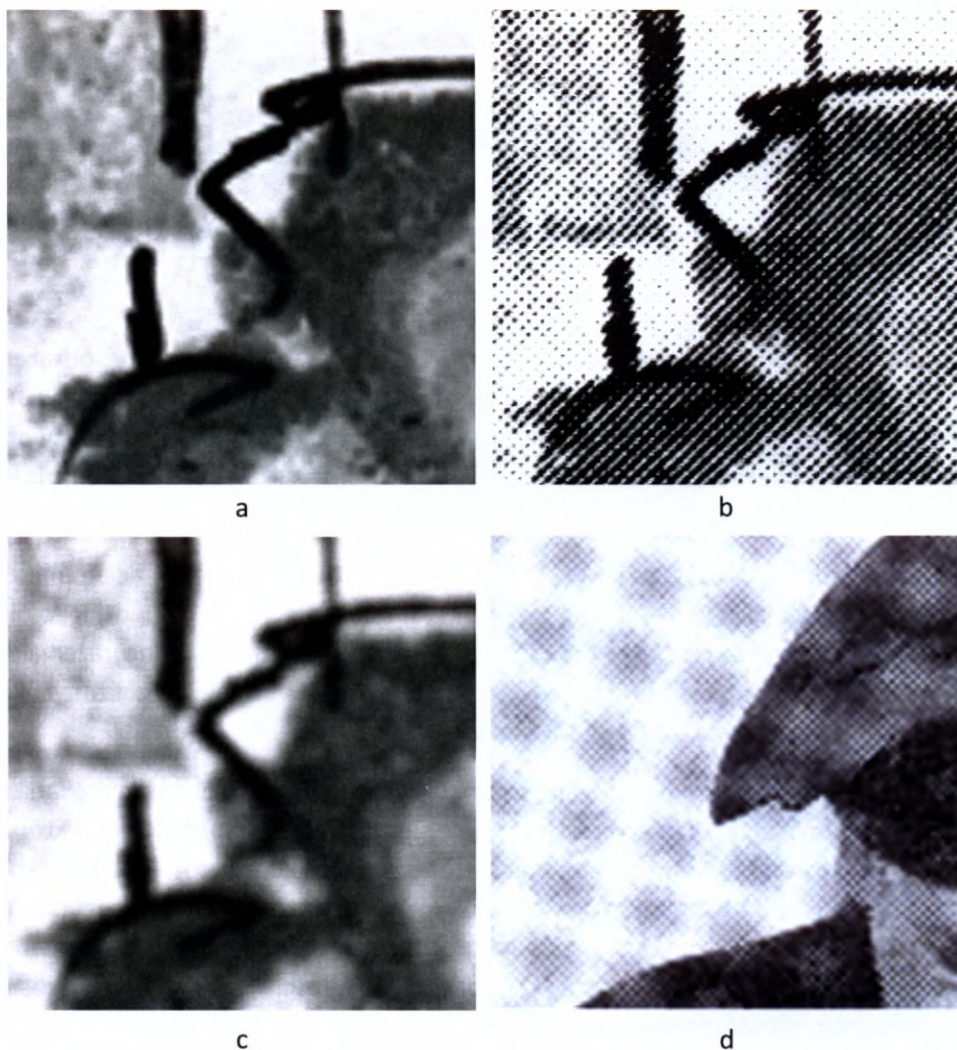
Jeżeli wydrukowany obraz wskanujemy, to możemy próbować odtworzyć obraz początkowy. Pewnych szczegółów nie uda się odtworzyć, ale wynik pokazany na il. 3.11 należy uznać za całkiem dobry. Użyte w tym celu narzędzia: filtr uśredniający oraz medianowy, operacje LUT i wyostrenie metodą ograniczającą szum, są opisane w kolejnych rozdziałach.

Przeprowadzane w obrębie rastra stochastycznego losowe przesunięcia punktów rastra w niewielkich granicach nie powodują zmian widocznych dla ludzkiego oka, a zabezpieczają przed zjawiskiem mory, które polega na pojawianiu się wrażenia dodatkowych struktur, np. prążków lub siatki, w wyniku nakładania się zbliżonych lub lekko obróconych względem siebie wzorów (il. 3.12).

Raster stochastyczny daje dobrą jakość druku, ale stawia przed sprzętem drukarskim większe wymagania. Dlatego częściej stosuje się tzw. raster amplitudowy, w którym punkty rastra są równomiernie rozmieszczone w siatce kwadratowej, która z kolei jest tworzona pod różnymi kątami. W przypadku druku czarno-białego jest to kąt 45°. Wybór tego kąta podnosi jakość druku, gdyż tworzone przez raster ukośne linie są położone bliżej siebie. Ich odległość to $\sqrt{2}/2$ odległości linii przy rozmieszczeniu węzłów rastra wzdłuż linii pionowych i poziomych. Punkty rastra buduje się z wielu mniejszych punktów. Wystarczy użyć 64 punktów, co umożliwi imitowanie w druku 64 odcieni szarości. Ta liczba przekracza zdolność ludzkiego zmysłu wzroku i zapewnia wystarczającą jakość druku. Punkty rastra mogą mieć różny kształt. Dwa przykłady rastrów o różnych kształtach punktów pokazuje il. 3.13b oraz 3.13d.

Raster amplitudowy jest bardziej wrażliwy na powstawanie efektu mory. Jego elementy wyraźnie widać na il. 3.13d. Zeskanowane obrazy, wydrukowane za pomocą rastra amplitudowego, również można przynajmniej częściowo odtworzyć (il. 3.13c), ale należy to traktować jako ostateczność.

Bardzo ciekawym, ale jednocześnie trudnym obszarem do potencjalnej współpracy jest analiza obrazów medycznych. Podczas współpracy z lekarzami trzeba pamiętać,



Il. 3.13. Skanowanie obrazu drukowanego: a) fragment wyjściowego pliku, b) skan wydruku tego fragmentu, c) próba odtworzenia obrazu wyjściowego na podstawie zeskanowanego wydruku, d) fragment innego wydruku, w którym użyto innego wzoru rastra amplitudowego; widoczne zjawisko mory



Il. 3.14. Fragment zdjęcia rentgenowskiego, obrazującego wszczepioną endoprotezę stawu biodrowego (zdjęcie otrzymane dzięki uprzejmości Uniwersyteckiego Szpitala Ortopedycznego w Zakopanem)

że skupiają się w swojej pracy na zagadnieniach medycznych, natomiast kwestie techniczne, które mogą decydować o sukcesie lub porażce analizy obrazu, są dla nich mniej istotne. Dlatego na osobach chcących analizować obrazy medyczne cięży obowiązek dbałości o kwestie techniczne.

Niezależnie od rozwoju metod tomograficznych wciąż bardzo ważnym elementem diagnostyki medycznej pozostaje wykonywanie zdjęć rentgenowskich (il. 3.14).

Pokazany powyżej fragment zdjęcia jest duży dlatego, żeby pokazać, jak wiele szczegółów potrafi zarejestrować współczesna aparatura rentgenowska. Trzeba jednak pamiętać, że jest ona optymalizowana pod kątem wygody oceny zdjęcia przez lekarza, a nie przez komputer. Dodatkowo operator sprzętu może w pewnych granicach zmieniać parametry pracy urządzenia, aby dostosować je np. do wielkości badanych organów, gdyż nie można badać przy tych samych parametrach kości małego dziecka i dorosłej osoby. Dlatego, jeżeli chcemy próbować wyciągać wnioski na podstawie oceny poziomów szarości zdjęcia rentgenowskiego, musimy zapewnić sobie



Il. 3.15. Zdjęcie rentgenowskie regeneratu kostnego powstającego podczas wydłużania kończyn. Widoczny jest aparat Ilizarowa stabilizujący kończynę oraz, z prawej strony, dodatkowy znacznik do korekty obrazu (zdjęcie otrzymane dzięki uprzejmości Uniwersyteckiego Szpitala Ortopedycznego w Zakopanem)

porównywalne parametry jego struktury. W tym celu najlepiej jest używać dodatkowego znacznika, którego przykład widać z prawej strony il. 3.15.

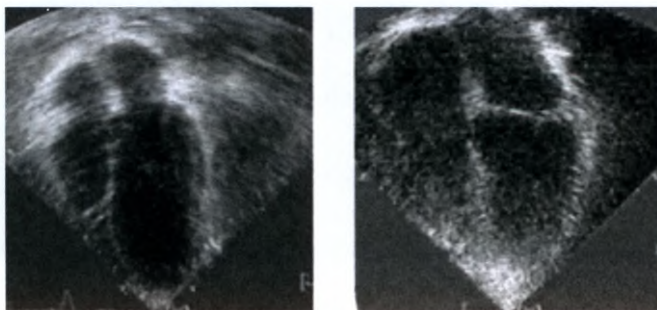
Widoczny na il. 3.15 znacznik powinien wyglądać identycznie na wszystkich zdjęciach. Jeżeli tak nie jest, to zdjęcie można łatwo skorygować, stosując odpowiednie przekształcenie LUT. Opis takiego przekształcenia znajduje się w następnym rozdziale, poświęconym m.in. operacjom punktowym.

Podczas analizy obrazów medycznych należy pamiętać o dwóch istotnych ograniczeniach. Po pierwsze obrazy powinny być zanonimizowane, czyli powinny być z nich trwale usunięte dane pozwalające osobom trzecim zidentyfikować pacjenta. Jednocześnie współpracujący z nami lekarz powinien tak zakodować opis zdjęcia w swoim systemie, aby mógł zidentyfikować pacjenta, np. w celu pobrania dodatkowej dokumentacji. Kolejną sprawą jest uzyskanie zgody komisji bioetycznej na przeprowadzenie naszych badań. Jest ona konieczna nawet wtedy, gdy korzystamy z materiałów archiwalnych, a pacjent nie jest poddawany żadnym nowym badaniom. Rozpoznawcze badania możemy prowadzić w uzgodnieniu z lekarzem bez takiej zgody, ale jest ona niezbędna podczas składania do druku publikacji w poważnych periodykach lekarskich.

Osobnym problemem jest analiza danych pochodzących z tomografii komputerowej lub rezonansu magnetycznego. Producenci sprzętu zazwyczaj bardzo niechętnie udostępniają pełną wersję zarejestrowanego materiału do analizy przez niezależne oprogramowanie. Jest to o tyle zrozumiałe, że chronią w ten sposób swoje zyski, które mogłyby zostać uszczuplone w przypadku pojawienia się konkurencyjnych systemów analizy. Co prawda pacjent dostaje płytkę z wynikami badania, ale zwykle jest na niej informacja, że zawartość nie nadaje się do celów diagnostycznych i ma jedynie poglądowy charakter.

Niemal zawsze dane z badań obrazowych są zapisywane w standardzie DICOM (ang. *Digital Imaging and Communications in Medicine*), który służy do ujednolicenia zapisu oraz wymiany danych medycznych związanych z diagnostyką obrazową. Bogactwo metod badań oraz specyfika poszczególnych narządów sprawiają, że zalecenia DICOM są bardzo rozbudowane i dokumentacja zapisana w tym standardzie wymaga często specjalizowanych programów do odczytu danych. O skali problemów związanych z danymi medycznymi może świadczyć fakt różnej wizualizacji tych samych danych, obserwowanych za pomocą różnych przeglądarek.

Łatwiej jest otrzymać dobrej jakości obrazy z badań ultrasonograficznych, gdyż producenci ultrasonografów często umożliwiają ich niezakłócony zapis na zewnętrznych nośnikach informacji. Niestety, dużym problemem jest interpretacja tych obrazów, gdyż zwykle wymagają do właściwej oceny solidnej wiedzy medycznej i doświadczenia. Ilustracją możliwych problemów jest il. 3.16, która przedstawia dwa obrazy serca w projekcji czterojamowej, czyli przy takim ustawieniu głowicy, że widoczne są jednocześnie oba przedsionki i obie komory. Różnica pomiędzy obrazami jest ewidentna. Ultrasonogram z lewej strony, przedstawiający serce dziecka w wieku ok. 10 lat, jest zdecydowanie wyraźniejszy i wydaje się być łatwiejszy do późniejszej analizy. Z kolei prawy obraz serca dotyczy osoby w wieku ok. 45 lat i sprawia wrażenie, jakby składał się z wielu

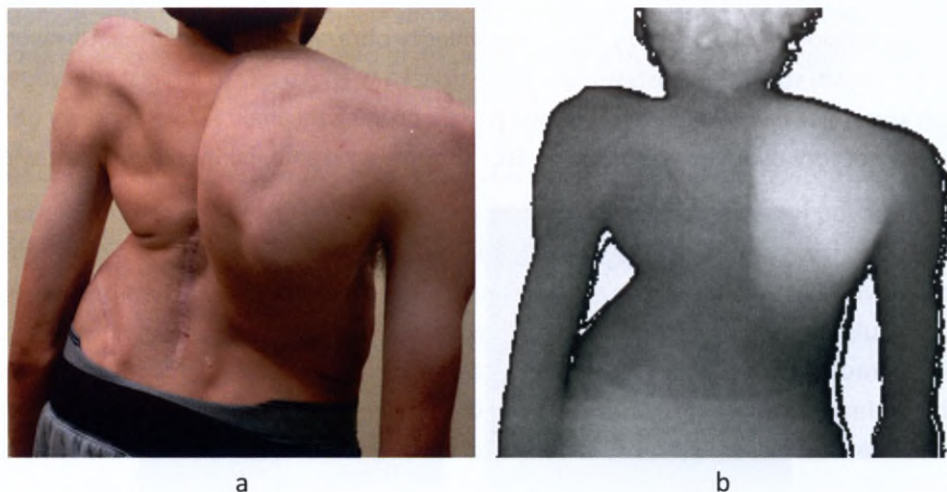


Il. 3.16. Obrazy ultrasonograficzne dwóch serc w tej samej projekcji. Różnice w wyglądzie wynikają z różnic w wieku pacjentów

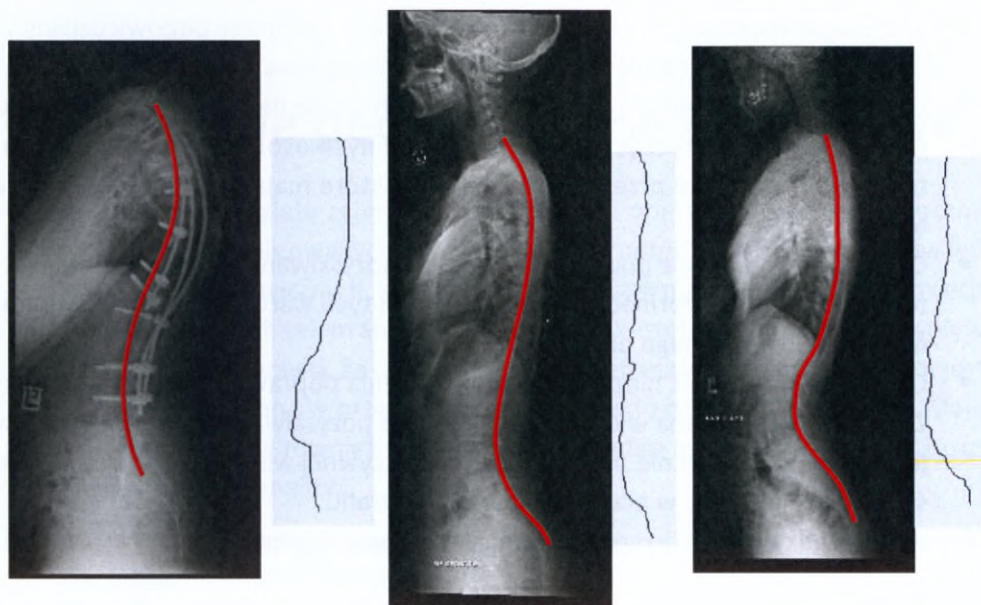
drobnych ziarenek. Jest to wynikiem naturalnych zmian zachodzących wraz z wiekiem w tkankach, na które nakłada się m.in. inna wielkość organu i grubość tkanki tłuszczowej. Jeżeli dodamy do tego różnych producentów sprzętu oraz świadomość dużego wpływu operatora ultrasonografu na wygląd obrazu, to okazuje się, że liczba czynników istotnie wpływających na obraz ultrasonograficzny praktycznie uniemożliwia budowę jednego, uniwersalnego, automatycznego programu do analizy obrazów ultrasonograficznych.

Obrazy mogą służyć do zapisu różnych danych, niekiedy wręcz abstrakcyjnych, jak np. kąty orientacji krawędzi lub powszechnie używane mapy, na których kolor danego punktu koduje jego wysokość nad poziomem morza. Obecnie dostępne są czujniki, np. Kinect, umożliwiające precyzyjną rejestrację odległości, która jest następnie pokazywana w postaci obrazu z poziomami szarości proporcjonalnymi do odległości obiektu od sensora (il. 3.17). Jakość wyników badań uzyskanych za pomocą tego typu urządzeń jest zaskakująco wysoka. W przypadku pokazanym na il. 3.17 dokładność pomiaru odległości wynosi ok. 1 mm. Jest to widoczne w dolnej części il. 3.17b, gdzie wyraźnie odcięty został obraz ubioru pacjenta, którego obecność zmienia odległość od sensora zaledwie o kilka milimetrów. Zatem można w ten sposób budować model 3D nieregularnej powierzchni, w tym przypadku pleców.

Porównanie kształtu środkowej linii kręgosłupa, wyznaczonego na podstawie zdjęcia rentgenowskiego wykonanego w projekcji bocznej (czerwone linie na il. 3.18), z zarysami uzyskanymi dzięki analizie obrazu z sensora Kinect (linie na prawo od zdjęć rentgenowskich) dowodzi możliwości wykorzystania tego typu obrazów do



Il. 3.17. Pacjent z bardzo silną deformacją kręgosłupa: a) widok pleców pacjenta, b) obraz uzyskany za pomocą sensora Kinect. Opis w tekście



Il. 3.18. Porównanie kształtu kręgosłupa wyznaczonego manualnie na podstawie zdjęcia rentgenowskiego (czerwona linia) oraz w wyniku analizy obrazu odległości otrzymanego z sensora Kinect (zdjęcia rentgenowskie otrzymane dzięki uprzejmości Uniwersyteckiego Szpitala Ortopedycznego w Zakopanem)

diagnostyki medycznej. Oczywiście, informacja o samej powierzchni jest dużo uboższa od tej, którą zawiera zdjęcie rentgenowskie, ale sensor odległości wykorzystujący promieniowanie podczerwone jest nieszkodliwy dla pacjenta i dlatego tego typu badanie może być wykonywane częściej, np. w celu śledzenia zmian zachodzących w trakcie długotrwałej terapii.

Problemy akwizycji obrazów są bardzo rozległe i zasługują na oddzielną książkę. Zostały omówione jedynie bardzo skrótowo, aby pokazać skalę problemów i ułatwić dyskusję z osobami dostarczającymi obrazy do analizy. Jest to konsekwencją założenia, że najczęściej będziemy pracowali z obrazami zarejestrowanymi przez kogoś innego, natomiast naszym zadaniem będzie ich analiza. Założenie to ma silne podstawy praktyczne. Trudno sobie wyobrazić, abyśmy zostali dopuszczeni do obsługi skomplikowanej medycznej aparatury diagnostycznej, wymagającej często długotrwałego szkolenia i mogącej przy nieprawidłowej obsłudze zaszkodzić pacjentowi. Podobnie będzie w przypadku przemysłowych systemów kontroli jakości. Z drugiej strony, zagadnienia analizy obrazu też nie są łatwe i zazwyczaj osoby obsługujące aparaturę badawczą lub diagnostyczną z chęcią przyjmują pomoc przy analizie uzyskiwanych obrazów.

Rozpoczynając współpracę w zakresie analizy obrazu z osobami odpowiedzialnymi za ich akwizycję, należy pamiętać, aby przedyskutować kilka ważnych kwestii:

- Czy pozyskiwanie obrazów do celów naszych analiz nie jest ograniczone aktami prawnymi, dotyczącymi np. ochrony danych osobowych, praw pacjenta lub wewnętrznych przepisów danej firmy, które mają chronić jej tajemnice technologiczne?
- Czy mamy zapewnione powtarzalne warunki pozyskiwania obrazów, a jeżeli nie, to czy posiadamy informację o wpływie zmian tych warunków na zmiany parametrów rejestrowanego obrazu?
- Czy mamy możliwość niezależnego sprawdzenia poprawności naszych analiz? Zdarza się, że używana wyrafinowana metoda pozyskiwania pewnych obrazów jest jedyną możliwą i nie ma możliwości obiektywnej weryfikacji jej poprawności. Tak dzieje się np. w przypadku nanotomografii.
- W przypadku wszystkich obrazów, niezależnie od źródła ich pochodzenia, należy omówić problem ich rozdzielczości. Powinna być na tyle duża, aby zapewnić możliwość detekcji poszukiwanych obiektów oraz ich pomiary z wymaganą dokładnością, a jednocześnie możliwie mała, aby nie wydłużać niepotrzebnie czasu analizy.

W ramach tego rozdziału nie były omawiane parametry techniczne ani zasady doboru sprzętu do pozyskiwania obrazów. Skupiamy się bowiem nie na akwizycji obrazów, ale na kolejnym kroku, jakim jest ich analiza. Poza tym należy pamiętać o ciągłym pojawianiu się coraz to nowszych urządzeń generujących obrazy. Również parametry techniczne pozornie takiego samego sprzętu ulegają nieustannym modyfikacjom i opisywanie konkretnych urządzeń narażałoby tekst na szybką dezaktualizację. Przykładem mogą być parametry cyfrowych aparatów fotograficznych. Około roku 2000 czołowe firmy oferowały aparaty cyfrowe z matrycą 1200x1600 pikseli i możliwością zapisu ok. 24 zdjęć na karcie pamięci. Obecnie ponaddziesięciokrotnie większa matryca 6000x4000 pikseli nie robi większego wrażenia, a karta pamięci pozwala zarejestrować nawet kilka tysięcy zdjęć. Należy też pamiętać o wpływie optyki obiektywu na jakość zdjęć. Niezależnie od ogromnego postępu w jakości zdjęć wykonywanych za pomocą smartfonów, trudno oczekiwać takiej samej jakości zdjęcia jak wykonanego przez lustrzanke z najwyższej jakości profesjonalnym obiektywem, który kosztuje tyle, ile kilka wysokiej klasy smartfonów.

4. Operacje geometryczne i punktowe

W ramach tego rozdziału zajmiemy się szeregiem operacji, które mają znaczny wpływ na wygląd analizowanych obrazów. Fakt ten odróżnia je od wielu filtrów lub przekształceń morfologicznych, których zadaniem jest wzmocnienie wybranych cech lub obiektów będących celem analizy, a nie dbałość o estetykę obrazów. Silny wpływ na wygląd obrazów sprawia, że część operacji geometrycznych oraz punktowych jest powszechnie stosowana w przetwarzaniu obrazów do celów artystycznych. Mają jednak duże znaczenie również w analizie obrazu, gdzie są używane bardzo często na wstępnym etapie analizy.



a



b

Il. 4.1. Wykorzystanie przekształceń geometrycznych do korekcy reprodukcji rysunku: a) obraz zarejestrowany aparatem fotograficznym, b) końcowy wynik

Jak sama nazwa wskazuje, przekształcenia geometryczne służą do korekty geometrii obrazów. Obrazy możemy przesuwac w różnych kierunkach, obracać, zwiększać lub zmniejszać, wykonywać lustrzane odbicia, jak również zmieniać ich proporcje, czyli skracać, wydłużać, poszerzać lub zwężać. Do przekształceń geometrycznych możemy też zaliczyć wycinanie części obrazu, czyli branie do dalszych prac jedynie jego fragmentu, który zawiera elementy potrzebne nam do analizy.

Często zachodzi potrzeba korekty kształtu fragmentu obrazu, co pokazuje przykładowo rys. 4.1. Widać na nim (il. 4.1a) reprodukcję rysunku wykonanego na kartce formatu A4. Kartka nie mogła w całości wypełnić kadru ze względu na różnice w proporcjach. Kartka A4 ma wymiar 210x297 mm, czyli krótszy bok stanowi ok. 70,7% dłuższego boku. Natomiast aparat fotograficzny wykonywał zdjęcia w formacie 4000x6000 punktów, czyli krótszy bok zarejestrowanego obrazu stanowi ok. 66,7% dłuższego boku. Dodatkowo zdjęcie było wykonywane w mało komfortowych warunkach, gdy trzeba było trzymać aparat w rękach. Dlatego nie udało się zachować prostopadłości osi optycznej do fotografowanego obiektu i jego kształt odbiega od prostokąta. Aby uniknąć utraty części obrazu fotografowanego rysunku, nie zajmuje on całego zdjęcia i wokół kartki papieru widać margines, zawierający obraz fragmentów podłogi.

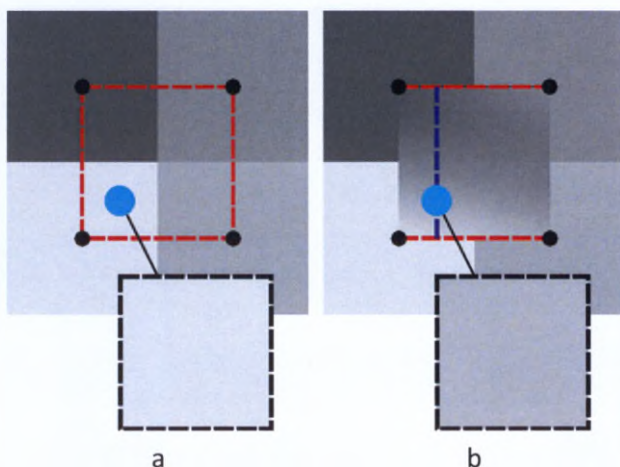
W celu korekcji obrazu z il. 4.1a wycięto fragment o proporcjach formatu A4, zawierający cały rysunek. Następnie, przesuwając rogi reprodukcji, doprowadzono do poprawnego końcowego wyniku (il. 4.1b). Potrzebne w tym zadaniu zniekształcenia (ang. *distortions*) można nazwać wypaczeniem (ang. *warping*). Mówiąc obrazowo, tego typu przekształcenia traktują obraz jak arkusz bardzo elastycznej gumy, którą możemy niemal dowolnie naciągać lub ścisnąć, pod warunkiem jednak, że zniekształcany arkusz cały czas znajduje się w jednej płaszczyźnie. Zatem nie ma mowy o przestrzennym zniekształcaniu obrazu.

Jak łatwo się domyślić, przemieszczenia poszczególnych pikseli zwykle nie będą wielokrotnościami odległości pomiędzy pikselami początkowego obrazu. Pojawia się zatem problem, jak wyznaczyć nowe wartości pikseli po przekształceniu. Stosuje się w tym celu różne metody interpolacji. Najpopularniejsze są trzy z nich:

- metoda najbliższego sąsiada (ang. *nearest neighbor*, il. 4.2a),
- metoda dwuliniowa (ang. *bilinear*, il. 4.2b),
- metoda dwusześcienna (ang. *bicubic*, il. 4.3).

Analizowane przez nas obrazy są mozaikami, które obserwowane przy znacznym powiększeniu stają się zbiorami przylegających do siebie kwadratów o jednolitym zabarwieniu lub poziomie szarości. Każdy kwadrat odpowiada pojedynczemu pikselowi. Podczas dalszej dyskusji o sposobach wyznaczania nowych wartości pikseli będziemy brali pod uwagę i nazywali pikselami punkty będące środkami ciężkości tych kwadratów.

Najprostsza jest metoda najbliższego sąsiada. Załóżmy, że mamy wyznaczyć nową wartość piksela, którego położenie symbolizuje małe niebieskie kółko na il. 4.2a. Nowy piksel otoczony jest czterema pikselami z aproksymowanego obrazu, oznaczonymi za pomocą nieco mniejszych czarnych kółek na il. 4.2a. Narysowany czerwoną przerywaną

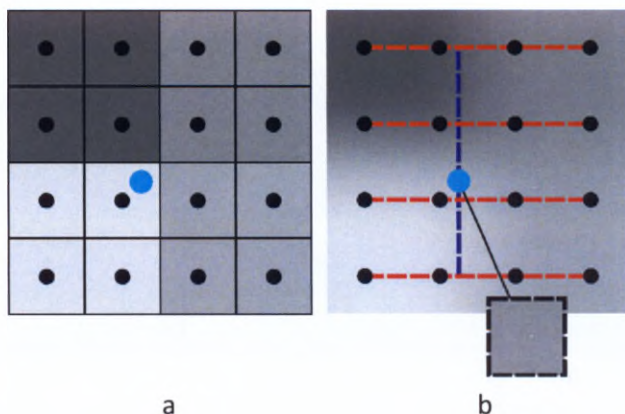


Il. 4.2. Różne metody interpolacji: a) metoda najbliższego sąsiada, b) metoda dwuliniowa.
Opis w tekście

linią kwadrat zaznacza swoimi narożnikami piksele brane pod uwagę podczas aproksymacji i symbolizuje obszar, w którym znajduje się nowy piksel. Każdy początkowy piksel znajduje się w środku kwadratu, który ma poziom szarości danego piksela i jednocześnie wyznacza obszar, w którym odległość do tego piksela jest mniejsza niż do jakiegokolwiek innego. Nowa wartość piksela odpowiada poziomowi szarości tej strefy najbliższej odległości, w której znalazł się nowy piksel. Całość jest bardzo prosta, natomiast wytłumaczenie sposobu wyboru nowej wartości jest już nieco trudniejsze. Poziom szarości nowego piksela pokazuje kwadrat otoczony czarną przerywaną linią.

W metodzie dwuliniowej najpierw wyznacza się dwa równania liniowe, opisujące przejście pomiędzy dwoma parami pikseli (połączonych czerwonymi przerywanymi liniami na il. 4.2b). Następnie wyznacza się równanie opisujące zmiany poziomów szarości wzdłuż kolejnej linii, prostopadłej do poprzednich i jednocześnie przechodzącej przez piksel, którego wartość mamy wyznaczyć (ciemnoniebieska linia na il. 4.2b). W przestrzeni pomiędzy czarnymi kółkami na il. 4.2b znajduje się kwadrat wypełniony gradientowo, czyli zmieniającym się poziomem szarości. Ilustruje on, jak w metodzie dwuliniowej zmienia się wartość aproksymowanego piksela w zależności od jego położenia względem pikseli dających wartości do aproksymacji. Teraz można bez problemu wyznaczyć nową wartość piksela. Jak widać, jest ona inna od wartości, którą otrzymaliśmy, używając metody najbliższego sąsiada.

Najbardziej skomplikowana jest metoda dwusześcienna. W jej przypadku bierzemy pod uwagę nie 4, tylko 16 pikseli z otoczenia aproksymowanego punktu (il. 4.3a). W pierwszej kolejności wyznacza się cztery funkcje aproksymujące



Il. 4.3. Interpolacja dwusześcienna: a) konfiguracja pikseli, b) schemat interpolacji.
Opis w tekście

zamiany wartości sąsiednich pikseli. Przebieg tych funkcji symbolizują cztery odcinki, narysowane na il. 4.3b czerwoną przerywaną linią. Następnie, podobnie jak w przypadku metody dwuliniowej, wyznacza się kolejną funkcję aproksymującą, obejmującą aproksymowany piksel. Jako funkcji aproksymujących używa się najczęściej wielomianów, ale nie jest to jedyne możliwe rozwiązanie. Zmienny poziom szarości tła na il. 4.3b symbolizuje charakter zmian wartości dawanych przez funkcje aproksymujące.

Opisane powyżej metody aproksymacji dobrze ilustrują problemy pojawiające się podczas redukcji (zmniejszania) wymiarów obrazu. Czasami jednak potrzebujemy powiększyć obraz. Postępowanie jest wtedy analogiczne. Jeżeli nowy piksel znajdzie się w obszarze brakujących pikseli, czyli wirtualnej dziury wywołanej oddalaniem się pikseli podczas rozciągania powiększanego obrazu, to bierzemy pod uwagę wartości pikseli znajdujących się w bezpośrednim sąsiedztwie tego obszaru.

Na il. 4.4 pokazano wyniki aproksymacji omawianymi powyżej metodami podczas czterokrotnego powiększenia niewielkiego obrazu widocznego na il. 4.4a. Obraz referencyjny, czyli taki, do jakiego dążymy w tym zadaniu, zawiera il. 4.4b. Na il. 4.4c-e pokazano kolejno wyniki aproksymacji metodami: najbliższego sąsiada, dwuliniową i dwusześcienną. Żadna z metod nie pozwoliła na powstanie obrazu zbliżonego do ideału pokazanego na il. 4.4b. Trudno jednak się temu dziwić, skoro liczba pikseli w początkowym obrazie (il. 4.4a) wynosi zaledwie 6,25% liczby pikseli obrazu z il. 4.4b. Przekształcenia, których wyniki pokazano na il. 4.4, są używane w tzw. zoomie cyfrowym, spotykanym w niektórych kompaktowych aparatach fotograficznych lub kamerach. Jak widać, narzędzie to tylko pozornie pozwala zastąpić fizyczną



a



b



c



d



e

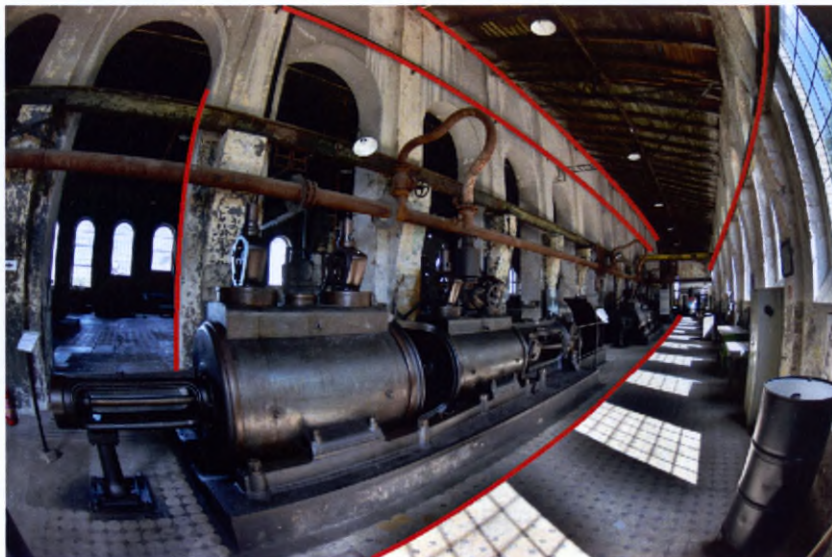
Il. 4.4. a) niewielki fragment obrazu podlegający procesowi czterokrotnego powiększenia, b) obraz referencyjny, którego uzyskanie byłoby idealnym rozwiązaniem, c) aproksymacja metodą najbliższego sąsiada oraz metodami: d) dwuliniową i e) dwusześcienną

zmianę ogniskowej aparatu. Może jednak okazać się akceptowalnym rozwiązaniem w przypadku nagrywania plików wideo, gdzie ze względu na bezwładność zmysłu wzroku tolerujemy gorszą jakość obrazu bez odczuwalnego dyskomfortu.

Metoda najbliższego sąsiada dała najgorszy wizualnie wynik, zwłaszcza w obszarze biegnących lekko skośnie rurek balustrady (il. 4.4c). Z drugiej strony metoda ta posiada pewną cechę, która może okazać się istotna podczas analizy. Otóż podczas wyznaczania nowych wartości pikseli nie wprowadza się do obrazu żadnych wartości, które w nim nie występowały przed przekształceniem. Może to stanowić istotny element w niektórych analizach.

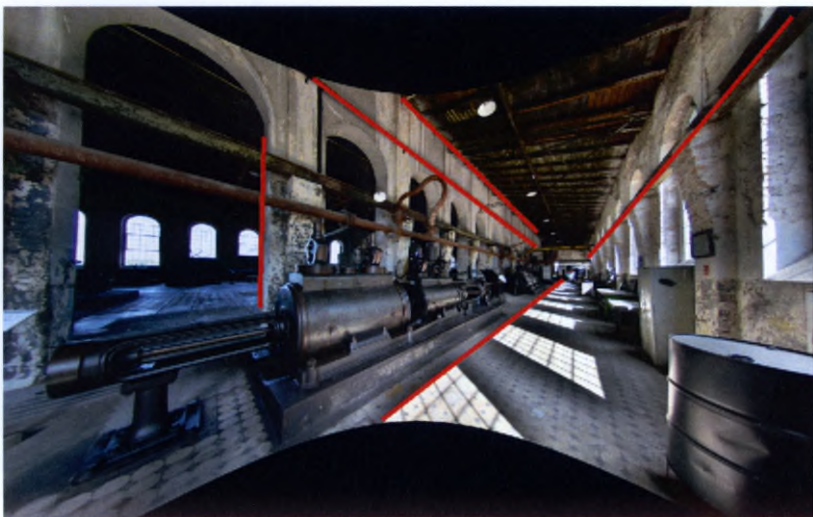
Powstaje oczywiście pytanie, na ile celowe jest sztuczne powiększanie liczby pikseli obrazu, skoro prowadzi ono do tak poważnych błędów. Trzeba zwrócić uwagę, że w przypadku widocznym na il. 4.4 zakres zmian był bardzo duży. Zrobiono to w celu wyraźnego pokazania różnic pomiędzy różnymi metodami aproksymacji. Najczęściej niezbędne korekty sprowadzają się do zniekształceń rzędu 5-10% (jak np. na il. 4.1) i wtedy metody aproksymacji sprawdzają się bardzo dobrze.

Ważnym obszarem zastosowania metod aproksymacyjnych jest korygowanie wadliwego odwzorowania obiektów wynikającego z wad układów optycznych. Na il. 4.6 pokazano wynik korekcji obrazu wykonanego z użyciem obiektywu typu rybie oko (il. 4.5). W przypadku zaawansowanych aparatów cyfrowych wraz z danymi obrazowymi zapisywane są parametry ekspozycji, a wśród nich rodzaj obiektywu i długość jego ogniskowej. Dzięki temu możliwe jest automatyczne dopasowanie sposobu



Il. 4.5. Zdjęcie fragmentu zabytkowej odlewni cynku, wykonane obiektywem typu rybie oko o kącie widzenia blisko 180°. Bardzo szeroki kąt widzenia jest okupiony dużymi zniekształceniami geometrycznymi, powodującymi zakrzywienie prostych linii.

Kilka przykładowych zakrzywionych linii zaznaczono na czerwono

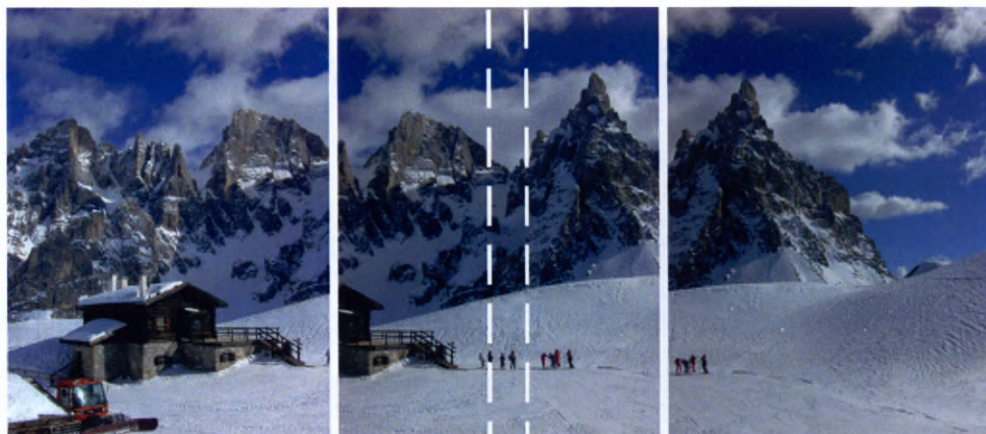


Il. 4.6. Korekcja obrazu z il. 4.5. Pierwotnie zakrzywione linie zostały wyprostowane. Czarne obszary u dołu i na górze zdjęcia są miejscami, dla których pierwotne zdjęcie nie zarejestrowało żadnej informacji

korekcji do wad wprowadzanych przez optykę. Odpowiednie profesjonalne oprogramowanie bywa udostępniane bez dodatkowych opłat przez producentów sprzętu fotograficznego. Ze skorygowanego obrazu z il. 4.6 można wyciąć fragment nieobejmujący pól pozbawionych informacji, co prowadzi do wyniku pokazanego na il. 4.7. Takiego obszaru nie da się sfotografować standardowym obiektywem.



Il. 4.7. Końcowy wynik korekcji. Skorygowany obraz obejmuje kąt widzenia niedostępny dla standardowych obiektywów



Il. 4.8. Trzy zdjęcia wykonane w Dolomitach. Po lewej stronie wybudowane w 1936 r. schronisko Baita Segantini (ok. 2200 m n.p.m.), z prawej szczyt Pale di San Martino (3192 m n.p.m.), a w środku zdjęcie obejmujące fragmenty obu skrajnych fotografii



Il. 4.9. Końcowy wynik składania

Do operacji geometrycznych można zaliczyć również składanie wielu obrazów w jeden. Dość efektowny przykład pokazuje połączenie trzech zdjęć z il. 4.8 w jeden fotomontaż, pokazany na il. 4.9. Ze względu na ok. 1000 metrów różnicy wysokości konieczne było wykonanie zdjęć w układzie pionowym. Końcowe zdjęcie (il. 4.9) obejmuje kąt ok. 120° , czyli $1/3$ widnokregu. Obecnie niektóre aparaty fotograficzne, wbudowane np. w smartfony, pozwalają na wykonanie takiego panoramicznego zdjęcia w sposób automatyczny. Wystarczy wolno się obracać i utrzymywać aparat w przybliżeniu na jednym poziomie.

Podczas wykonywania tego typu zdjęć należy unikać obiektów w pobliżu fotografującego, gdyż zostaną bardzo silnie zniekształcone (il. 4.10). Dobre efekty da natomiast wykonanie panoramy górskiego krajobrazu ze szczytu wieży widokowej. Podczas



Il. 4.10. Panoramiczne zdjęcie portu w Aarhus (Dania). Fragment nabrzeża w kształcie łuku jest w rzeczywistości prostoliniowy

analizy zdjęć, obejmujących szeroki zakres kątów, należy pamiętać o wprowadzonych przez nie zniekształceniach, które mogą mieć istotny wpływ na wyniki końcowe. Jest to szczególnie ważne, gdy jednym z celów analizy jest zbadanie wielkości zarejestrowanych w obrazie obiektów.

Składanie wielu obrazów w jeden nie ogranicza się do zastosowań artystycznych. Często jest wykonywane w celu uzyskania obrazu większego, niż pozwalają na to warunki techniczne, np. pole widzenia mikroskopu. Połączenie wielu obrazów umożliwia analizę obiektów, których rozmiary nie pozwalają na objęcie go tylko jedną fotografią. Takie rozwiązanie ogranicza również błędy ilościowej oceny, spowodowane przecięciem części obiektów przez brzeg obrazu.

Z obrazami będącymi złożeniem wielu mniejszych fragmentów mamy do czynienia podczas korzystania z takich programów jak np. Google Earth (il. 4.11). Najczęściej



Il. 4.11. Fragment obrazu okolic Olszanicy koło Krakowa, dostępny w Google Earth (dostęp: 23.08.2020). Przebiegająca przez środek obrazu lekko rozmyta granica rozdziela obszary o różnej rozdzielczości. W części widocznej po prawej stronie znajduje się znacznie więcej szczegółów.

nie dostrzegamy miejsc łączenia poszczególnych zdjęć. Niekiedy staje się to jednak dość wyraźne, zwłaszcza w miejscach, gdzie autorzy postanowili ograniczyć szczegółowość udostępnianych danych.

Przejdziemy teraz do całego szeregu operacji, nazywanych często operacjami punktowymi. Określenie pochodzi stąd, że o nowej wartości przekształcanego punktu (piksela) decyduje wyłącznie początkowa wartość tego punktu, niezależnie od tego, jakie wartości przyjmują sąsiedzi. Jest to inna sytuacja niż podczas przekształceń geometrycznych, gdzie nowa wartość jest zwykle wyznaczana na podstawie wartości kilku sąsiadów (il. 4.2-4).

Zacniemy od operacji modyfikujących histogram. Histogram obrazu to wykres pokazujący liczebności pikseli o poszczególnych wartościach. Początkowy obraz, który będziemy przekształcali, jest dosyć ciemny i nie widać na nim białych obszarów (il. 4.12a). Analiza histogramu tego obrazu tylko to potwierdza, gdyż mamy dużo pikseli o wartościach nieprzekraczających 50 i jednocześnie nie mamy żadnych pikseli o wartościach powyżej 180 (il. 4.12c).

Stan ten możemy łatwo poprawić, przeprowadzając normalizację obrazu. Jest to przekształcenie, w którym wartości poszczególnych pikseli zmieniamy zgodnie z funkcją liniową, tak aby najciemniejsze punkty osiągnęły wartość zerową, natomiast najjaśniejsze osiągnęły wartość 255. Zakładamy przy tym, że przekształcamy 8-bitowy obraz szary. Jeżeli najniższa wartość piksela w obrazie wynosi $p_{\min}$, a najwyższa wartość wynosi $p_{\max}$, to dowolny piksel z tego zakresu, mający wartość p , uzyska po normalizacji nową wartość p_n , taką że:

$$p_n = 255 \frac{p - p_{\min}}{p_{\max} - p_{\min}} \quad (4.1)$$

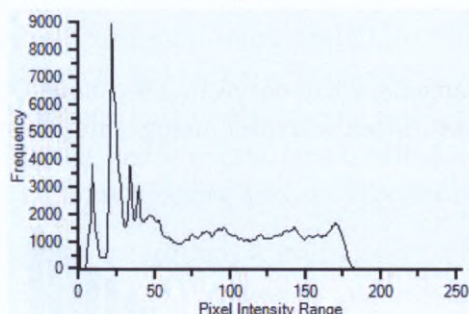
Wartość p_n powinna zostać zaokrąglona do najbliższej liczby całkowitej. Obraz po normalizacji widoczny jest na il. 4.12b. Ze względu na rozciągnięcie zakresu poziomów szarości od czerni do bieli, szczegóły na obrazie są lepiej widoczne. Histogram poziomów szarości (il. 4.12d) zajmuje teraz pełny zakres od 0 do 255. Ponieważ początkowych wartości, użytych do normalizacji, było mniej niż 256 (było to ok. 180 wartości, co widać na il. 4.12c), to histogram po normalizacji wygląda nieco podobnie do grzebienia, gdyż posiada sporo miejsc o wartości zerowej. Nie stanowi to jednak żadnej przeszkody dla naszego zmysłu wzroku, gdyż rozróżniamy nie więcej niż ok. 50 poziomów szarości. Przeprowadzona normalizacja obrazu pokazuje zalety zapisywania obrazów szarych w postaci 8-bitowej. Nadmiar poziomów szarości



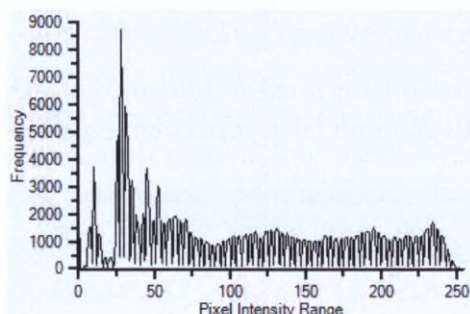
a



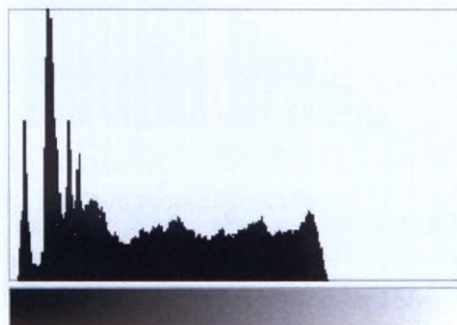
b



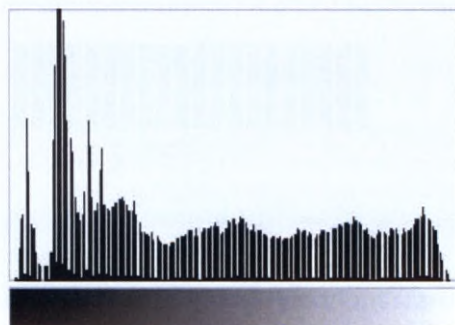
c



d



e



f

Il. 4.12. Przekształcenia punktowe: a) obraz początkowy, b) obraz po normalizacji, c) histogram obrazu (a), d) histogram obrazu (b), e-f) inne formy graficzne histogramów z il. (c) oraz (d)

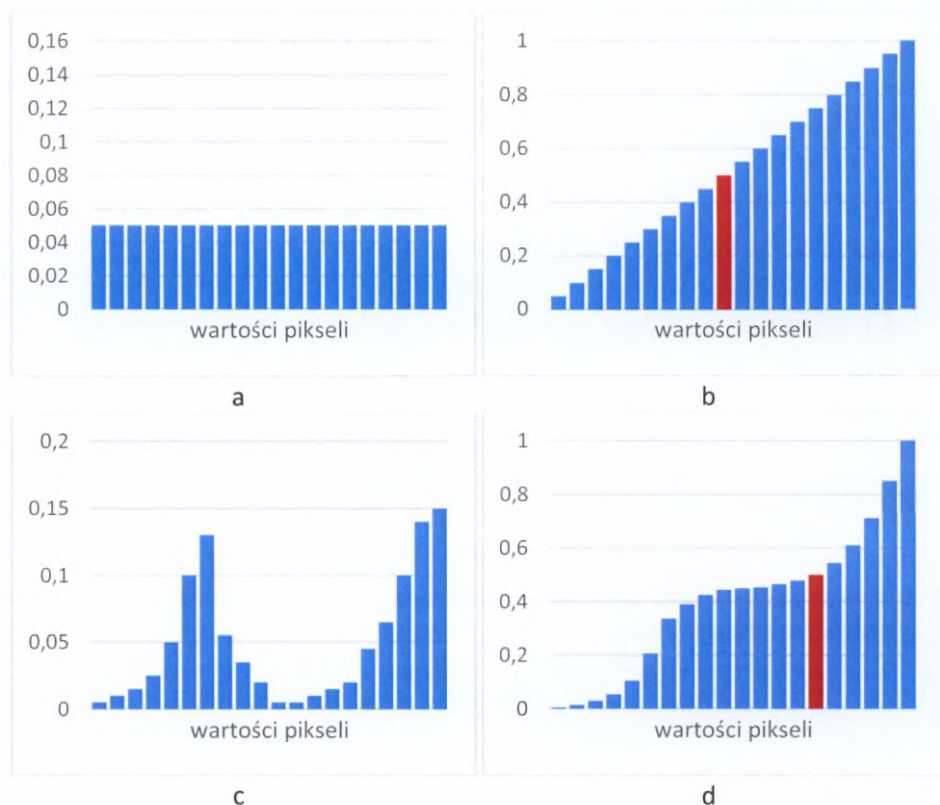
w stosunku do możliwości ludzkiego oka pozwala na różne operacje, powodujące powstanie przerw w histogramie bez wizualnej utraty jakości obrazu.

Normalizację przeprowadza się często jako jeden z pierwszych kroków podczas przetwarzania obrazu. Wprawdzie dla komputera wizualna poprawa obrazu nie ma żadnego znaczenia, ale przeprowadzenie normalizacji może ułatwić dalszą obróbkę

obrazu. Jeżeli będziemy analizowali serię obrazów przedstawiających elementy podobne, lecz zarejestrowane w różnych warunkach, to mogą być one ciemniejsze lub jaśniejsze i mniej lub bardziej kontrastowe. Normalizacja sprawi, że w przybliżeniu będą te obrazy wyglądały tak, jakby zostały zarejestrowane w tych samych warunkach. Umożliwia to użycie jednego algorytmu do obróbki całej serii obrazów, bez potrzeby jego modyfikacji.

W kolejnych rysunkach będziemy używali nieco innej formy graficznego przedstawienia histogramów (il. 4.12e-f). Jest ona wprawdzie pozbawiona osi z wartościami liczbowymi, ale wyraźniej widać np. nieciągłości histogramów (por. il. 4.12d i 4.12f). Ponieważ nie odczytujemy konkretnych wartości z histogramów, to wybór tej formy graficznej wydaje się być uzasadniony.

Przejdziemy teraz do kolejnego przekształcenia, które nazywamy wyrównaniem histogramu (ang. *histogram equalization*). Jeżeli wartości histogramu obrazu



Il. 4.13. Przykładowe rozkłady oraz ich dystrybuanty: a) rozkład równomierny, b) dystrybuanta rozkładu równomiernego, c) rozkład nierównomierny i d) jego dystrybuanta

podzielimy przez całkowitą liczbę pikseli w obrazie, to otrzymamy wartości prawdopodobieństwa przyjęcia przez piksel danej wartości, a histogram (bez zmiany swojego kształtu) będzie przedstawiał rozkład prawdopodobieństwa. Jeżeli wszystkie wartości pikseli występują z takim samym prawdopodobieństwem, to histogram ma wszystkie słupki tej samej wysokości (il. 4.13a) i mówimy, że rozkład prawdopodobieństwa jest równomierny.

Można ten rozkład przedstawić w nieco zmienionej formie, która jest nazywana dystrybuantą (il. 4.13b). Dystrybuanta dla danego poziomu szarości jest sumą prawdopodobieństwa występowania tej wartości oraz prawdopodobieństw dla wszystkich wartości mniejszych od tego poziomu szarości. Jak łatwo przewidzieć, wykres dystrybuanty rozkładu równomiernego jest linią prostą (il. 4.13b). Każda dystrybuanta, niezależnie o rozkładu prawdopodobieństwa, jest funkcją niemalejącą, czyli jej wartość dla kolejnych poziomów szarości będzie albo rosła, albo pozostawała stała, natomiast nigdy nie będzie malała. Poza tym maksymalna wartość dystrybuanty zawsze wynosi 1. Można też powiedzieć, że dystrybuanta jest całą rozkładu prawdopodobieństwa.

Rzeczywiste rozkłady wartości pikseli w obrazie zwykle nie są równomierne, co ukazuje np. il. 4.12. Wyrównywanie histogramu to operacja, która ma przybliżyć rzeczywisty histogram obrazu do równomiernego. Operacja ta zmienia wartości poszczególnych pikseli w sposób, który można sobie wyobrazić na podstawie analizy il. 4.13. Gdyby piksele o nierównomiernym rozkładzie (il. 4.13c) miały rozkład równomierny, to wartość piksela, dla którego dystrybuanta osiąga wartość 0,5 (czerwony słupek na il. 4.13d) powinna być taka sama, jak wartość piksela odpowiadająca czerwonemu słupkowi na il. 4.13b. Przeliczenie wartości poszczególnych pikseli jest niezbyt skomplikowane.

Zakładamy, że w obrazie można zapisać n stopni szarości, od wartości 0 do $n - 1$. W obrazie 8-bitowym będzie to 256 wartości od 0 do 255. Każdy poziom szarości i jest reprezentowany przez p_i pikseli, przy czym mogą to być również wartości zerowe. Niech jednocześnie j oznacza najniższą wartość poziomu szarości, który jest reprezentowany przez niezerową liczbę punktów. Możemy to zapisać w postaci warunku:

$$\text{jeżeli } i < j, \quad \text{to } p_i = 0 \quad (4.2)$$

Najpierw wyznaczmy wartości dystrybuanty $D(i)$ dla każdej wartości piksela:

$$D(i) = \frac{\sum_{k=0}^i p_k}{\sum_{k=0}^{n-1} p_k} \quad (4.3)$$

Następnie wyznaczamy dla każdego poziomu szarości i jego nową wartość, którą możemy oznaczyć jako $LUT(i)$:

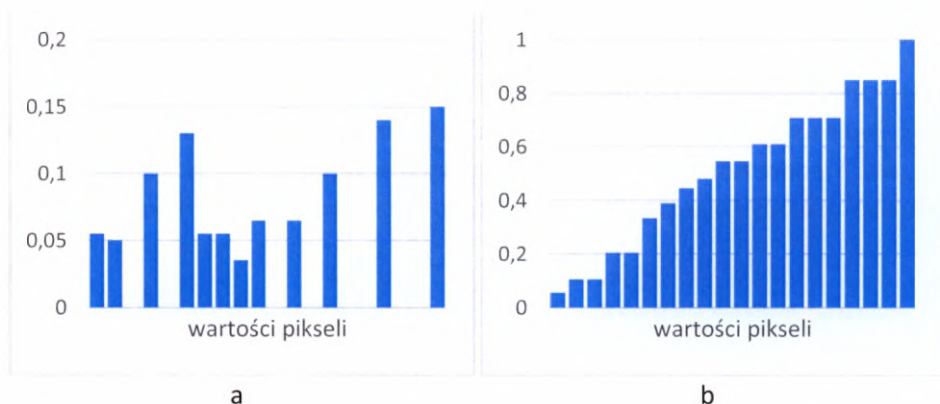
$$LUT(i) = \begin{cases} 0 & \text{dla } i < j \\ \frac{D(i) - D(j)}{1 - D(j)}(n-1) & \text{dla } i \geq j \end{cases} \quad (4.4)$$

Wynik należy zaokrąglić do wartości całkowitej. W dalszej części rozdziału wyjaśnimy, dlaczego wynik równania (4.4) oznaczyliśmy symbolem LUT . Może się zdarzyć, że dla różnych wartości i otrzymamy taką samą wartość $LUT(i)$. Nie zdarzy się jednak tak, żeby jeden punkt jaśniejszy od drugiego przed wyrównaniem histogramu stał się od niego ciemniejszy po wyrównaniu histogramu. Można to zapisać w postaci następującego twierdzenia:

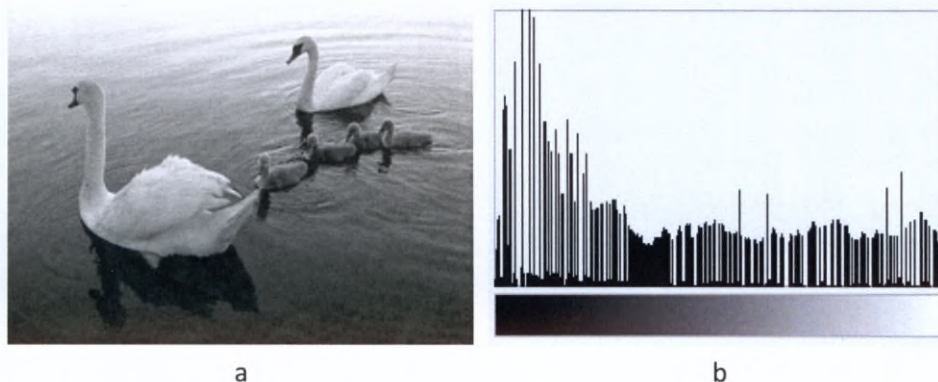
$$\text{jeżeli } LUT(k) > LUT(i), \quad \text{to } k > i \quad (4.5)$$

Histogram po wyrównaniu może wykazywać znaczne nieciągłości i lokalnie wysokie wartości, tak jak pokazano to na il. 4.14a, który przedstawia wyrównany histogram z il. 4.13c. Warto jednak zwrócić uwagę, że pomimo wyraźnej nieregularności, dystrybuenta po wyrównaniu histogramu (il. 4.14b) jest bliższa histogramowi rozkładu równomiernego, niż dystrybuenta widoczna na il. 4.13d.

Obraz po wyrównaniu histogramu pokazuje il. 4.15a. Zarówno obraz, jak i jego histogram wykazują znaczne podobieństwo do wyników normalizacji, pokazanych na il. 4.12b i 4.12d, ale nie jest to wcale normą. Często spotykamy obrazy, które nie



Il. 4.14. Wyrównanie histogramu: a) rozkład na il. 4.13d po wyrównaniu histogramu i b) jego dystrybuenta



Il. 4.15. Przekształcenia punktowe: obraz z il. 4.12a po wyrównaniu histogramu (a) i jego histogram (b)

ulegają zmianie w wyniku normalizacji, gdyż ich piksele wykazują wartości zarówno 0, jak i 255. Wystarczy jeden piksel o wartości 255, aby cały obraz, który jest bardzo ciemny, nie uległ jakiegokolwiek zmianie po normalizacji. Za to duże zmiany może przynieść wyrównanie histogramu.

Najczęściej wyrównanie histogramu uwidacznia szczegóły, które nie były wcześniej widoczne z powodu słabego kontrastu. Obraz wydaje się bardziej kontrastowy i dynamiczny (il. 4.16a-b). Zdarza się jednak, zwłaszcza w przypadku obrazów przedstawiających bardzo ciemne obiekty na bardzo jasnym tle, że wyrównanie histogramu istotnie pogorszy wygląd obrazu i utrudni jego dalszą obróbkę (il. 4.16c-d). Nie zmienia to faktu, że wyrównanie histogramu jest często bardzo pomocne w analizie obrazu. Bliżej zagadnienie to omówimy w rozdziale poświęconym binaryzacji.

Przejdziemy teraz do kolejnej grupy przekształceń, które najczęściej określa się mianem LUT. W istocie są to również przekształcenia histogramu, podobnie jak omówione już: normalizacja i wyrównanie histogramu. Jest jednak pewna istotna różnica. Normalizacja, tak samo jak wyrównanie histogramu, wymaga indywidualnych obliczeń dla każdego obrazu. Natomiast te przekształcenia, które zaczynamy omawiać, mogą mieć uniwersalny algorytm dla wszystkich obrazów, o ile tylko zostały zarejestrowane z taką samą głębią bitową. Zatem operacje te mogą być wykonane bez żadnych modyfikacji na przykład na każdym 8-bitowym obrazie szarym, jeżeli tylko zostały zdefiniowane dla takiego obrazu.

Nazywanie działania omawianych teraz operacji modyfikowaniem histogramu jest nieco mylące, gdyż w istocie modyfikujemy wartości poszczególnych pikseli, a zmiana histogramu jest jedynie konsekwencją tych modyfikacji. Czasami dochodzi



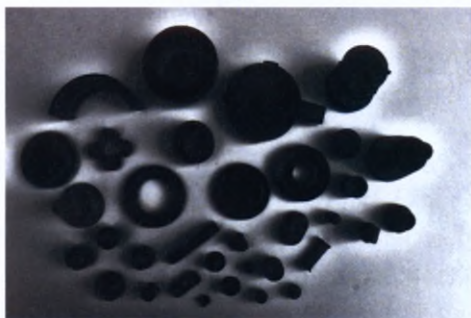
a



b



c



d

Il. 4.16. Nocne zdjęcie ulicy (a) z dobrze uwidocznionymi szczegółami po wyrównaniu histogramu (b) oraz obraz bardzo ciemnych wyrobów gumowych na jasnym tle (c), który został niekorzystnie zmieniony w wyniku wyrównania histogramu (d)

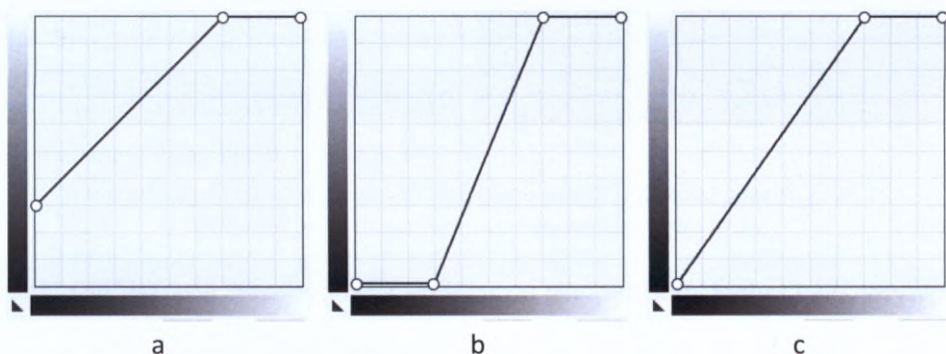
do paradoksalnych sytuacji, na przykład wtedy gdy operacja nazywana wyrównywaniem histogramu prowadzi do jeszcze bardziej nieregularnej jego postaci (por. il. 4.13c i 4.14a).

Nazwa LUT jest akronimem angielskiego terminu *Look-Up Table*, co można przetłumaczyć jako „tabela sposobu wyświetlania” lub „tablica podstawień”. Pierwsze tłumaczenie oddaje efekt zastosowania LUT: obraz wygląda inaczej. Natomiast drugie tłumaczenie lepiej odzwierciedla fakt, że sprawa nie ogranicza się do wyświetlania, gdyż w obrazie wartości pikseli zostają trwale zmienione. Wyjaśnienia wymaga jeszcze powód pojawienia się słowa: tabela.

Jeżeli przekształcamy wartości pikseli 8-bitowego obrazu szarego w zestaw nowych wartości, które również mieszczą się w zakresie 8 bitów, to możemy do tego celu użyć różnych, nawet bardzo złożonych funkcji. Podstawowym ograniczeniem tych funkcji jest to, że dla argumentów będących liczbami całkowitymi z zakresu 0-255 wynik funkcji również ma być liczbą całkowitą z tego zakresu.

Wyliczanie nowych wartości dla każdego piksela może być w przypadku złożonych funkcji czasochłonne. Całą procedurę można jednak znacznie uprościć. Mamy ograniczoną liczbę 256 argumentów funkcji i taką samą liczbę jej wartości. Wystarczy zatem stworzyć dwukolumnową tabelę, zawierającą wartości początkowe i końcowe dla wszystkich 256 wartości poziomów szarości i zapamiętać ją. Wyznaczenie nowych wartości pikseli nie wymaga już wyliczania wartości funkcji. Wystarczy znacznie prostsze i szybsze podstawienie odpowiedniej wartości z tabeli podstawień. Ponieważ cała operacja zmienia wygląd obrazu, to nazwano ją Look-Up Table.

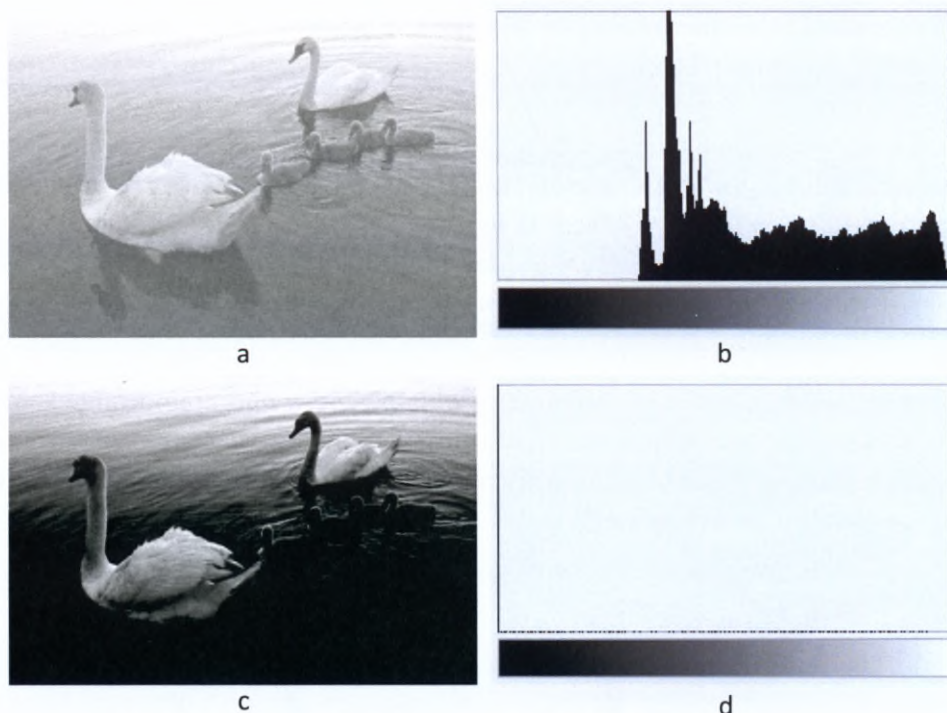
Przekształcenia LUT są bardzo szybkie i mogą być wykonywane nawet w czasie rzeczywistym. Odpowiednie tabele LUT są wbudowane w oprogramowanie kamer i aparatów fotograficznych, dzięki czemu możemy wybrać zdjęcia z zimniejszymi lub cieplejszymi kolorami, z kolorami o intensywnym lub neutralnym nasyceniu lub dopasować balans bieli do oświetlenia.



Il. 4.17. Przebiegi funkcji stanowiących podstawę LUT: a) zwiększającej jasność, b) zwiększającej kontrast i c) zwiększającej intensywność

W dalszej części przebiegi funkcji stanowiących podstawę LUT będą przedstawiane na wykresach takich jak na il. 4.17. Na osi odciętych oraz na osi rzędnych mamy wartości od 0 do 255. Aby wykresy były bardziej poglądowe, zamiast liczbowych wartości opisujących osie mamy paski ze zmiennym poziomem szarości, odpowiadającym wartościom liczbowym.

Il. 4.17a pokazuje przebieg funkcji realizującej rozjaśnienie obrazu. Polega ono po prostu na dodaniu do każdego poziomu szarości tej samej liczby. Jeżeli suma przekroczy wartość 255, to nową wartością zostaje właśnie ta liczba. Możemy zatem stracić część informacji zapisanej w najjaśniejszych partiach obrazu. Efekt rozjaśnienia obrazu z il. 4.12a jest pokazany na il. 4.18a. Ponieważ początkowy obraz nie zawierał pikseli o wartości powyżej 180 (il. 4.12b), to nie osiągnęły one po rozjaśnieniu wartości



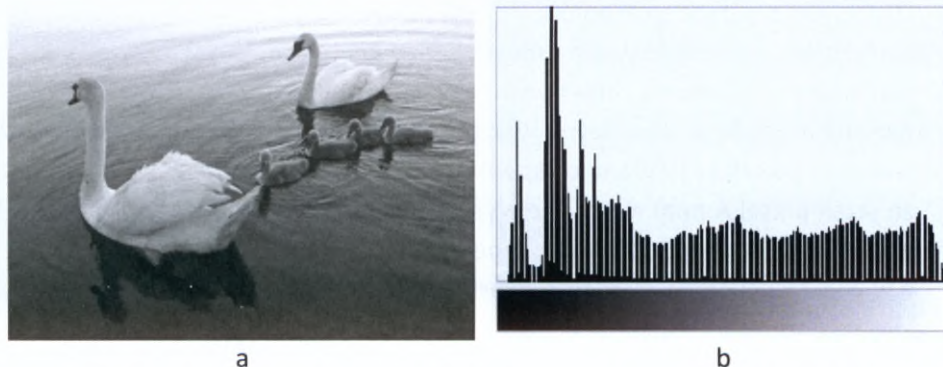
Il. 4.18. Przekształcenia LUT: a-b) obraz o podniesionej jasności i jego histogram, c-d) obraz o podwyższonym kontraście i jego histogram

powyżej 255 i żadnej informacji nie utraciliśmy. W konsekwencji kształt histogramu poziomów szarości nie uległ w tym przypadku zmianie i został on tylko przesunięty w stronę wyższych wartości (il. 4.18b). Oczywiście, można przeprowadzić odwrotną operację, czyli przyciemnienie obrazu, nazywane też obniżaniem jasności. Wtedy od każdej wartości piksela odejmujemy tę samą liczbę, a tym pikselem, dla których różnica jest liczbą ujemną, przypisujemy wartość 0. Możemy zatem stracić część informacji zapisanej w najciemniejszych partiach obrazu.

Kolejną operacją wykonywaną z użyciem LUT jest zmiana kontrastu. Funkcja pokazana na il. 4.17b ilustruje podwyższenie kontrastu. W tym przekształceniu piksel o wartości 128 dla obrazu 8-bitowego, nazwijmy go środkowym, nie ulega zmianie. Piklele o wartościach wyższych od środkowego mają wartości podwyższane proporcjonalnie do różnicy wartości tego piksela i piksela środkowego. W analogiczny sposób są zmniejszane wartości pikseli o wartości mniejszej od piksela środkowego. Podobnie jak w przypadku jasności, nowe wartości są obcinane na poziomie 255 dla wartości większych od 255 lub na poziomie 0 dla wartości mniejszych od 0. Możemy zatem utracić część informacji zarówno w najciemniejszych, jak i w najjaśniejszych obszarach obrazu.

W przypadku obrazu widocznego na il. 4.18d straciliśmy wszystkie szczegóły w najciemniejszych jego obszarach. Jest jednak i dobra wiadomość – podniesienie kontrastu sprawia, że obraz wydaje się ostrzejszy. Histogram obrazu z podniesionym kontrastem (il. 4.18d) sprawia wrażenie pozbawionego wartości. Tak jednak nie jest. Ze względu na przyjęcie przez bardzo dużą część pikseli (jest to blisko połowa z nich) wartości zerowej, tworzą one bardzo wysoki i wąski pik na początku histogramu. Maksimum osi rzędnych na il. 4.18d jest ok. 15-krotnie wyższe niż na il. 4.18b, ale nie widzimy tego, gdyż prezentowane histogramy nie są opisane liczbowo. Wysokości pozostałych słupków histogramu są w porównaniu z wartością początkową tak małe, że na histogramie na il. 4.18d stały się niewidoczne.

Często zdarza się, że w obrazie mamy wartości pikseli bliskie 0, natomiast maksymalne wartości są wyraźnie niższe niż 255. Obraz na il. 4.12a jest tego dobrym przykładem. W tym przypadku możemy dokonać korekcji, będącej połączeniem wzrostu jasności i kontrastu (il. 4.17c). Nazwiemy ją zmianą intensywności. Ponieważ zmiana wartości pikseli ma charakter liniowy do momentu osiągnięcia przez piksele wartości 255, a jednocześnie obraz z il. 4.12a zawiera piksele o wartościach z zakresu ok. 4-180, to zarówno obraz, jak i histogram na il. 4.19 są bardzo zbliżone do wyniku normalizacji obrazu (il. 4.12b i 4.12d).

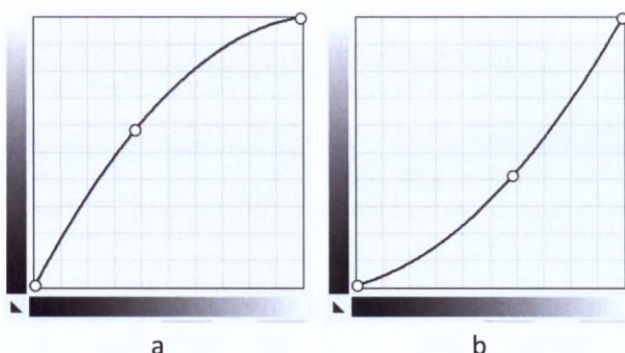


Il. 4.19. Przekształcenia LUT: obraz o podniesionej intensywności (a) i jego histogram (b)

Przejdziemy teraz do omawiania LUT, dla których wykresy tworzących je funkcji mają przebieg krzywoliniowy. Zaczniemy od gamma modulacji (il. 4.20). Ta bardzo popularna metoda korekcji powstała na potrzeby korekcji obrazów tworzonych przez skaningowe mikroskopy elektronowe, gdy jeszcze nie było komputerowego przetwarzania obrazu.

Jeżeli początkową wartość pikseli oznaczmy przez x , a wynik korekcji przez y , to przekształcenie odbywa się dla obrazów 8-bitowych zgodnie z zależnością:

$$y = 255 \left(\frac{x}{255} \right)^{\frac{1}{\gamma}} \quad (4.6)$$



Il. 4.20. Krzywe korekcji gamma dla: a) $\gamma > 1$, b) $\gamma < 1$

Z analizy il. 4.20 i wzoru (4.6) wynikają następujące cechy gamma modulacji:

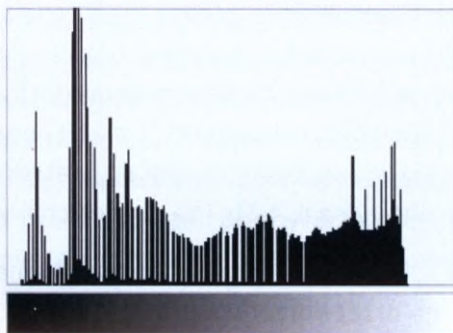
- skrajne wartości zakresu poziomów szarości (0 oraz 255) nie ulegają zmianie,
- wartości $\gamma > 1$ powodują rozszerzenie początkowego zakresu poziomów szarości w ciemnych fragmentach obrazu, tym samym rozjaśniają cały obraz i uwypuklają szczegóły w cieniach, co dzieje się kosztem obszarów jasnych, gdzie możemy stracić pewną część szczegółów,
- wartość $\gamma = 1$ nie wprowadza żadnych zmian w obrazie,
- wartości $\gamma < 1$ wywołują efekt odwrotny w stosunku do $\gamma > 1$,
- gamma modulacja, również w wersji używanej w cyfrowych obrazach, gdzie wartości pikseli są liczbami całkowitymi, jest funkcją niemalejącą, co oznacza, że jeżeli piksel A miał w obrazie wyjściowym wartość większą niż piksel B, to po przekształceniu na pewno nie będzie miał wartości mniejszej niż piksel B, a w większości przypadków jego wartość nadal będzie większa.

Szczególnie warta uwagi jest ostatnia wymieniona cecha gamma modulacji. Dzięki niej wynik przetwarzania obrazu wygląda bardzo naturalnie.

Zastosowanie gamma modulacji z wartością $\gamma > 1$ (il. 4.21a) uwypukliło grę świateł na powierzchni wody, która była najciemniejszym elementem il. 4.12a. Nie straciliśmy informacji w jasnych partiach obrazu, gdyż obraz początkowy nie zawierał bardzo jasnych punktów. Rozszerzenie zakresu ciemnych tonów znalazło swoje odbicie w histogramie (il. 4.12b), który utracił ciągłość po swojej lewej stronie. Ponieważ nasze oko rozróżnia do ok. 50 poziomów szarości, to te nieciągłości w histogramie nie przekładają się na dostrzegalne błędy w obrazie.



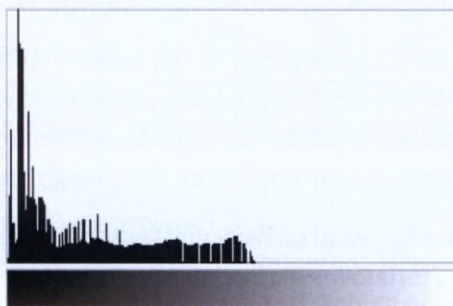
a



b



c



d



e



f

Il. 4.21. Gamma modulacja obrazu z il. 4.12a oraz histogramy: a-b) $\gamma > 1$, c-d) $\gamma < 1$, e-f) $\gamma < 1$, przekształcenie wykonane na obrazie po normalizacji

Wspomniany powyżej brak bardzo jasnych pikseli w początkowym obrazie (il. 4.12a) sprawia, że obraz po modulacji z wartością $\gamma < 1$ stał się ciemniejszy i widać na nim zdecydowanie mniej szczegółów (il. 4.21c). Potwierdza to obserwacja histogramu (il. 4.21d) pokazującego, że zdecydowana większość pikseli ma wartości odpowiadające miejscom ciemniejszym niż 50% szarości.

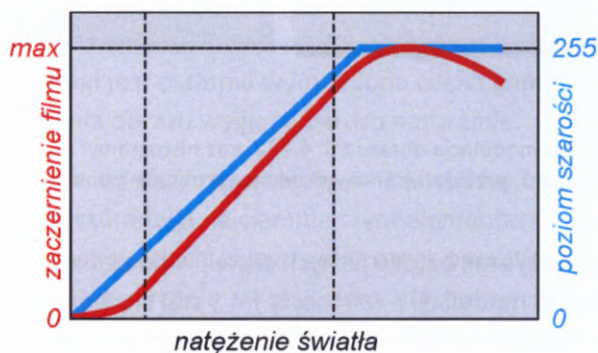
Jeżeli jednak przed gamma modulacją z wartością $\gamma < 1$ wykonamy normalizację, to uzyskamy interesujący efekt, pokazany na il. 4.21e. Co prawda, woda stanie się jeszcze ciemniejsza, ale za to możemy dostrzec subtelną grę światła i cieni na powierzchni piór łabędzi. Histogram (il. 4.21f) potwierdza rozszerzenie zakresu tonów jasnych. Dzięki wcześniejszej normalizacji zakres tonów ciemnych jest szerszy niż w przypadku obrazu na il. 4.21c i na powierzchni wody widać nieco więcej szczegółów.

W komputerowej analizie obrazu gamma modulację możemy wykorzystać do wstępnego przetworzenia obrazów w celu ułatwienia późniejszej manualnej detekcji obiektów. Może być też przydatna w ujednolicaniu serii obrazów, tak aby mogły być dalej przetwarzane za pomocą tego samego algorytmu.

Gamma modulacja jest używana w sterowaniu sposobem wyświetlania kolorów na monitorach komputerów. Zmieniając wartości gamma (domyślna wartość to 2,2), możemy skorygować sposób wyświetlania kolorowych obrazów. Zagadnienie to wykracza jednak poza ramy niniejszego podręcznika.

Chwilę uwagi poświęcimy teraz omówieniu reakcji na działanie światła tradycyjnych materiałów światłoczułych oraz matryc we współczesnych aparatach cyfrowych. Możemy to prześledzić na schemacie tzw. krzywej densytometrycznej (il. 4.22).

Na początku musimy pamiętać, że cała analiza, którą przeprowadzimy za chwilę, dotyczy stałego, niezmiennego czasu naświetlania. W przypadku klasycznej błony fotograficznej i bardzo słabego światła nie następują żadne przemiany, nawet gdybyśmy bardzo wydłużyli czas ekspozycji. Inaczej mówiąc, nie zarejestrujemy żadnego obrazu. Stan ten odpowiada lewemu końcowi czerwonej krzywej na il. 4.22. W miarę wzrostu natężenia światła błona fotograficzna zaczyna zmieniać swój poziom zaczernienia (ujawniamy to dopiero po wywołaniu filmu) i w pewnym zakresie



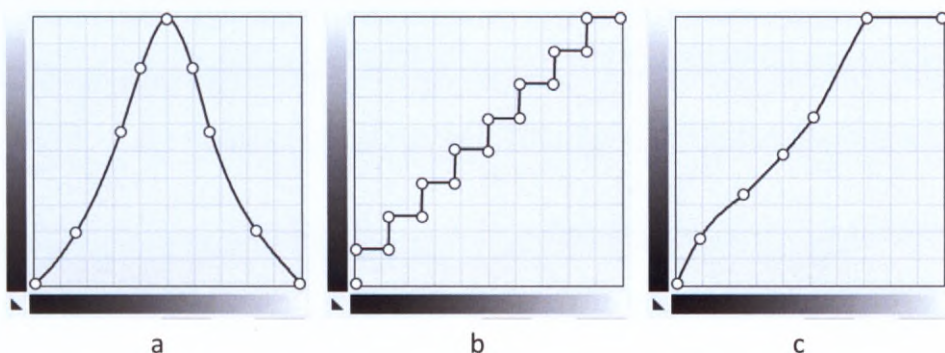
Il. 4.22. Schemat reakcji na światło klasycznego filmu (czerwona krzywa) oraz światłoczułej matrycy aparatu cyfrowego (niebieska linia)

jej zaczernienie jest w przybliżeniu proporcjonalne do natężenia światła. Ten fragment krzywej densytometrycznej, znajdujący się na il. 4.22 między pionowymi przerywanymi liniami, starano się wykorzystywać w klasycznej fotografii. Dostarczenie większej ilości promieniowania świetlnego prowadzi do osiągnięcia stanu nasycenia, a następnie do odwrócenia zaczernienia, czyli powstania częściowo obrazu pozytywowego. Dalsze zwiększanie natężenia światła prowadzi do tzw. przeświecenia błony fotograficznej, na której niczego już nie zarejestrujemy. Zjawisko częściowego odwrócenia zaczernienia prześwieczonego filmu nazywamy solaryzacją i bywało wykorzystywane do celów artystycznych.

W przypadku aparatów cyfrowych ich charakterystyka jest zbliżona do przedstawionej na il. 4.22 za pomocą niebieskiej linii. Charakterystyka ta jest w całości liniowa, a po przekroczeniu granicznego poziomu natężenia światła mamy zarejestrowaną po prostu białą plamę. Zjawisko solaryzacji tutaj nie zachodzi, ale możemy je zasymulować (il. 4.23a).

Liniowa charakterystyka sprawia, że możemy rejestrować obrazy już przy bardzo słabym oświetleniu, co było niemożliwe w przypadku klasycznej fotografii. Niestety, przy słabym oświetleniu występuje duży stosunek szumu do sygnału i dlatego obrazy rejestrowane w tych warunkach mają wysoki poziom szumów, co obserwowaliśmy w przypadku il. 3.2a. Można temu zaradzić przez wydłużenie czasu naświetlania, które uśrednia szumy i daje poprawę jakości obrazu (por. il. 4.31), ale nie można wtedy rejestrować szybko zmieniających się obiektów, bo zostaną zapisane jako rozmazane.

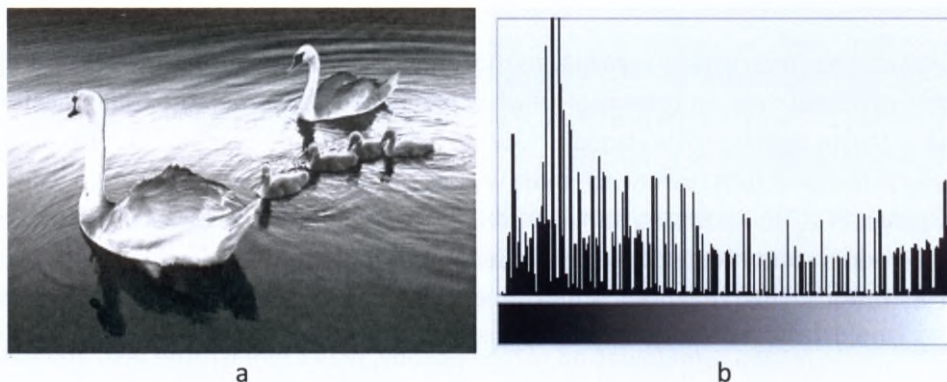
Solaryzację można symulować na różne sposoby, przy czym istotne jest, aby funkcja wykorzystana do tworzenia LUT miała maksimum dla średnich wartości poziomu szarości, co umożliwi uzyskanie obrazu częściowo negatywowego (il. 4.23a).



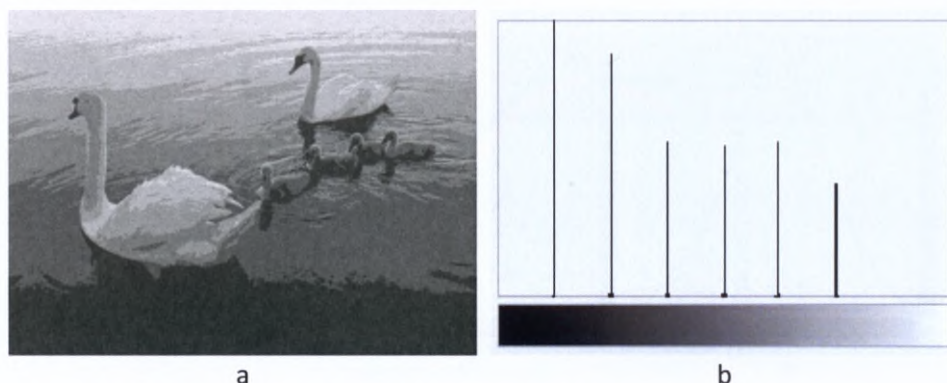
Il. 4.23. Przebiegi funkcji stanowiących podstawę LUT: a) symulacja solaryzacji, b) ograniczenie liczby poziomów szarości, c) manualna korekcja obrazu

Zastosowanie tego przekształcenia daje silny efekt graficzny (il. 4.24a), co bywa wykorzystywane przez artystów fotografików. Możliwości wykorzystania tego przekształcenia w analizie obrazu wydają się być bardzo ograniczone, a odpowiedni histogram (il. 4.24b) również nie wnosi informacji, które byłyby bardzo przydatne w analizie obrazu. Pewną ciekawostką może być to, że na granicy przejścia pomiędzy obrazem pozytywowym i negatywowym pojawia się cienka, jasna linia, która odpowiada maksymalnej wartości pikseli obrazu.

Kolejnym przekształceniem jest ograniczenie liczby poziomów szarości, co uzyskuje się dzięki LUT, opracowanemu na podstawie schodkowej funkcji modyfikującej wartości pikseli (il. 4.23b). Przekształcenie to zamienia obraz na szereg plam o jednolitym poziomie szarości (il. 4.25a) lub zabarwieniu w przypadku obrazów kolorowych. Na histogramie widzimy jedynie odizolowane piki o wysokości proporcjonalnej do wielkości obszarów zajętych przez odpowiednie poziomy szarości (il. 25b).



Il. 4.24. a) obraz po symulacji solaryzacji oraz b) histogram tego obrazu



Il. 4.25. Drastyczne ograniczenie liczby poziomów szarości (a) oraz odpowiadający mu histogram poziomów szarości (b)

Omawiane przekształcenie jest często wykorzystywane przez grafików, zwłaszcza podczas projektowania plakatów. Z tego względu w oprogramowaniu do przetwarzania obrazu występuje pod nazwą posteryzacji (ang. *poster* – ‘plakat’). Nie jest ono jednak wymysłem grafików komputerowych. Blisko 90 lat temu, w roku 1932, polski fotograf i profesor chemii, Witold Romer, opracował metodę pozwalającą na ograniczenie liczby odcieni w fotografii. Metoda otrzymała nazwę izohelii. Polegała na wykonaniu z jednego negatywu wielu bardzo kontrastowych odbitek o różnych czasach naświetlania. Następnie były one fotografowane, najlepiej na błonach o możliwie dużym formacie. Tak przygotowane negatywy nakładało się na siebie i z takiego pakietu wykonywało ostateczną odbitkę. Była to dość czasochłonna metoda, bo filmy i odbitki w tradycyjnej fotografii należy suszyć, a ponadto na każdym etapie trzeba poświęcić czas na wywoływanie, utrwalanie i płukanie. W praktyce wykonanie jednej izohelii zajmowało cały dzień pracy, a obecnie ten sam efekt można uzyskać jednym kliknięciem myszy...

Zdjęcia z ograniczeniem liczby tonów, czyli dawne izohelie, dają bardzo efektowne wyniki, zwłaszcza przy silnym ograniczeniu liczby odcieni (il. 4.26). Ze względu na polskie korzenie tego przekształcenia poświęciliśmy mu więcej miejsca, choć jest



a



b



c

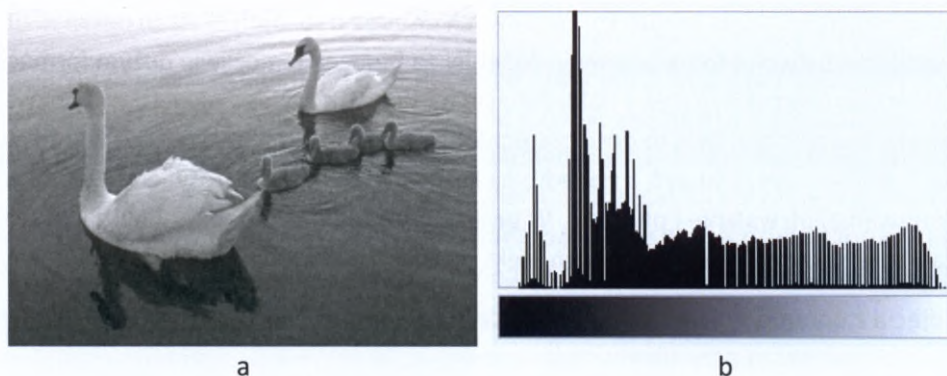


d

Il. 4.26. Przykłady obrazów z silnie zredukowaną liczbą odcieni szarości:

a) 4 odcienie, b) 6 odcieni, c) 4 odcienie, d) 5 odcieni

wykorzystywane przede wszystkim w celach artystycznych. Na potrzeby analizy obrazu można je wykorzystać do ograniczenia ilości szczegółów widocznych w obrazie, co może pomóc w detekcji obiektów. Obrazy z ograniczoną liczbą przejść tonalnych są też wykorzystywane do przedstawienia na jednym obrazie kilku rodzin obiektów. Wyglądają – co prawda – jak izohelie, ale powstają w zupełnie inny sposób.



Il. 4.27. Wykorzystanie LUT: a) obraz zoptymalizowany pod kątem odbioru przez człowieka, b) histogram tego obrazu

Il. 4.27a przedstawia obraz z il. 4.12a, przetworzony tak, aby człowiek mógł na nim zobaczyć możliwie dużo szczegółów, a jednocześnie oddać nastrój stworzony przez łabędzie z młodymi, płynące po spokojnej wodzie na krótko przed zachodem słońca. Histogram tego obrazu (il. 4.27b) pokazuje, że wykorzystano niemal cały zakres poziomów szarości z wyjątkiem niewielkiej liczby tonów bliskich czerni. Tabela LUT została stworzona na podstawie graficznie modelowanej funkcji o przebiegu pokazanym na il. 4.23c.



Il. 4.28. Schemat S-kształtnej funkcji używanej do wstępnej korekty obrazów. Opis w tekście

Wykres funkcji pokazany na il. 4.23c ma dość nieregularny przebieg, który wynika z potrzeby dopasowania się do konkretnego obrazu. Jeżeli poszukujemy możliwie uniwersalnej metody korekcji obrazów na potrzeby przeglądania ich przez człowieka, to zazwyczaj dobre wyniki daje wykorzystanie S-kształtnej krzywej, pokazanej schematycznie na il. 4.28. Dokładny kształt należy dopasować do konkretnego obrazu. Można jednak wskazać na pewne cechy ogólne tego przekształcenia:

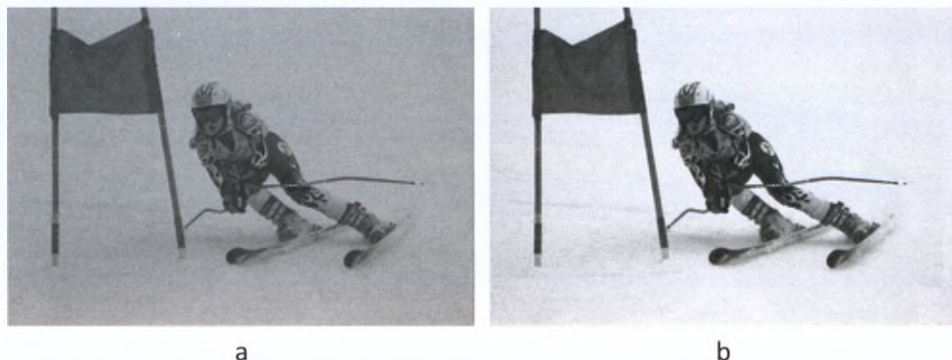
W zakresie najciemniejszych tonów mamy przekształcanie przypominające gamma modulację z $\gamma < 1$. Przyciemniamy obszary o potencjalnie najgorszej jakości spowodowanej obecnością szumów. Dodatkowo przybliżamy się do nieliniowej charakterystyki naszego zmysłu wzroku.

Środkowa, w przybliżeniu liniowa, część zwiększa kontrast obrazu, przez co wydaje się on ostrzejszy.

Wygięcie w obszarze najjaśniejszych tonów kompensuje lekki efekt niedoświetlenia, który stosuje się w celu uniknięcia utraty szczegółów w najjaśniejszych partiach obrazu.

Jak widać, mamy w jednym przekształceniu całkiem dobry „kombajn” do poprawy wyglądu naszych zdjęć, który przybliży cyfrowy oryginał do zdjęcia wykonanego za pomocą klasycznej fotografii analogowej. Efekt takiego przekształcenia obrazu pokazuje il. 4.29. Jest to kolejna metoda korekcji, która ma znaczenie przede wszystkim do celów artystycznych. W przypadku komputerowej analizy obrazu może być wykorzystana do korekcji obrazów przeznaczonych do interaktywnej analizy.

Przechodzimy teraz do operacji arytmetycznych, które potrafią zaskoczyć, pomimo swojej pozornej prostoty. Zaczynamy od negatywu, który powstaje w wyniku pomnożenia wartości wszystkich pikseli przez (-1) i następnie dodania liczby 255 (il. 4.30a).



Il. 4.29. Wykorzystanie S-kształtnej funkcji do korekty obrazu: a) obraz początkowy, b) obraz po korekcie

Histogram negatywu (il. 4.30b) jest lustrzanym odbiciem histogramu wyjściowego obrazu. Negatyw ułatwia człowiekowi dostrzeżenie niektórych szczegółów, które były niewidoczne w wersji pozytywowej. W analizie obrazu bywa wykorzystywany np. w korekcji cienia. Najczęściej negatyw obrazu jest używany podczas przekształcania obrazów binarnych i wtedy jest tożsamy z operacją logiczną NOT. Pozwala w prosty



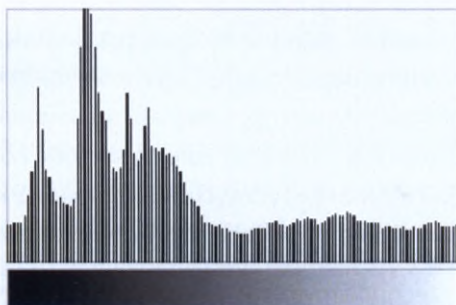
a



b



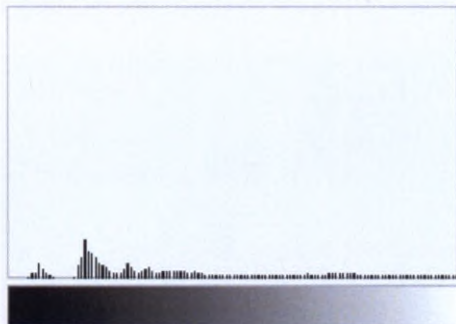
c



d



e

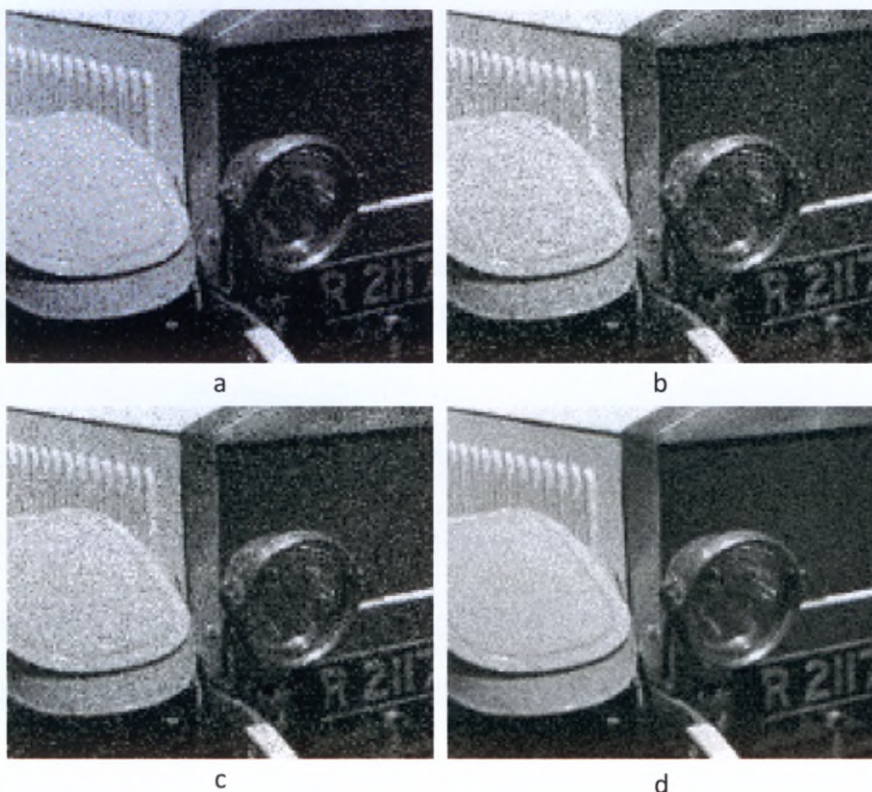


f

Il. 4.30. Obrazy po prostych przekształceniach arytmetycznych obrazu z il. 4.12a oraz ich histogramy: a-b) negatyw, c-d) dodawanie obrazu do samego siebie z ograniczeniem wyniku do 8 bitów, e-f) analogiczna operacja w programie Corel Photo-Paint. Opis w tekście

sposób zamienić np. obraz obiektów na obraz ich tła. Warto zwrócić uwagę, że wykonanie negatywu jest jedyną operacją, która na pewno nie zmienia zawartości obrazu. Powtórne wykonanie negatywu przywraca w niezmienionej formie obraz wyjściowy.

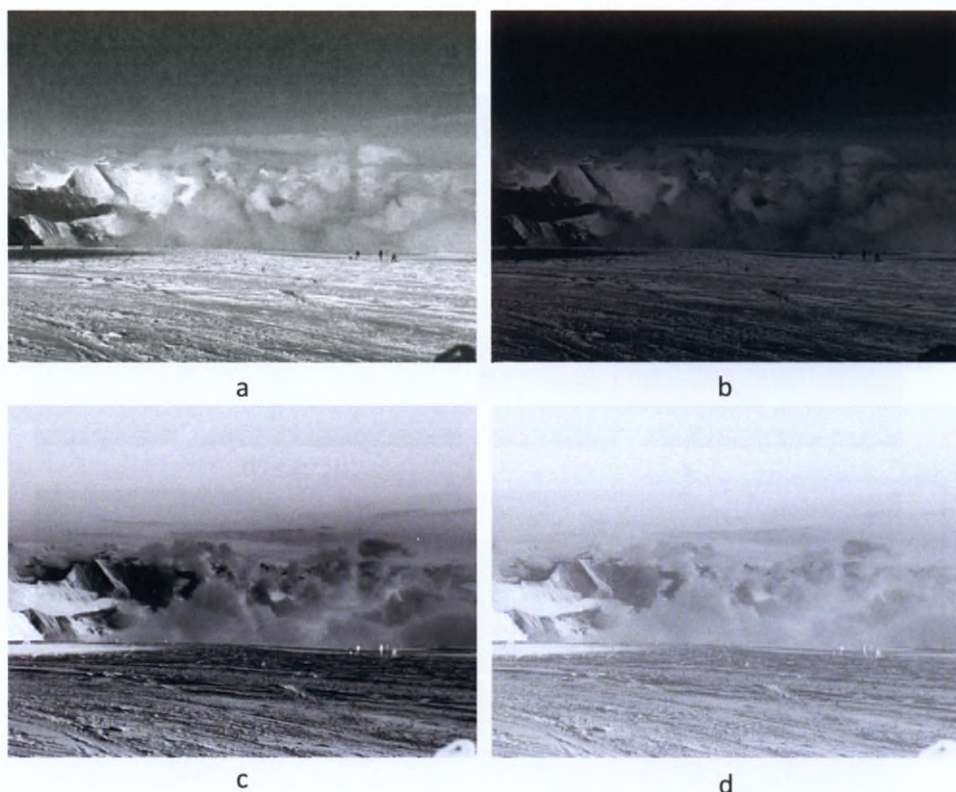
Wydaje się, że trudno o prostszą operację niż dodawanie dwóch obrazów. Trzeba tylko zadbać o zgodność formatów liczb opisujących piksele. Nie można bowiem sumować np. liczb zespolonych i całkowitych. Możemy jednak otrzymać zaskakujące wyniki z powodu ograniczeń głębi bitowej poszczególnych obrazów. Jeżeli dodajemy dwa obrazy 8-bitowe z pikselami o wartościach powyżej 127, to wynik dodawania może wykroczyć poza zakres 0-255. Reakcja oprogramowania może być różna. Profesjonalne systemy analizy obrazu podzielą sumę przez 255 i jako wynik dodawania zapiszą resztę z tego dzielenia. W konsekwencji w najjaśniejszych miejscach sumy obrazów możemy zobaczyć ciemne plamy (Il. 4.30c), a w histogramie podniosą się słupki odpowiadające najniższym wartościom (il. 4.30d). Inaczej problem rozwiązują niektóre programy graficzne, w których wartości przekraczające maksimum, przewidziane



Il. 4.31. Sprytnie ograniczenie poziomu szumów: a-c) obrazy tego samego obiektu z przypadkowym szumem, d) średnia arytmetyczna 4 obrazów z szumem

przez format graficzny, przyjmują wartość 255, podobnie jak to się dzieło np. podczas rozjaśniania obrazu. W konsekwencji otrzymujemy obraz silnie rozjaśniony i pozbawiony szczegółów w najjaśniejszych miejscach (il. 4.30e-f). Aby uniknąć niespodzianek, należy w systemach analizy obrazu ustalać większą, np. 16-bitową głębokość bitową w obrazach, które mają zawierać wynik dodawania obrazów. Oprócz tego, w przypadku operacji (nie tylko dodawania) na dwóch obrazach, konieczne jest, aby miały one taki sam wymiar w pikselach.

Interesujący wynik dodawania obrazów pokazuje il. 4.31. Po zsumowaniu 4 obrazów z szumem (trzy z nich pokazano na il. 4.31a-c) i podzieleniu wyniku przez 4 otrzymujemy końcowy obraz, na którym poziom szumu jest zdecydowanie mniejszy (il. 4.31d). Dzieje się tak z powodu przypadkowego charakteru szumu. Wykorzystuje się tę cechę szumu podczas fotografowania słabo oświetlonych obiektów, np. w nocy. Ustala się



Il. 4.32. Odejmowanie obrazów: a) obraz początkowy, b) obraz początkowy z zakresem zmian poziomów szarości dwukrotnie ograniczonym, c) różnica obrazów (b) i (a) w przypadku ustalenia, że wynik jest obrazem 16-bitowym z liczbami ujemnymi, na potrzeby wizualizacji początkowy zakres wartości od -127 do 0 został przeliczony na zakres 0-255, d) różnica obrazów (b) i (a) przy założeniu, że wynik jest obrazem 8-bitowym

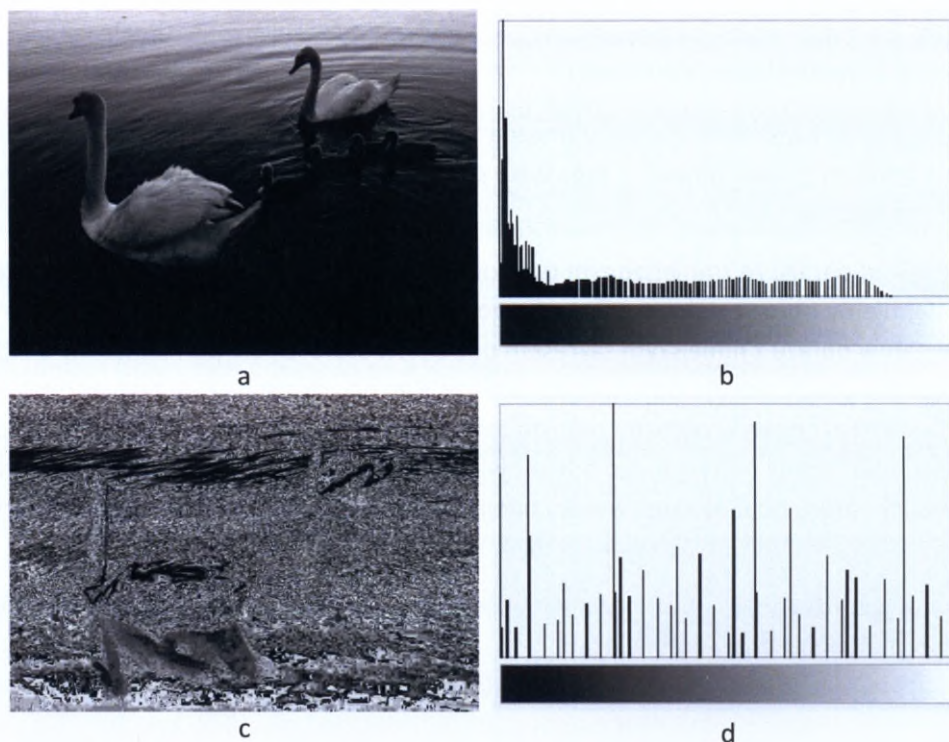
niski poziom czułości, a czasy naświetlania wydłuża do kilku sekund. Wynikowy obraz jest praktycznie wolny od szumów.

Skoro możemy dodawać obrazy (można też oczywiście dodać do obrazu liczbę), to możemy też je odejmować. Trzeba tylko pamiętać, że wynik odejmowania może być liczbą ujemną.

Jeżeli od obrazu ze zmniejszonym o połowę zakresem poziomów szarości (il. 4.32b) odejmiemy obraz z pełnym zakresem poziomów szarości (il. 4.32a), to otrzymamy negatyw obrazu z mniejszym zakresem poziomów szarości, ale pomniejszony o 128, czyli w wartościach ujemnych. Przeskalowanie go do zakresu 0-255, czyli normalizacja, da nam negatyw obrazu z pełnym zakresem poziomów szarości (il. 4.32c). Jeżeli natomiast wynik opisywanego odejmowania będzie obrazem 8-bitowym, to system analizy obrazu do ujemnego wyniku odejmowania będzie dodawał liczbę 255. Otrzymamy zatem rozjaśniony negatyw (il. 4.32d).

Kolejnym działaniem, które omówimy, jest mnożenie. Można mnożyć obraz przez obraz lub obraz przez liczbę. Trzeba tylko pamiętać, żeby wynik mnożenia zapisywać w postaci liczby co najmniej 16-bitowej (il. 4.33a). Oczywiście, mnożenie obrazu samego przez siebie jest w istocie podnoszeniem wartości pikseli do kwadratu i dlatego rozszerza się zakres tonów jasnych, a zacieśnia zakres tonów ciemnych. Przypomina to nieco gamma modulację z wartością $\gamma < 1$. Histogramy na il. 4.21d i 4.32b mają zbliżony kształt, ale trzeba pamiętać, że zakres wartości po mnożeniu jest aż 255 razy większy. W przypadku zapisywania wyników mnożenia w obrazie 8-bitowym wszystkie wyniki powyżej 255 są dzielone przez 255 i jako wynik podawana jest reszta z tego dzielenia (il. 4.33c). Histogram tego obrazu potwierdza, że 8-bitowy wynik mnożenia wygląda jak obraz szumu. Omawianego tutaj problemu nie obserwujemy w oprogramowaniu do przetwarzania obrazu do celów artystycznych, gdyż wynik mnożenia jest dzielony przez 255 przed zapisaniem nowego obrazu. Dzięki temu wynik mnożenia dwóch obrazów 8-bitowych pozostaje obrazem 8-bitowym. Do celów artystycznych takie rozwiązanie jest prawidłowe, ale w systemach analizy obrazu zależy nam na przetwarzaniu informacji, a nie na efektach wizualnych. Dlatego należy pamiętać o zapewnieniu odpowiedniej głębi bitowej obrazu dla wyników mnożenia.

W dotychczasowych przykładach oba dodawane lub mnożone obrazy były w istocie tym samym obrazem. Na il. 4.34 pokazano wynik mnożenia różnych obrazów. Ciemny obraz sylwetki wiatraka na tle chmur (il. 4.34a) pomnożono przez obraz lodowca i panoramy gór z il. 4.32a. Jeżeli wynik zapiszemy w obrazie 8-bitowym, to końcowy obraz (il. 4.34b) będzie zawierał przede wszystkim szum, podobnie jak il. 4.33c.

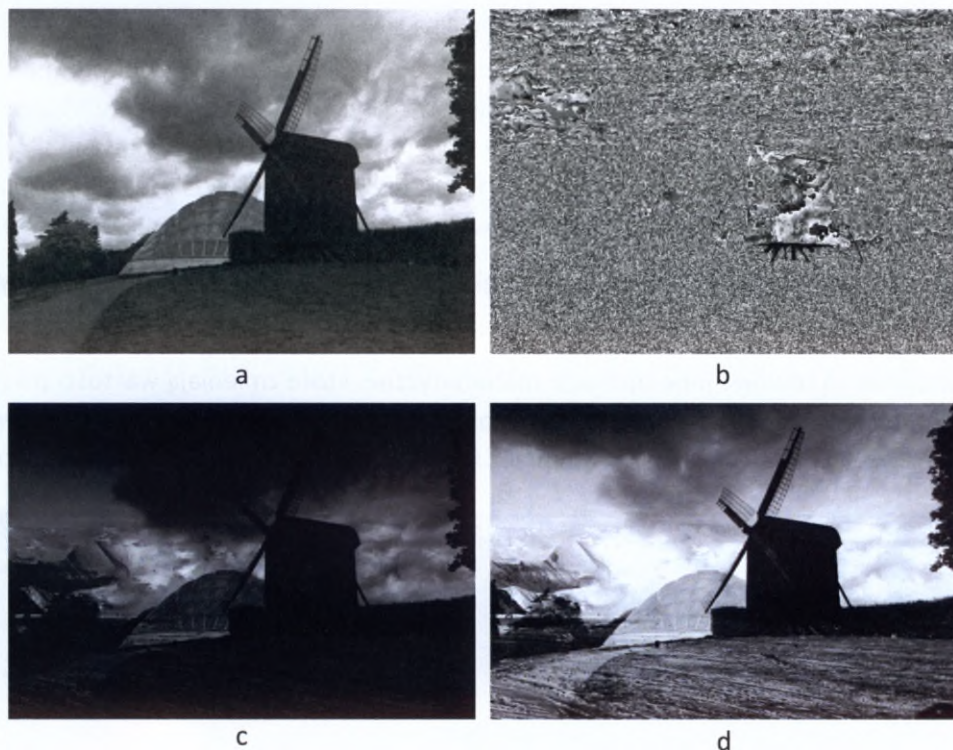


Il. 4.33. Mnożenie obrazów: a) 16-bitowy wynik mnożenia obrazu z il. 4.12a przez samego siebie, b) histogram obrazu (a), c) 8-bitowy wynik mnożenia tych samych obrazów, d) histogram obrazu (c)

Zapis wyniku w obrazie 16-bitowym daje, zgodnie z oczekiwaniami, bardzo ciemny obraz (il. 4.34c). Jeżeli jednak wykonamy jeszcze wyrównanie histogramu, to uzyskamy dość ciekawy efekt wizualny (il. 4.34d).

Mnożenie obrazów daje nieprzewidywalne wyniki, jeżeli mnożymy przez siebie przypadkowe obrazy. Może być jednak skutecznym narzędziem, gdy odpowiednio przygotuje się obraz. Dobry przykład zostanie omówiony w rozdziale poświęconym korekcji cienia. Warto też pamiętać, że np. normalizacja to w istocie złożenie odejmowania liczby od obrazu i mnożenia go przez liczbę.

Dzielenie jest związane z mnożeniem, podobnie jak dodawanie z odejmowaniem. Jednak w przypadku dzielenia zapisywanie wyniku wymaga jeszcze innego rodzaju obrazów. Wystarczy zakres od 0 do 255, ale muszą to być obrazy, w których piksele przyjmują wartości zmiennoprzecinkowe. Jednak nawet i wtedy wynik dzielenia obrazów będzie zazwyczaj trudny do interpretacji. Załóżmy, że mamy dwa obrazy, zawierające piksele o równomiernym rozkładzie w zakresie od 0 do 255. Dodatkowo



Il. 4.34. Mnożenie obrazów: a) jeden z mnożonych obrazów, b) 8-bitowy wynik mnożenia obrazu (a) i obrazu z il. 4.32a, c) 16-bitowy wynik mnożenia obrazu (a) i obrazu z il. 4.32a, d) wyrównanie histogramu obrazu (c)

zakładamy, że rozmieszczenie pikseli o konkretnych wartościach jest w obu obrazach przypadkowe. Jeżeli podzielimy przez siebie te obrazy, to wynikowy obraz będzie miał piksele w zakresie od 0 do 255, ale aż połowa z nich obejmie zakres zaledwie od 0 do 1. Tylko połowa pikseli będzie miała wartości z zakresu 1 do 255.

Obrazy zamieszczone na il. 4.35a i 4.35b pokazują, że również w przypadku obrazów dzielenie nie jest przemienne. Ze względu na bardzo dużą niejednorodność obrazów po dzieleniu, w celu poprawy czytelności poddano je normalizacji, a następnie wyrównaniu histogramu. Z kolei il. 4.35c i 4.35d potwierdzają, że bardzo duża część informacji w obrazach po dzieleniu znajduje się w zakresie wartości pikseli od 0 do 1. W rozdziale poświęconym korekcji cienia zostanie pokazana metoda wykorzystująca dzielenie obrazów, która wśród kilku różnych metod okaże się najskuteczniejsza.

Podsumowując, operacje arytmetyczne, które pozornie wydają się być bardzo proste i posiadać bardzo ograniczony zakres stosowania w analizie obrazu, mogą dać nam wyniki odpowiadające m.in. efektom takich operacji jak:

- normalizacja,
- wykonanie negatywu,
- rozjaśnienie lub przyciemnienie,
- zmiana kontrastu,
- ograniczenie liczby poziomów szarości (izohelia),
- zmiana zakresu poziomów szarości.

Jeżeli w ten sposób spojrzymy na te przekształcenia, to okażą się one całkiem silnym narzędziem przetwarzania obrazów.

Możliwe są również inne operacje matematyczne, które zmieniają wartości pikseli zgodnie z różnymi funkcjami, np. trygonometrycznymi, logarytmicznymi itp. Ciekawe możliwości daje mnożenie obrazu przez macierz współczynników. Jest to przydatne



a



b



c



d

Il. 4.35. Dzielenie obrazów: a) wynik dzielenia obrazu z il. 4.34a przez obraz z il. 4.32a, poddany normalizacji i wyrównaniu histogramu, b) wynik dzielenia obrazu z il. 4.32a przez obraz z il. 4.34a, poddany normalizacji i wyrównaniu histogramu, c) zakres od 0 do 1 wyniku dzielenia obrazu z il. 4.34a przez obraz z il. 4.32a, poddany normalizacji, d) zakres od 0 do 1 wyniku dzielenia obrazu z il. 4.32a przez obraz z il. 4.34a, poddany normalizacji

w przypadku obrazów posiadających kilka składowych, jak np. obrazy kolorowe. Kolumny macierzy odpowiadają za składowe obrazu początkowego, a wiersze za składowe obrazu wynikowego (il. 4.36a-c). Jeżeli zastosujemy macierz z il. 4.36b, to uzyskamy cykliczną zmianę składowych obrazu RGB: składowa czerwona stanie się zieloną, zielona stanie się niebieską, a niebieska stanie się czerwoną (il. 4.36e i 4.36h). Z kolei macierz z il. 4.36c powoduje zamianę składowej niebieskiej z czerwoną (il. 4.36f i 4.36i).

W przypadku obrazów na il. 4.36g oraz 4.36i nie obserwujemy drastycznych zmian w kolorach. Wynika to stąd, że składowa zielona, która dominuje w liściach, nie uległa zmianie, a za fioletowy kolor kwiatów odpowiadają w zbliżonych proporcjach

	B	G	R
B	1	0	0
G	0	1	0
R	0	0	1

a

	B	G	R
B	0	0	1
G	1	0	0
R	0	1	0

b

	B	G	R
B	0	0	1
G	0	1	0
R	1	0	0

c



d



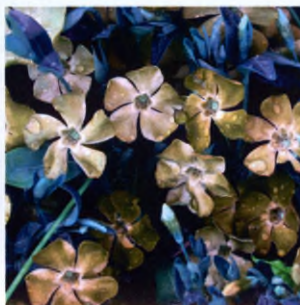
e



f



g



h



i

Il. 4.36. Mnożenie obrazu RGB przez macierz: a) macierz niezmieniająca kolorów, b) macierz cyklicznej zamiany: R na G, B na B i B na R, c) macierz zamiany R na G i G na R, d) model RGB, e-f) model RGB przekształcany zgodnie z macierzami (b) i (c), h) wyjściowy obraz kwiatów, h-i) obraz kwiatów przekształcany zgodnie z macierzami (b) i (c)



	B	G	R
B	0	0	0,5
G	0,5	1	0
R	1	0	0



Il. 4.37. Od lewej: obraz początkowy, macierz mnożenia, obraz wynikowy. Opis w tekście

składowe czerwona i niebieska. Można oczywiście wykorzystać bardziej skomplikowane macierze. Dodając do składowej niebieskiej 50% zielonej i osłabiając czerwoną do 50%, otrzymamy kolory pokazane na il. 4.37.

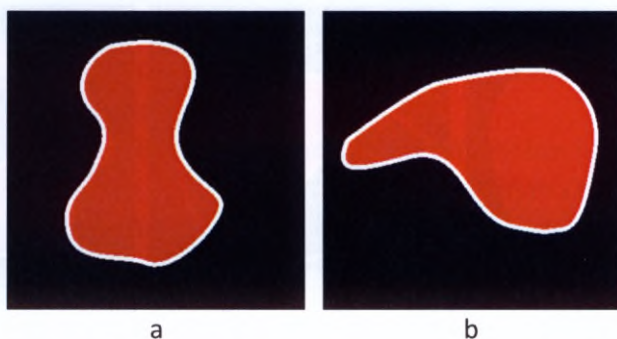
Omawiane wcześniej operacje arytmetyczne są oczywiście również operacjami matematycznymi. Główna różnica polega na tym, że w operacjach arytmetycznych zwykle ograniczamy się do podstawowych działań na liczbach. Należy jednak pamiętać, że ten podział jest umowny i służy przede wszystkim uporządkowaniu przekształceń.

Dotychczasowe rozważania dotyczyły przede wszystkim obrazów szarych lub, rzadziej, kolorowych. Zajmiemy się teraz operacjami logicznymi, które są wykonywane głównie na obrazach binarnych. Ich stosowanie w stosunku do innych rodzajów obrazów jest wprawdzie możliwe, ale zdarza się to bardzo rzadko.

Tabela 4.1. Zestawienie wyników operacji logicznych

wartości pikseli obrazów A i B oraz wyniki operacji logicznych											
obraz A	obraz B	A OR B	A AND B	A XOR B	A NXOR B	A NAND B	A NOR B	A – B	B – A	NOT (A – B)	NOT (B – A)
0	0	0	0	0	1	1	1	0	0	1	1
0	1	1	0	1	0	1	0	0	1	1	0
1	0	1	0	1	0	1	0	1	0	0	1
1	1	1	1	0	1	0	0	0	0	1	1

Możliwości operacji logicznych przedstawia tabela 4.1. W czterech wierszach tabeli znajdują się wyniki operacji logicznych dla czterech możliwych kombinacji pikseli z obrazów wyjściowych: 00, 01, 10 i 11. Na podstawie danych zawartych w tabeli 4.1



Il. 4.38. Obrazy binarne użyte do zilustrowania operacji logicznych: a) obraz A, b) obraz B.
Opis w tekście

trudno jest wyobrazić sobie, jaki efekt daje dane przekształcenie. Dlatego wyniki poszczególnych operacji zostały pokazane również w sposób graficzny. Punktem wyjścia są dwa obrazy binarne, oznaczone literami A i B (Il. 4.38). Są one wyświetlane w konwencji przyjętej w systemach analizy obrazu: piksele o wartości 0 są wyświetlane jako czarne, a piksele o wartości 1 wyświetlane są jako czerwone. Dodatkowo, w celu zwiększenia przejrzystości ilustracji zawsze rysowany jest białą linią kontur obiektu.

Pomimo wielkiej prostoty obrazów binarnych, które zawierają tylko dwie wartości, operacje logiczne oferują aż 12 różnych wyników, które pokazuje il. 4.39. Są to:

- OR, czyli suma logiczna (il. 4.39a). Przypisuje wartość 1 tym pikselom wynikowego obrazu, które miały wartość 1 w obrazie A lub w obrazie B. Pozostałe punkty pozostają z wartością 0.
- AND, czyli iloczyn logiczny (il. 4.39b). Przypisuje wartość 1 tym pikselom wynikowego obrazu, które miały wartość 1 w obu obrazach A i B. Pozostałe punkty mają wartość 0, nawet jeśli w jednym z obrazów miały wartość 1.
- XOR, czyli suma rozłączna (ang. *eXclusive OR*, il. 4.39c). Przypisuje wartość 1 tym pikselom wynikowego obrazu, które miały różne wartości w obrazie A i w obrazie B. Pozostałe punkty pozostają z wartością 0.
- NOR (il. 4.39d). Negatyw operacji OR. Przypisuje wartość 1 tym pikselom wynikowego obrazu, które miały w obu obrazach wartości 0. Pozostałe punkty pozostają z wartością 0.
- NAND (il. 4.39e). Negatyw operacji AND. Przypisuje wartość 1 tym pikselom wynikowego obrazu, które miały przynajmniej w jednym obrazie wartość 0. Pozostałe punkty pozostają z wartością 0.
- NXOR (il. 4.39f). Negatyw operacji XOR. Przypisuje wartość 1 tym pikselom wynikowego obrazu, które miały w obu obrazach te same wartości, nieważne, czy było to 1, czy też 0. Pozostałe punkty pozostają z wartością 0.



a: $A \text{ OR } B$



b: $A \text{ AND } B$



c: $A \text{ XOR } B$



d: $A \text{ NOR } B$



e: $A \text{ NAND } B$



f: $A \text{ NXOR } B$



g: $\text{NOT } A$



h: $A - B$



i: $B - A$



j: $\text{NOT } B$



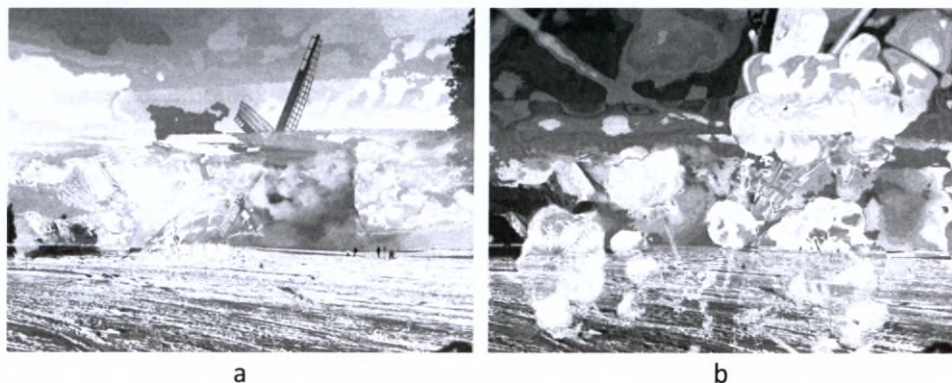
k: $\text{NOT}(A - B)$



l: $\text{NOT}(B - A)$

- NOT (il. 4.39g oraz 4.39j). Jest to po prostu negatyw obrazu binarnego. Piksele z początkową wartością 1 otrzymują wartość 0, a w pozostałych przypadkach działanie jest przeciwne.
- $A - B$ lub $B - A$ (il. 4.39h oraz 4.39i). W pierwszym obrazie binarnym zamienia się wartości pikseli z 1 na 0 w tych miejscach, w którym drugi obraz również wykazuje wartość 1. Można też powiedzieć, że jest to usunięcie z obrazu części wspólnej z drugim obrazem.
- $\text{NOT}(A - B)$ lub $\text{NOT}(B - A)$ (il. 4.39k i 4.39l). Są to negatywy obrazów 4.39h oraz 4.39i.

Liczba możliwych operacji logicznych dla dwóch obrazów binarnych (il. 4.39) jest zaskakująco duża. Wykorzystuje się je do modyfikacji wykrytych obiektów, zapisanych w postaci obrazów binarnych. Część obrazów może zawierać jedną rodzinę obiektów, część inne obiekty, a kolejny obraz linie (będzie o nich mowa przy okazji operacji morfologicznych), które rozdzielają sklejone obiekty. Dzięki operacjom logicznym możemy uzyskać poprawny wynik końcowy. Czytelnika zapraszam do uruchomienia swojej wyobraźni w celu znalezienia dobrego zastosowania różnych operacji logicznych.



Il. 4.40. Suma logiczna (OR) obrazów szarych: a) obrazów z il. 4.32a i 4.34a, b) obrazów z il. 4.32a i 4.41a

W przypadku obrazów szarych operacje logiczne przeprowadza się na podstawie zapisu w układzie dwójkowym dla każdego bitu oddzielnie. Na przykład operacja OR na pikselach o wartości 00110111 (czyli 55) oraz 100001100 (czyli 140) daje w wyniku 10111111 (czyli 191). Taki wynik jest oczywiście jednoznaczny, ale praktycznie niemożliwy do przewidzenia na podstawie obserwacji wyjściowych obrazów. Zaskakujące wyniki operacji logicznych na obrazach szarych potwierdzają przykłady pokazane na il. 4.40. Z tego właśnie powodu operacje logiczne na obrazach innych niż binarne są przeprowadzane bardzo rzadko.

Fakt używania układu dwójkowego do zapisu liczb w systemach komputerowych można łatwo wykorzystać. Poszczególne pozycje cyfr liczby zapisanej w układzie dwójkowym odpowiadają kolejnym bitom obrazu, co umożliwia proste operacje na poszczególnych bitach. Przykładem może być tzw. przesuwanie bitów w lewo i w prawo. Jeżeli liczbę 157 zapiszemy w układzie dwójkowym, to dostaniemy:

10011101.

Jeżeli usuniemy trzy ostatnie cyfry, czyli przesuniemy liczbę o trzy bity w prawo, to otrzymujemy:

10011,

co w układzie dziesiętnym zapisujemy jako 19. Jest to wynik dzielenia bez reszty liczby 157 przez 8. Jeżeli teraz naszą liczbę przesuniemy o trzy bity w lewo, to otrzymamy:

10011000,

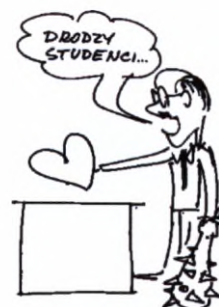
co odpowiada 152 w układzie dziesiętnym, czyli iloczynowi 19 i 8. Podsumowując, za pomocą przesunięć bitów o jedną pozycję w prawo lub w lewo wykonujemy w istocie dzielenie lub mnożenie przez 2. Operacje na pojedynczych bitach można wykorzystać np. do szybkiej redukcji szumów w obrazach.

Na zakończenie przedstawimy trochę nietypowe zastosowanie omawianych przekształceń. Obraz kwiatów na il. 4.41a nie zdradza niczego szczególnego. Okazuje się jednak, że ostatnie bity kodują binarną wersję pewnego rysunku (il. 4.41b). Można go odzyskać na il. 4.41a w dość prosty sposób:

- dzielimy, a następnie mnożymy obraz z il. 4.41a przez 2,
- wynik odejmujemy od początkowego obrazu,
- otrzymany wynik mnożymy przez 255.



a



b

Il. 4.41. a) pozornie nieukrywający niczego obraz, b) inny obraz, który udaje się odczytać z obrazu (a). Szczegóły w tekście



a



b

Il. 4.42. a) pozornie nieukrywający niczego obraz, b) inny, 3-bitowy obraz, który udaje się odczytać z obrazu (a). Szczegóły w tekście

Gdyśmy zapisali ten obraz jako jeden z pierwszych bitów, to oczywiście pewne jego elementy byłoby widać od razu. Jeśli jest to ostatni bit, to zmienia on wartość piksela najwyżej o 1, co jest przy odpowiednio bogatym w zróżnicowane szczegóły obrazie niewidoczne dla ludzkiego oka. Ponieważ rozróżniamy tylko kilkadziesiąt stopni szarości, to na ukryty obraz można poświęcić nawet trzy bity (il. 4.42).

Jak widać, można w obrazie ukryć pewną informację i nie będzie ona widoczna dla przypadkowego użytkownika. Może ją jednak dość łatwo odczytać osoba, która zna techniki ukrywania informacji w obrazach, czyli steganografii. Dlatego nie należy w ten sposób przysyłać wrażliwych informacji, chyba że zrobimy to w sposób jeszcze trudniejszy do odkrycia lub wiadomość dodatkowo zaszyfrujemy.

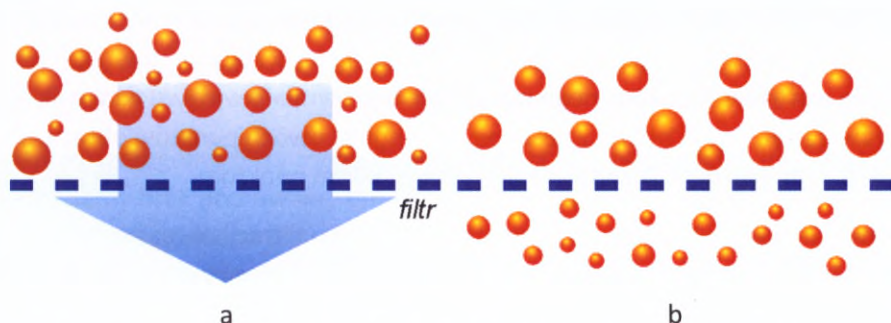
5. Filtry

Różnego rodzaju filtry są obecne w wielu otaczających nas urządzeniach. Spotykamy je w ekspresach do kawy, pralkach, odkurzaczach, układach uzdatniania wody, klimatyzacji itp. Prostym filtrem jest też każde sito czy też cedzak, „oddzielający” makaron od wody, w której go gotowano. Wiele filtrów pracuje w samochodach. Są to filtry powietrza, oleju, paliwa czy też przeciwpyłkowe. Wspólną cechą tych filtrów jest to, że oddzielają filtrowane elementy przede wszystkim w sposób mechaniczny. Inaczej mówiąc, umożliwiają przepływ cieczy lub gazu, ale ze względu na rozmiary otworów lub kanałów uniemożliwiają transport cząstek większych od rozmiarów tych mechanicznych przeszkód. Schemat działania takich filtrów ukazuje il. 5.1.

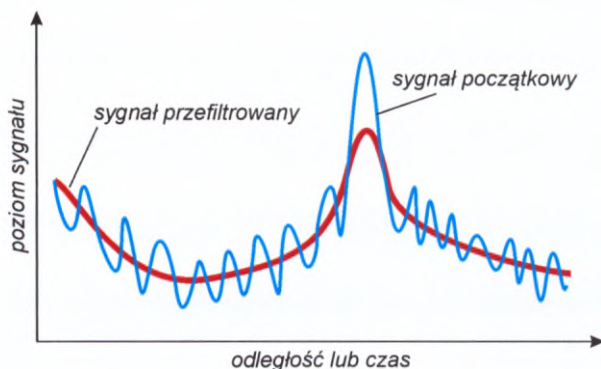
Z całej rodziny cząstek przez filtr przechodzą tylko te, których rozmiary są mniejsze od otworów w filtrze. W ten sposób możemy odseparować duże i małe cząstki albo pozbyć się niechcianych zanieczyszczeń.

Niewidoczne dla naszego oka, ale równie ważne, są filtry działające w urządzeniach elektronicznych. Służą one oczyszczaniu sygnału z różnego rodzaju szumów i zakłóceń. Dzięki tego typu filtrom możemy swobodnie rozmawiać przez telefon komórkowy, nawet stojąc w hałaśliwym miejscu, czy też oglądać dobrej jakości obraz w telewizorze. W przypadku filtracji sygnałów nie mamy fizycznie istniejącego sita. Jego rolę spełniają albo odpowiednie układy elektroniczne, albo oprogramowanie (il. 5.2).

Procesu filtracji używamy w analizie obrazu do wyodrębnienia potrzebnej nam informacji. W praktyce taką filtrację wykonuje nasz zmysł wzroku, gdy wypatrujemy



Il. 5.1. Schemat procesu mechanicznej filtracji: a) cząsteczki przed lub nad filtrem, b) rozdzielenie cząstek na dwie frakcje



Il. 5.2. Idea procesu filtracji sygnału

w tłumie przechodniów osoby, na którą właśnie czekamy. Jeżeli wiemy, że ta osoba będzie ubrana w czerwoną kurtkę, to nasz zmysł wzroku będzie skupiał się przede wszystkim na czerwonych elementach. Dzięki odpowiednim algorytmom filtracji poddajemy też obrazy analizowane komputerowo.

W poprzednim rozdziale omówiliśmy wiele operacji, które istotnie wpływają na obraz. Możemy zmieniać jego kolor, jasność, kontrast, zamieniać obrazy kolorowe na szare lub binarne, zmieniać liczbę poziomów szarości, wykonywać operacje logiczne itd. Operacje te mają jedną wspólną cechę: ich wynik zależy tylko od wartości konkretnego, przetwarzanego punktu. Potocznie można powiedzieć, że są to funkcje jednoargumentowe. Określenie to jest jednak nie do końca słuszne, bo np. w przypadku obrazów kolorowych jeden punkt jest opisywany za pomocą trzech (np. model RGB), czterech (np. model CMYK) lub nawet większej liczby zmiennych. Ujmując rzecz ściśle, są to operacje punktowe (wynik dla jednego punktu zależy od funkcji i jednego punktu). Nawet jeżeli w tych operacjach używaliśmy dwóch obrazów, to wynik zależał tylko od jednego punktu w każdym obrazie.

W przeciwieństwie do operacji punktowych filtry są operacjami kontekstowymi, czyli wynik filtracji dla jakiegoś punktu zależy od większej liczby punktów. Może się zdarzyć, że wartość przetwarzanego punktu nie będzie miała wpływu na wynik filtracji, ale zawsze ten wynik będzie zależał od co najmniej dwóch punktów obrazu, a ściślej mówiąc – od wartości liczbowych przypisanych tym punktom.

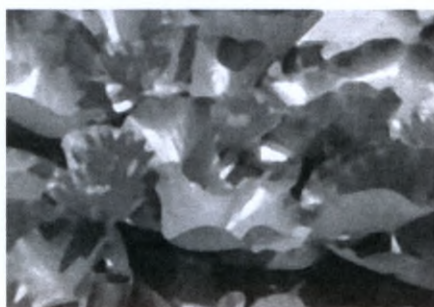
Oczywiście nie ma jednego, uniwersalnego sposobu filtracji obrazów. Kilka przykładów działania różnego rodzaju filtrów pokazuje il. 5.3 (obraz początkowy poddawany następnie filtracji oznaczono literą a). Najogólniej rzecz biorąc, wynik działania filtru jest po prostu wynikiem pewnych działań algebraicznych na liczbach opisujących kolor lub stopień szarości punktów obrazu, leżących w sąsiedztwie analizowanego



a



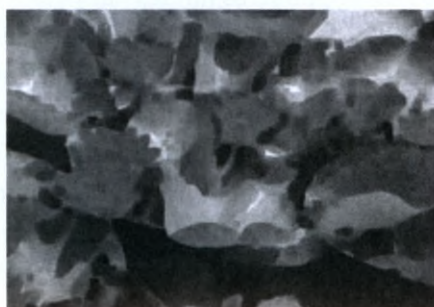
b



c



d



e

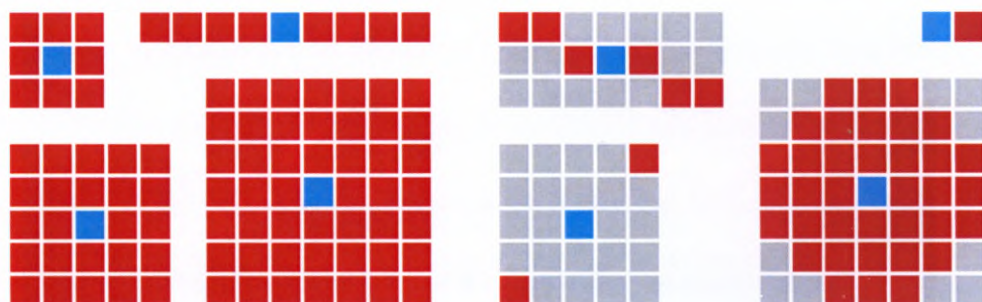


f

Il. 5.3. Efekt działania wybranych filtrów: a) obraz wyjściowy, b) obraz uśredniony, c) wynik filtracji medianowej, d) wyostrenie, e) filtr minimalny, f) filtr maksymalny.

Wymienione rodzaje filtrów zostaną opisane w dalszej części rozdziału

punktu. Poszczególne filtry mogą działać z różną siłą. Najczęściej, im większy obszar będziemy brali pod uwagę, tym silniej filtr będzie zmieniał początkowy obraz. Wszystkie przekształcenia zilustrowane na il. 5.3 zostały specjalnie wykonane z dużą siłą, aby wyraźnie pokazać wyniki ich działania. Nawet pobieżna analiza il. 5.3 wskazuje, że wyniki działania filtrów mogą być silnie zróżnicowane. Niektóre filtry, np. uśredniające (il. 5.3b), mogą doprowadzić do powstania obrazu, który ocenimy jako obraz o znacznie gorszej jakości niż obraz początkowy (podobnie jest z il. 5.3e



Il. 5.4. Różne matryce filtrów. Opis w tekście

oraz 5.3f). Czasami takie działanie jest jednak potrzebne. Odpowiednie przykłady zostaną pokazane podczas omawiania bardziej złożonych algorytmów analizy obrazu.

Czas na zdefiniowanie filtrów w komputerowej analizie obrazu. Filtrami będziemy nazywali wieloargumentowe funkcje przekształcające obraz. Wartość punktu po przekształceniu jest funkcją wartości tego punktu oraz wartości jego sąsiadów, przy czym sąsiedztwo to należy rozumieć bardzo ogólnie. Sąsiadami w sensie filtracji mogą być piksele znacznie oddalone od analizowanego piksela.

Sąsiedztwo analizowanego punktu brane pod uwagę podczas filtracji nazywamy matrycą filtra. W każdej z matryc pokazanych na il. 5.4 przetwarzany punkt jest zaznaczony na niebiesko, a pozostałe punkty brane pod uwagę podczas filtracji zaznaczono na czerwono. Matryce definiujemy w kwadratowym lub prostokątnym otoczeniu. Jeżeli matryca ma inny kształt, to części punktów nie bierzemy pod uwagę. Takie punkty zaznaczono na il. 5.4 jako jasnoszare. Jak widać na tym rysunku, punkty brane pod uwagę podczas filtracji nie muszą być najbliższymi sąsiadami. Może być ich dużo, ale to wydłuża proces filtracji. W skrajnym przypadku mogą to być dwa punkty. Najczęściej używane są matryce kwadratowe o rozmiarach 3x3, 5x5 lub 7x7, w których przetwarzany jest ich centralny punkt (co tłumaczy, dlaczego ich rozmiar wyrażany jest liczbą nieparzystą).

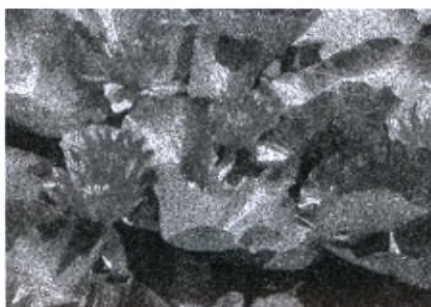
Intuicyjne rozumienie filtrów każe nam oczekiwać, że usuną one z obrazu coś niepożądanego lub poprawią jego jakość. Przykład takiego działania pokazuje il. 5.5. Z silnie zakłóconego szumem obrazu (il. 5.5b) jesteśmy w stanie uzyskać obraz bardzo dobrej jakości (il. 5.5d).

Należy podkreślić, że zastosowania filtrów w przetwarzaniu oraz analizie obrazu mogą być bardzo zróżnicowane. Podstawowe obszary wykorzystywania filtrów to:

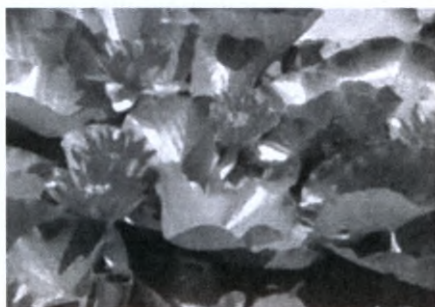
- redukcja szumów,
- wyostrażanie obrazów,



a



b



c



d

Il. 5.5. Dobrej jakości obraz (a) może być często skutecznie odtworzony na podstawie obrazu zawierającego szum (b). Szum likwiduje filtracja medianowa (c), która jednak daje obraz lekko rozmyty. Zaawansowane wyostżanie prowadzi do końcowego obrazu (d), który jest bardzo bliski idealnemu obrazowi (a)

- detekcja krawędzi,
- efekty specjalne.

Redukcja szumów jest często potrzebna, gdy obrazy są rejestrowane przy słabym poziomie sygnału. Wielu osobom wydaje się, że jest to najważniejsze zastosowanie filtrów. Pogląd ten jest jednak błędny, bo mamy wiele innych, co najmniej równie istotnych zakresów zastosowania filtrów w przetwarzaniu obrazów.

Wyostżanie obrazów jest powszechnie stosowane w urządzeniach do ich rejestracji, często bez informowania użytkownika o tym działaniu. W analizie obrazu stosuje się je bardzo rzadko, bo zazwyczaj generuje dodatkowy szum, utrudniający analizę. Wyostżanie zostanie jednak omówione, gdyż ma duże znacznie praktyczne poza obszarem analizy obrazu. Wyjaśnimy wtedy, jak powstał obraz z il. 5.5d.

Bardzo istotnym, ale jednocześnie trudnym (jak się okaże nieco później) obszarem zastosowania filtrów jest detekcja krawędzi. Ludzki zmysł wzroku doskonale wykrywa krawędzie i dzięki temu skutecznie omijamy przeszkody, chwytamy przedmioty itp.



Il. 5.6. Obiekty na znakach drogowych rozpoznajemy bez problemu, gdyż zawierają uproszczony obraz dobrze nam znanych krawędzi tych obiektów

Od niepamiętnych czasów przedstawiamy graficznie uproszczone rysunki, zawierające ogólne krawędzie obiektów, i nie mamy żadnych problemów z rozpoznaniem, o co chodziło autorowi rysunku, zawierającego w istocie tylko krawędzie. Dobrym przykładem takiego uproszczenia są znaki drogowe (il. 5.6).

Efekty specjalne obejmują bardzo rozległą grupę filtrów. Część z nich może być wykorzystana w analizie obrazu, ale większość służy do celów artystycznych. Takie filtry pozwalają zamienić zdjęcie na obraz malowany akwarelami lub rysunek wykonany ołówkiem, przetworzyć słoneczny dzień na ciemną noc itp. Tego typu filtrami jednak nie będziemy się zajmowali w ramach niniejszego podręcznika.

Często spotykanym podziałem jest podział na filtry liniowe i nieliniowe. Trzeba wyraźnie podkreślić, że filtry liniowe nie są ani gorsze, ani lepsze od innych rodzajów filtrów. Nie mają również obszaru zastosowań niedostępnego dla innych filtrów. Jedynym wyróżnikiem filtrów liniowych jest to, że można je opisać w postaci funkcji liniowej, czyli wielomianu pierwszego stopnia.

p_{11}	p_{12}	p_{13}	a_{11}	a_{12}	a_{13}
p_{21}	p_{22}	p_{23}	a_{21}	a_{22}	a_{23}
p_{31}	p_{32}	p_{33}	a_{31}	a_{32}	a_{33}

Il. 5.7. Wielkości niezbędne do filtrowania liniowego: a) wartości pikseli w obrazie, b) maczyca współczynników

Jeżeli w obrazie mamy szereg pikseli o wartościach $p_{11}, p_{12}, \dots, p_{33}$ (il. 5.7a), to możemy zdefiniować filtr liniowy F :

$$F(p_{11}, p_{12} \dots p_{33}) = a_0 + a_{11} \cdot p_{11} + a_{12} \cdot p_{12} + \dots + a_{33} \cdot p_{33} \quad (5.1)$$

gdzie: $a_0, a_{11}, a_{12}, \dots, a_{33}$ – predefiniowane stałe.

Filtr liniowy można zapisać za pomocą prostej funkcji liniowej (równanie 5.1), będącej średnią ważoną analizowanych punktów. Inaczej mówiąc, każdemu punktowi przypisuje się pewien współczynnik, a wynik działania filtru to suma iloczynów wszystkich branych pod uwagę punktów oraz odpowiadających im współczynników. Bardzo często takie filtry definiuje się, rysując tabelę wizualizującą lokalną konfigurację punktów i w każdej komórce tej tabeli wpisując wartości odpowiednich współczynników (il. 5.7b).

Takie rozwiązanie umożliwia łatwe projektowanie filtru. Budujemy odpowiednią macierz (tabelę) współczynników, a oprogramowanie wykonuje filtrację zgodnie z naszym projektem. Poniżej zamieszczono tabelę o rozmiarze 3×3 , wypełnioną współczynnikami o stałej wartości, wynoszącej $1/9$. Możemy potraktować ją jako definicję filtru, który wyznacza średnią arytmetyczną z obszaru 3×3 punkty. Wartość, będącą wynikiem działania filtru, przypisuje się środkowemu punktowi. Tak zdefiniowany filtr można, przez analogię do statystycznej analizy wyników badań, wykorzystać do redukcji szumów w obrazie. Niestety, uśrednianie rozmywa obraz (il. 5.3b). Taki filtr uśredniający w systemach analizy obrazu wykorzystujących język angielski jest nazywany *box filter*.

1/9	1/9	1/9
1/9	1/9	1/9
1/9	1/9	1/9

1/16	1/8	1/16
1/8	1/4	1/8
1/16	1/8	1/16

Filtry rozmywające obraz mogą mieć również matryce ze współczynnikami o różnej wartości, jak w tabeli powyżej (z prawej strony). Zawsze jednak ich suma wynosi 1. W pokazanym powyżej przykładzie wpływ poszczególnych pikseli na wynik filtracji zmniejsza się wraz z odległością od centralnego punktu. Wartości współczynników są często tak dobierane, aby w matrycach o różnych rozmiarach przybliżyły znaną z rachunku prawdopodobieństwa krzywą rozkładu Gaussa, nazywanego też rozkładem normalnym. Takie filtry nazywamy filtrami Gaussa.

Znacznie lepszym przekształceniem do redukcji szumu jest mediana, czyli wartość środkowa (wartości poszczególnych punktów układu się w ciąg rosnący lub niemalejący w przypadku takich samych wartości liczb i wybiera wartość zajmującą środkową pozycję). Filtr medianowy odrzuca zatem wartości skrajne, które zazwyczaj są poszukiwanym przez nas szumem. Ponadto mediana nie wprowadza, w odróżnieniu



Il. 5.8. Filtrowanie obrazów: a) zdjęcie motocyklisty, które będzie poddawane różnym filtracjom, b) działanie filtra uśredniającego (*box filter*)

od poprzedniego filtra, nowych wartości do obrazu. Dzięki tym cechom jest nie tylko skutecznym narzędziem do redukcji szumu, ale jednocześnie dość dobrze zachowuje krawędzie przetwarzanego obrazu (il. 5.3c i 5.5c oraz 5.9b), nie rozmywając go tak, jak filtr oparty na średniej arytmetycznej.

Warto zwrócić uwagę, że filtr medianowy jest filtrem nieliniowym. Kolejnymi przykładami filtrów nieliniowych mogą być filtry: minimalny oraz maksymalny (il. 5.9c i 5.9d). W przypadku filtracji minimalnej wybiera się najniższą wartość z pikseli leżących w obrębie matrycy filtra. Analogicznie, podczas filtracji maksymalnej wybiera się najwyższą wartość. Każdy z tych filtrów eliminuje z obrazu pewną liczbę drobnych szczegółów. Dodatkowo powodują odpowiednio przyciemnienie lub rozjaśnienie obrazu.

Wszystkie te filtry, czyli medianowy, minimalny i maksymalny, należą do grupy tzw. filtrów rankingowych. Ich działanie polega na tym, że wartości punktów obrazu (pikseli) sortuje się i przekształca w ciąg rosnący lub niemalejący, zaś analizowany punkt przyjmuje wartość odpowiadającą pozycji w rankingu (il. 5.10). W przypadku filtra minimalnego jest to pierwsza wartość, dla filtra medianowego wartość środkowa, a dla filtra maksymalnego ostatnia wartość z posortowanego ciągu.

Omówione dotąd filtry nieliniowe są bardzo proste, ale pozwalają na uzyskanie dość wyraźnie zróżnicowanych obrazów. Filtr minimalny powoduje zwiększenie rozmiarów ciemnych obiektów i zmniejszenie obiektów jasnych. Filtr maksymalny działa wręcz odwrotnie. Można np. powiedzieć, że filtr maksymalny to sekwencja trzech operacji: negatyw – filtr minimalny – negatyw. Jeżeli złożymy te dwa filtry, to uzyskamy obraz o podobnym rozmiarze obiektów jak w obrazie wyjściowym, ale zmodyfikowanych. Złożenie filtra maksymalnego i następnie minimalnego spowoduje



a



b



c



d

Il. 5.9. Filtrowanie obrazów, matryca 3x3: a) obraz wyjściowy, b) wynik filtracji medianowej, c) wynik działania filtra minimalnego i d) maksymalnego



a



b



c



d



e



f

Il. 5.10. Filtr rankingowy. W rzędach od góry od lewej: obraz wyjściowy oraz wynik filtracji rankingowej o rozmiarze 5x5. Odpowiednio pozycje w rankingu: 1, 7, 13, 19 oraz 25



Il. 5.11. Filtrowanie obrazów: a) złożenie filtra maksymalnego i minimalnego (zamknięcie), b) złożenie filtra minimalnego i maksymalnego (otwarcie). Matryca filtru: 5x5

jak gdyby zalanie drobnych ciemnych elementów. Takie złożenie filtrów nazywamy zamknięciem (il. 5.11a).

Z kolei złożenie filtra minimalnego i następnie maksymalnego spowoduje usunięcie drobnych jasnych elementów. Obraz sprawia wrażenie lekko przyciemnionego. Takie złożenie filtrów nazywamy otwarciem (il. 5.11b). W tym momencie nazwy otwarcie i zamknięcie są zupełnie niezrozumiałe. Zostaną one bliżej wyjaśnione podczas analizy tzw. przekształceń morfologicznych. Filtry minimalny i maksymalny są tożsame z dwoma przekształceniami morfologicznymi: erozją i dylatacją. Nie będziemy jednak w tym momencie bliżej rozwijać tego zagadnienia.

Filtry minimalny i maksymalny są powszechnie używane do wygładzania konturów, detekcji lokalnych ekstremów w obrazie oraz do eliminacji części szumów. Oba filtry rzadko są używane samodzielnie, natomiast bardzo często stanowią istotny krok w procesie przetwarzania obrazu.

Przejdziemy teraz do bardzo ważnego zagadnienia, jakim jest wspomniana już detekcja krawędzi. Nasz zmysł wzroku radzi sobie doskonale z tym problemem. Dzięki temu, przechodząc przez nieznane nam wcześniej pomieszczenia nie obijamy się o znajdujące się tam osoby, meble czy też maszyny. Jeżeli zastanowimy się nad treścią prostych, odręcznych rysunków (il. 5.12a), to bez trudu rozpoznamy drzewo iglaste i liściaste, księżyc na krótko przed nowiem i słońce, a także głowę tygo człowieka noszącego okulary. Obraz krawędzi dał nam całkiem sporą porcję informacji, w dodatku zrozumiałą nawet dla osób komunikujących się w całkowicie różnych językach.

Podobną rolę pełnią emotikony, piktogramy oraz ikony. Symbole pokazane na il. 5.12.b odczytujemy bez problemu jako możliwość skorzystania z telefonu,



Il. 5.12. Wykorzystanie obrazu krawędzi do przekazywania informacji: a) rysunki odręczne, b) piktogramy oraz ikony



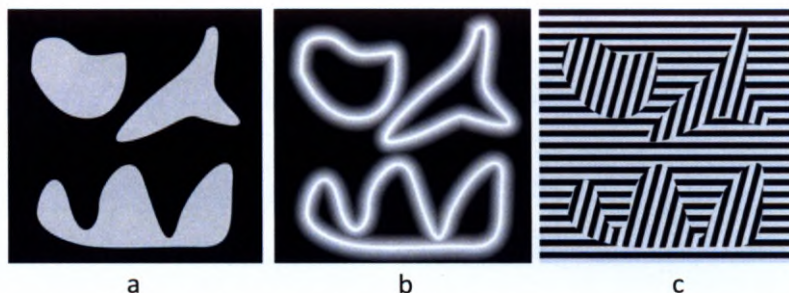
Il. 5.13. Złudzenie optyczne sprawiające, że mamy wrażenie istnienia na tle kółek białego kwadratu i trójkąta, których nie ma

konieczność pójścia w prawą stronę, możliwość utworzenia w jakimś programie wykresu, możliwość powiększenia obrazu lub wyszukania dodatkowych informacji, a także oznaczenie miejsc na pilocie do samochodu, których naciśnięcie spowoduje jego zamknięcie lub otwarcie.

Nasz zmysł wzroku we współpracy z mózgiem tworzy niekiedy wirtualne krawędzie, których w rzeczywistości nie ma. Dobrze ilustruje to złudzenie pokazane na il. 5.13. Takie tworzenie wymaginowanych krawędzi bywa bardzo pomocne w życiu, pomagając np. w ustawianiu mebli. Jest jednak bardzo trudnym zadaniem dla komputerów, które oceniają to, co jest na obrazie, a nie to, co mogłoby być. Wspominamy o tych problemach, żeby zrozumieć, dlaczego detekcja krawędzi jest tak trudna dla komputerowej analizy obrazu.

Krawędzie występujące w obrazach możemy podzielić na trzy podstawowe grupy, które zostały pokazane schematycznie na il. 5.14.

Zdecydowanie najłatwiej jest wykryć krawędzie wtedy, gdy poszczególne obszary silnie różnią się jasnością (il. 5.14a). W takim przypadku do detekcji krawędzi można



Il. 5.14. Schematyczne przedstawienie trzech rodzajów krawędzi spotykanych w obrazach:
a) gwałtowne przejście (silny gradient poziomów szarości) od miejsc jasnych do ciemnych,
b) lokalne maksimum (może czasami być to minimum), c) zmiana tekstury, czyli lokalnej konfiguracji pikseli



Il. 5.15. Detekcja krawędzi obrazu z il. 5.8a, uzyskana jako różnica wyników filtracji maksymalnej i minimalnej

wykorzystać nieliniowe filtry: maksymalny i minimalny. Ponieważ powodują one albo zwiększenie, albo zmniejszenie wielkości jasnych obszarów, to największe różnice wystąpią w miejscach gwałtownych zmian poziomów szarości. Wystarczy zatem od obrazu po filtracji maksymalnej odjąć obraz po filtracji minimalnej, aby otrzymać obraz krawędzi. Ukazuje to il. 5.15. Ten złożony z trzech operacji (filtr maksymalny, filtr minimalny, różnica) filtr do detekcji krawędzi nazywamy filtrem morfologicznym.

Wynik filtracji może ulegać znacznym zmianom wraz ze zmianą macierzy definiującej analizowany obszar. W przypadku il. 5.15 była to macierz o rozmiarze 3x3. Obraz krawędzi jest bardzo dobry dla elementów wykazujących duży kontrast z otoczeniem, jak np. przedni błotnik. Zdecydowanie gorzej jest z krawędziami kół, które są nieciągłe. Ten brak ciągłości w wykrywanych krawędziach to jeden z największych

niedostatków komputerowej analizy obrazu. Niestety, algorytm nie dobudowuje brakujących ogniw krawędzi, tak jak robi to zmysł wzroku z obrazem na il. 5.13.

Drugi model krawędzi – lokalne minima lub maksima (il. 5.14b) – jest stosunkowo łatwy do wykrycia i to nawet wtedy, gdy jasność linii krawędzi oraz ich otoczenia ulegają zmianie. Na ogół wykonuje się to w ten sposób, że od obrazu początkowego odejmujemy obraz po otwarciu lub od obrazu po zamknięciu odejmujemy jego wyjściową wersję.

W przypadku obrazu motocyklisty z il. 5.9a detekcja lokalnych ekstremów nie daje dobrych wyników, gdyż w tym obrazie dominują krawędzie odpowiadające modelowi z il. 5.14a.

Operacje te będą bardziej szczegółowo wyjaśnione podczas omawiania przekształceń morfologicznych, gdzie są nazywane *top hat*. Tam też Czytelnik znajdzie przykłady bardzo skutecznego wykorzystania detekcji lokalnych minimów lub maksimów.

Trzeci rodzaj krawędzi (il. 5.14c), rozumianych jako granica pomiędzy obszarami o lokalnie różnym układzie punktów (czyli obiekty ani nie różnią się wyraźnie poziomem jasności lub kolorem, ani nie są oddzielone wąskim obszarem lokalnie jaśniejszym lub ciemniejszym), jest często bardzo trudny do wykrycia. Nie można tu podać uniwersalnych metod rozwiązania tego problemu. Często jego skuteczne pokonanie wymaga użycia złożonych algorytmów. Na razie nie poznaliśmy jeszcze narzędzi, które to umożliwiają.

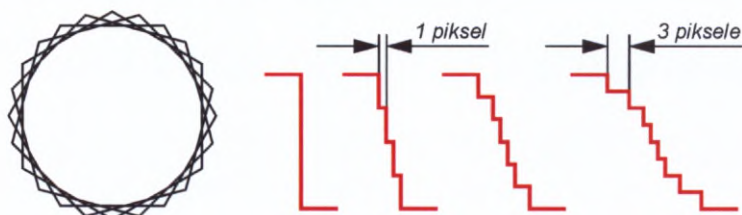
Odmianę tego typu krawędzi stanowią krawędzie rozmytych obiektów. Klasycznym przykładem są chmury (il. 5.16). Można na zdjęciu chmur wskazać obszary, gdzie na pewno nie ma chmur oraz obszary z pewnością zajęte przez chmury. Granica



Il. 5.16. Jednoznaczne wyznaczenie krawędzi chmur jest praktycznie niemożliwe

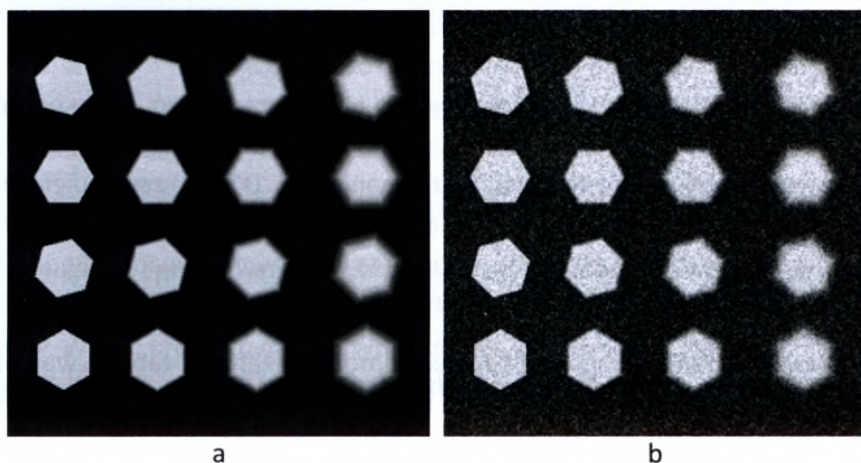
między tymi obszarami jest jednak rozmyta i różne narzędzia wykryją ją w różnych miejscach. W takich przypadkach trzeba wstępnie umówić się, jakie kryterium będziemy stosowali do wyznaczenia krawędzi obiektów. Kryteria oceny będą zależały od rodzaju problemu oraz analizowanego obrazu.

Krawędzie można wykrywać za pomocą dedykowanych filtrów liniowych, które wykrywają je przede wszystkim wtedy, gdy obiekty różnią się poziomem szarości (lub kolorem). Mówimy, że o obecności krawędzi decyduje w takim przypadku lokalny gradient poziomów szarości. Miara wielkości gradientu, czyli lokalnej szybkości jego zmian, jest jego pierwsza pochodna i filtry liniowe w przybliżeniu wyznaczają właśnie tę pierwszą pochodną. W miejscu osiągania przez nią maksymalnych wartości bezwzględnych (bo może przyjmować zarówno wartości dodatnie, jak i ujemne) powinna znajdować się poszukiwana przez nas krawędź.



Il. 5.17. Idea testowego obrazu do detekcji krawędzi. Opis w tekście

Ponieważ wykrywanie krawędzi jest trudne, przygotowano testowy obraz, zawierający specjalnie zaprojektowane krawędzie, na którym można badać skuteczność filtrów do ich detekcji. Jako obiekt do badań wybrano sześciokąt foremny, który zapewnia trzy kierunki krawędzi oraz dwie przeciwne sytuacje przy przecięciu krawędzi z dowolną siecią: przejście od tła do obiektu oraz przejście od obiektu do tła. W testowym obrazie zawarto cztery rodziny sześciokątów: początkową i trzy kolejne obracane: 15°, 30° i 45°. Daje to dobre przybliżenie wszystkich kierunków krawędzi w obrazie (patrz lewa strona il. 5.17). Krawędzie również zamodelowano na cztery sposoby, co pokazuje prawa strona il. 5.17. Pierwsza krawędź została zaprojektowana tak, by wyglądała jak „ucięta nożem”. Po prostu piksele obiektu sąsiadują bez żadnego pośrednictwa ze znacznie ciemniejszymi pikselami tła. W następnym przypadku w okolicach krawędzi wartości pikseli dochodzą do poziomu tła przez trzy wartości pośrednie. Każdy następny poziom szarości zajmuje o 1 piksel więcej. Następna odmiana krawędzi modeluje pewne jej rozmycie. Od obiektu do tła przechodzimy przez pięć wartości pośrednich, które dodają w kierunku prostopadłym do krawędzi kolejno: 2-1-1-1-2 punkty. Wreszcie w ostatniej modelowej krawędzi występuje siedem pośrednich wartości, które zwiększają figurę odpowiednio o 3-2-1-1-1-2-3 piksele.



Il. 5.18. Dwie odmiany modelowego obrazu do testowania metod wykrywania krawędzi. W wierszach stopień rozmycia krawędzi wzrasta od lewej do prawej, zgodnie ze schematem pokazanym na il. 5.17. Kolejne wiersze, patrząc od dołu, zawierają obiekty obrócone o 15° względem poprzedniego wiersza. Wyjściowy obraz testowy (a) występuje również w wersji z dodanym szumem (b)

Różnice w poziomach szarości obiektu i tła nie mają wpływu na skuteczność wykrywania krawędzi. Wyjściowy obraz testowy (il. 5.18a) uzupełniono wersją z dodanym szumem (il. 5.18b), co przybliży do rzeczywistych problemów.

Omówimy teraz w kolejności chronologicznej trzy najpopularniejsze filtry liniowe do detekcji krawędzi. W literaturze angielskojęzycznej wymienia się je jako: *Roberts filter*, *Prewitt filter* oraz *Sobel filter*. Język polski dopuszcza odmienianie nazwisk i stąd często czytamy o filtrach Roberta, Prewitta i Sobela. Jest to jednak poważny błąd, o czym za chwilę.

Najstarszy jest filtr Roberta, nazywany niekiedy krzyżem Roberta. Jego ideę zaproponował w 1963 r. Lawrence G. Roberts. Drugi z filtrów został opracowany w 1966 r. podczas analizy obrazów medycznych przez Judith M. Prewitt. Mówienie o filtrze Prewitta jest zatem poważnym błędem, gdyż powinniśmy używać formy filtr Prewitt. Najpóźniej, bo na konferencji w 1968 r., propozycję kolejnego filtra przedstawił (wspólnie z Jerome A. Feldman) Irwine Sobel.

Na potrzeby analizy filtrów do detekcji krawędzi uzupełnimy zapis filtra liniowego, przedstawionego równaniem (5.1), zapisem macierzy współczynników:

$$F(p_{11}, p_{12}, \dots, p_{33}) = \sum_{i=1}^3 \sum_{j=1}^3 p_{ij} \cdot a_{ij} = \begin{bmatrix} a_{11} & a_{21} & a_{31} \\ a_{12} & a_{22} & a_{32} \\ a_{13} & a_{23} & a_{33} \end{bmatrix} \quad (5.2)$$

Różnice w zapisie polegają na tym, że:

- Ograniczamy się chwilowo do matrycy o rozmiarze 3x3.
- Funkcja F jest obliczana dla środkowego piksela, czyli tego, który odpowiada wartości p_{22} . Pozycja ta jest zaznaczona szarym kolorem tła w tabeli poniżej.
- I najważniejsze – sumę iloczynów w równaniach (5.1) i (5.2) zastępujemy macierzą współczynników. Ten zapis formalnie nie odpowiada rachunkowi macierzowemu, dlatego trzeba pamiętać, co oznacza w naszym przypadku. Będzie jednak wygodny na potrzeby dalszej analizy filtrów do detekcji krawędzi.

Dzięki wprowadzeniu powyższej konwencji możemy zamiast tabeli z wartościami współczynników napisać po prostu ich macierz. Trzeba tylko pamiętać, że w tym przypadku macierz oznacza sumę iloczynów, tak jak pokazuje to równanie (5.2).

a_{11}	a_{21}	a_{31}
a_{12}	a_{22}	a_{32}
a_{13}	a_{23}	a_{33}

$$\begin{bmatrix} a_{11} & a_{21} & a_{31} \\ a_{12} & a_{22} & a_{32} \\ a_{13} & a_{23} & a_{33} \end{bmatrix}$$

W przypadku filtra Roberta definiujemy dwa gradienty $G1$ i $G2$, opisane zgodnie z naszą konwencją równaniem (5.3):

$$G1 = \begin{bmatrix} -1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix} \quad G2 = \begin{bmatrix} 0 & 0 & -1 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix} \quad (5.3)$$

Często można spotkać nieco inną definicję filtra Roberta, która macierz współczynników ilustruje dwoma tabelami, pokazanymi poniżej:

0	1
-1	0

1	0
0	-1

Powyższy zapis nie daje niestety informacji, jaka jest pozycja piksela, dla którego wyznaczamy wartość filtra. Zapis z równania (5.3) nie pozostawia żadnych wątpliwości, ponieważ umówiliśmy się, że obliczenia prowadzimy dla środkowego elementu, któremu w równaniu (5.3) przypisano współczynnik o wartości 1. To, że w jednym przypadku wartości (-1) są w wierszu powyżej wartości (1), a w drugim jest na odwrót, nie ma żadnego znaczenia dla wyniku, bo końcową wartość krawędzi $Z(R1)$ wyznacza się ze wzoru (5.4):

$$Z(R1) = \sqrt{(G1)^2 + (G2)^2} \quad (5.4)$$

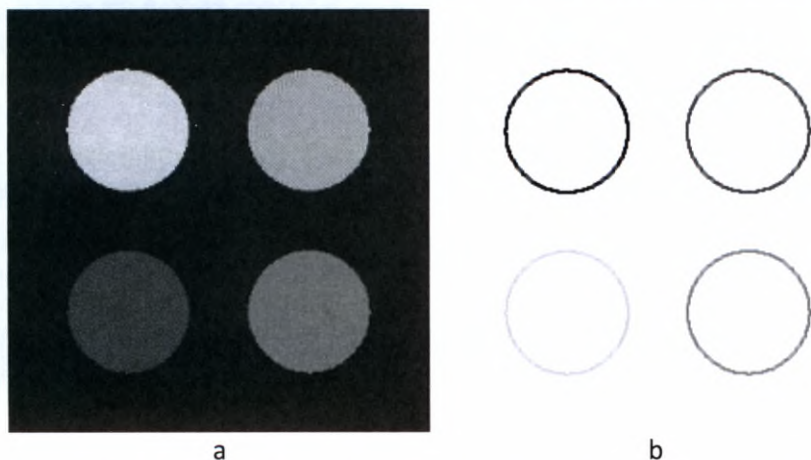
Czasami do wyznaczania wartości krawędzi korzysta się też z prostszej formuły, która jest mniej wymagająca obliczeniowo:

$$Z(R2) = |G1| + |G2| \quad (5.5)$$

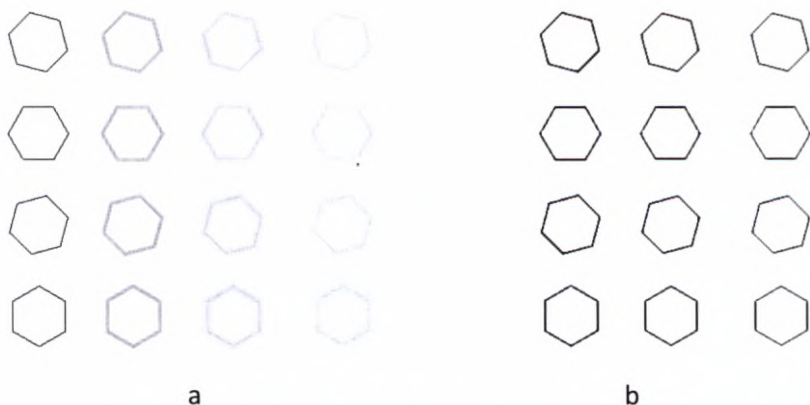
W tym drugim przypadku, gdy dodajemy po prostu bezwzględne wartości gradientów, otrzymamy nieco większe wartości, ale ogólnie wynik detekcji krawędzi będzie bardzo podobny. Wyniki prezentowane w tym rozdziale zostały wyznaczone zgodnie z równaniem (5.4), czyli na podstawie pierwiastka z sumy kwadratów dwóch gradientów.

Czasami, choć zdarza się to niezwykle rzadko, obraz zawiera tak proste elementy, że wyznaczenie ich krawędzi jest bardzo łatwe (il. 5.19). Filtry do detekcji krawędzi nawet w takim prostym przypadku różnicują krawędzie i dają dodatkową informację na temat siły lub wielkości krawędzi. W dalszej części będziemy używali terminu „wartość krawędzi”, gdyż im wyraźniejsza jest krawędź, tym większe wartości dają filtry do detekcji krawędzi. Można to zaobserwować na il. 5.19b. Jaśniejszym obiektom na il. 5.19a odpowiadają wyższe wartości krawędzi. Na il. 5.19b są one ciemniejsze, gdyż w celu poprawy czytelności obraz został znormalizowany i widzimy jego negatyw.

Jak widać na il. 5.20a, filtr Robertsa nie wykrywa dobrze krawędzi, jeżeli są rozmyte (prawa strona il. 5.20a). Detekcję dla rozmytych krawędzi można jednak znacznie poprawić przez prosty filtr maksymalny (il. 5.20b). Na oryginalnym obrazie krawędzi należałoby przeprowadzić filtrację minimalną, ale ponieważ do celów wizualizacji



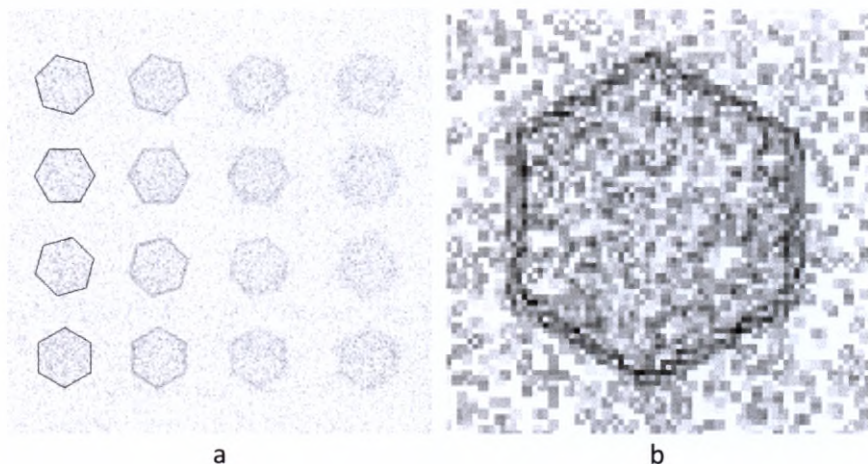
Il. 5.19. Detekcja krawędzi filtrem Robertsa: a) obraz początkowy, b) znormalizowany, negatywny obraz wykrytych krawędzi



Il. 5.20. Detekcja krawędzi filtrem Roberta na obrazie z il. 5.18a: a) wszystkie krawędzie, b) obraz (a) po filtracji maksymalnej i zwiększeniu kontrastu

pracujemy na negatywie obrazu, to przeprowadzamy filtrację maksymalną. Jak widać na il. 5.20b, można otrzymać dzięki tej prostej operacji perfekcyjny wynik detekcji krawędzi. Warto jednak zwrócić uwagę, że filtracja maksymalna spowodowała usunięcie obrazu krawędzi widocznych w lewej kolumnie il. 5.20a. Tego typu operacje wymagają dużej ostrożności, gdyż mogą prowadzić do błędów.

Detekcja krawędzi na obrazie testowym zakończyła się sukcesem (il. 5.20). Sprawdźmy teraz działanie filtracji Roberta na obrazie testowym z szumem (tys. 5.18b). Ponieważ filtr Roberta używa bardzo małej matrycy o rozmiarze 2×2 , to może wykryć mocną krawędź nawet w przypadku pojedynczego piksela o poziomie szarości wyraźnie



Il. 5.21. Detekcja krawędzi w obrazie z szumem (il. 5.18b) za pomocą filtru Roberta: a) znormalizowany negatyw obrazu wykrytych krawędzi, b) powiększony fragment drugiego od lewej elementu z dolnego wiersza

odbiegającym od tła. Znajduje to potwierdzenie w wyniku detekcji krawędzi, pokazanym na il. 5.21a, gdzie jedynie pierwsza kolumna obiektów z krawędziami bez jakiegokolwiek rozmycia została poprawnie i wyraźnie wykryta. Nawet niewielkie rozmycie krawędzi powoduje, że filtr Roberta praktycznie uniemożliwia poprawne wykrycie krawędzi na obrazie z szumem. Widać to wyraźnie na powiększonym fragmencie drugiego od lewej elementu w dolnym wierszu (il. 5.21b). W ten sposób poznaliśmy największą słabość filtra Roberta, którą jest bardzo duża wrażliwość na obecność szumu w obrazie.

Przejdziemy teraz do nieco późniejszego rozwiązania, czyli filtra Prewitt. Wyznacza on krawędzie zgodnie z zasadą opisaną równaniem (5.4) na podstawie dwóch gradientów, wykrywających krawędzie przebiegające poziomo i pionowo lub blisko tych kierunków. Gradienty te oznaczmy odpowiednio $G_x(P)$ oraz $G_y(P)$:

$$G_x(P) = \frac{1}{3} \cdot \begin{bmatrix} 1 & 1 & 1 \\ 0 & 0 & 0 \\ -1 & -1 & -1 \end{bmatrix} \quad G_y(P) = \frac{1}{3} \cdot \begin{bmatrix} 1 & 0 & -1 \\ 1 & 0 & -1 \\ 1 & 0 & -1 \end{bmatrix} \quad (5.6)$$

Gradienty Prewitt preferują odcinki krawędzi o długości trzech pikseli, co ogranicza wrażliwość na szumy, która jest główną wadą filtra Roberta. Dodatkowo jeden wiersz lub pojedyncza kolumna współczynników o wartości 0 ogranicza wpływ rozmycia krawędzi na wynik detekcji. Współczynnik $1/3$ we wzorze gradientu Prewitt wynika z dążenia do uzyskania porównywalnej wartości krawędzi, niezależnie od użytego filtra. W filtrze Roberta gradient wyznaczała pojedyncza różnica wartości pikseli sąsiadujących ze sobą na skos. W przypadku filtra Prewitt w gradientie dodajemy do siebie trzy różnice wartości pikseli. Stąd należało wartość gradientu podzielić przez 3.

Detekcja krawędzi w obrazie testowym za pomocą filtra Prewitt (il. 5.22a) daje wyniki zbliżone do filtra Roberta (il. 5.20a). Można nawet dopatrzeć się pewnej przewagi filtra Roberta, gdyż w drugiej kolumnie obiektów daje cieńsze krawędzie. Jest to oczywisty efekt nałożenia się sposobu zdefiniowania obrazu testowego (il. 5.17) i większej matrycy filtra. Nie jest to jednak różnica znacząca. Najistotniejszą przewagą filtra Prewitt jest jednak jego zdecydowanie mniejsza wrażliwość na szum, co widać wyraźnie, jeśli porównamy il. 5.22b i 5.21a.

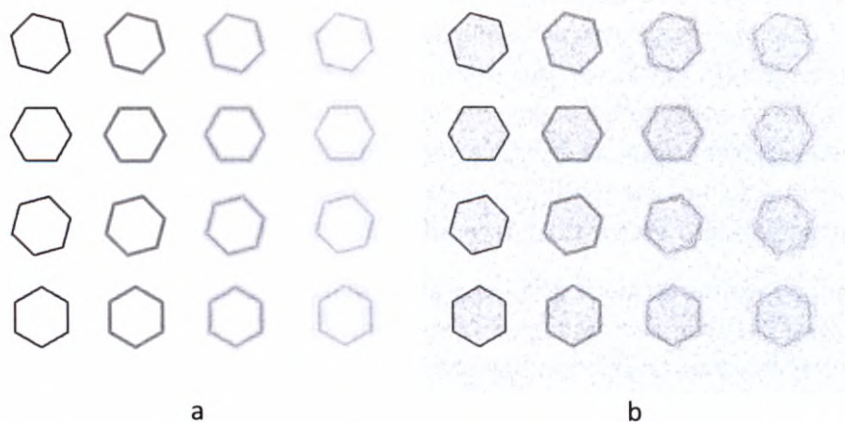
Można zdefiniować filtr bardzo podobny do filtra Prewitt, ale o wielkości nie 3×3 , tylko 5×5 . Matryce tych filtrów wyglądałyby następująco:

dla gradientu G_x :

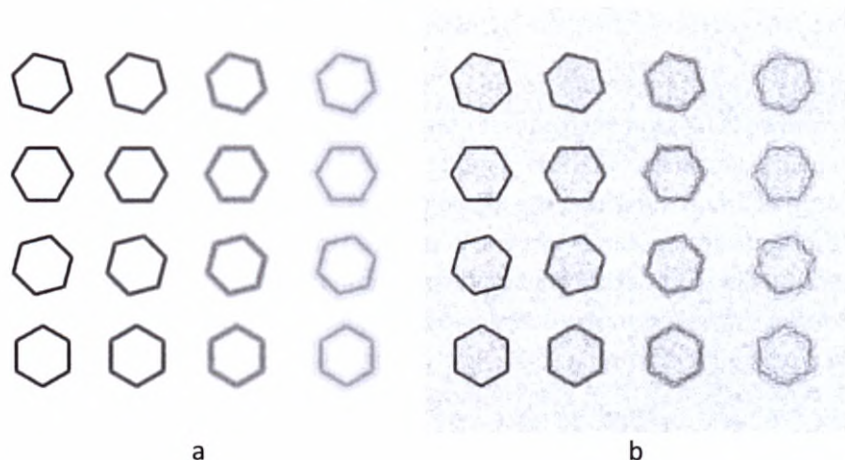
$$\begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 \\ -1 & -1 & -1 & -1 & -1 \\ -1 & -1 & -1 & -1 & -1 \end{bmatrix}$$

i dla gradientu G_y :

$$\begin{bmatrix} 1 & 1 & 0 & -1 & -1 \\ 1 & 1 & 0 & -1 & -1 \\ 1 & 1 & 0 & -1 & -1 \\ 1 & 1 & 0 & -1 & -1 \\ 1 & 1 & 0 & -1 & -1 \end{bmatrix}$$



Il. 5.22. Wyniki detekcji krawędzi obrazu testowego z il. 5.18, uzyskane za pomocą filtra Prewitt:
a) obraz testowy bez szumu, b) obraz testowy z szumem



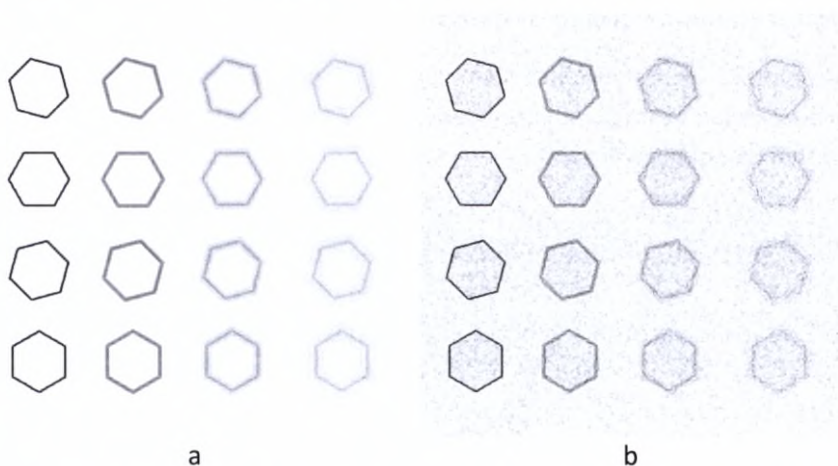
Il. 5.23. Wyniki detekcji krawędzi za pomocą filtra o konstrukcji podobnej do filtra Prewitt, ale o wielkości 5x5: a) detekcja na obrazie testowym bez szumu, b) detekcja na obrazie testowym z szumem

Współczynnik poprawkowy wyniósłby $1/10$ zamiast $1/3$, jak podano to w równaniu (5.6). Znaczna liczba branych pod uwagę pikseli powinna spowodować zmniejszenie wrażliwości na szum (il. 5.23). Zgodnie z oczekiwaniami tak zdefiniowany filtr jest mniej wrażliwy na szum, ale jednocześnie jest mniej czuły na drobne zmiany w przebiegu krawędzi.

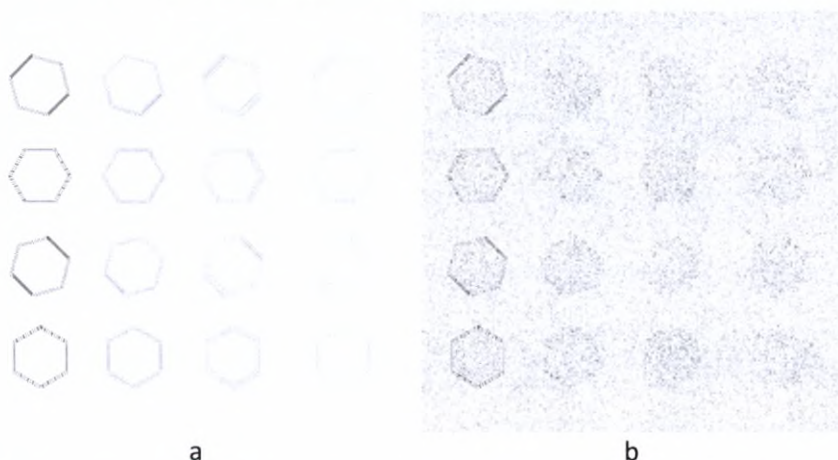
Można też zdefiniować filtr łączący niejako cechy filtrów Robertsa i Prewitt, który wykorzystuje matryce Prewitt, ale obrócone o kąt 45° . Nazwiemy go filtrem Prewitt 45° , a odpowiednie gradienty oznaczymy $G_{45}(P)$ i $G_{135}(P)$:

$$G_{45}(P) = \frac{1}{3} \cdot \begin{bmatrix} 1 & 1 & 0 \\ 1 & 0 & -1 \\ 0 & -1 & -1 \end{bmatrix} \quad G_{135}(P) = \frac{1}{3} \cdot \begin{bmatrix} 0 & -1 & -1 \\ 1 & 0 & -1 \\ 1 & 1 & 0 \end{bmatrix} \quad (5.7)$$

Wyniki detekcji filtrem Prewitt 45° pokazuje il. 5.24. Nie widać jednak istotnych różnic w porównaniu z wynikiem filtracji klasycznym filtrem Prewitt (il. 5.22). Ponieważ nasze odczucia mogą być subiektywne, należy zmierzyć różnice między tymi dwoma wynikami detekcji. il. 5.25 zawiera obrazy różnic pomiędzy omawianymi wynikami detekcji.



Il. 5.24. Wyniki detekcji krawędzi obrazu testowego z il. 5.18, uzyskane za pomocą filtra Prewitt 45°: a) obraz testowy bez szumu, b) obraz testowy z szumem



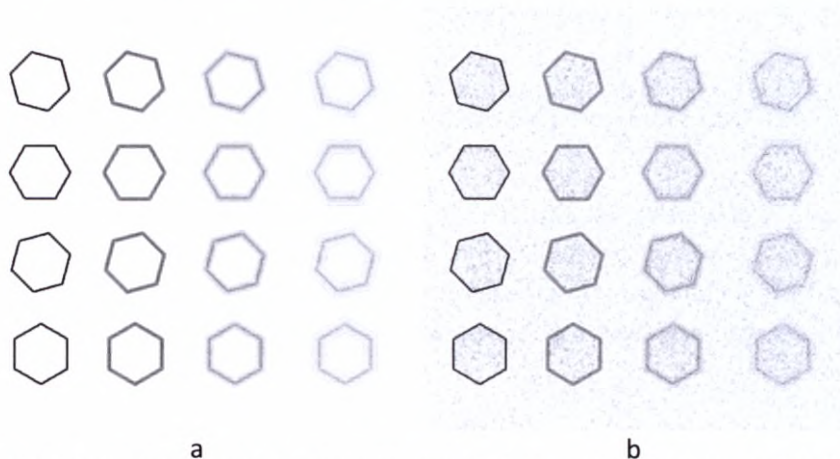
Il. 5.25. Różnica wyników detekcji krawędzi obrazu testowego z il. 5.18, uzyskana za pomocą filtrów Prewitt i Prewitt 45°. W celu poprawy czytelności zakres zmian poziomów szarości został rozszerzony pięciokrotnie: a) obraz testowy bez szumu, b) obraz testowy z szumem

W celu ich lepszego zwizualizowania zakresy poziomów szarości zostały zwiększone pięciokrotnie. Obserwacja obrazów różnic pomiędzy wynikami detekcji, widocznych na il. 5.25, potwierdza, że dodanie nowej konfiguracji maczy gradientów w filtrze Prewitt nie wprowadza istotnego postępu w jakości wykrywania krawędzi.

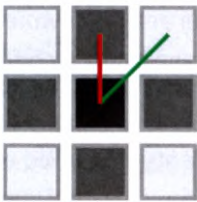
Ponieważ niewielkie różnice w wynikach detekcji jednak występują, to należy podjąć próbę wykorzystania jednocześnie wszystkich przedstawionych konfiguracji maczy filtra, opisanych wzorami (5.6) i (5.7). Wyniki takiej filtracji pokazano na il. 5.26. Wartości krawędzi widocznych na tym rysunku odpowiadają maksimum z detekcji klasycznym filtrem Prewitt oraz filtrem Prewitt 45° .

Podwojenie liczby maczy filtra nie przynosi, jak pokazuje il. 5.26, znaczących korzyści, a zwiększa złożoność obliczeniową filtra. Z tego względu w praktyce używa się klasycznej wersji filtra Prewitt.

W przypadku filtra Prewitt pod uwagę brane są wszystkie piksele w maczy 3×3 . Wartości gradientu oblicza się dla centralnego piksela, zaznaczonego czarnym kolorem na il. 5.27. Jeżeli weźmiemy pod uwagę odległości od centralnego piksela, to piksele leżące w narożnikach otoczenia 3×3 (jasne kwadraty na il. 5.27) mają większy wpływ na wartość krawędzi niż piksele położone najbliżej (ciemne kwadraty na il. 5.27). Jako miarę wagi danego piksela proponuje się kwadrat jego odległości od centralnego punktu. Odpowiednie odległości dla dwóch przykładowych pikseli zaznaczono na il. 5.27 w postaci czerwonego i zielonego odcinka. Jeśli kwadrat odległości najbliższego sąsiada wynosi 1, to kwadrat odległości narożnego piksela wynosi 2. Aby wyrównać wpływ poszczególnych pikseli na wartość krawędzi, w filtrze



Il. 5.26. Wyniki detekcji krawędzi obrazu testowego z il. 5.18, uzyskane za pomocą filtra Prewitt z czterema orientacjami gradientu: a) obraz testowy bez szumu, b) obraz testowy z szumem



Il. 5.27. Schemat odległości pikseli w matrycy filtra. Opis w tekście

Sobela przy obliczaniu gradientów zmieniono wartości współczynników na przedstawione we wzorze (5.8):

$$Gx(S) = \frac{1}{4} \cdot \begin{bmatrix} 1 & 2 & 1 \\ 0 & 0 & 0 \\ -1 & -2 & -1 \end{bmatrix}$$
$$Gy(S) = \frac{1}{4} \cdot \begin{bmatrix} 1 & 0 & -1 \\ 2 & 0 & -2 \\ 1 & 0 & -1 \end{bmatrix}$$

(5.8)

Konsekwencją zmian w matrycy gradientu stała się konieczność zmiany wartości współczynnika korygującego wartość gradientu na 1/4. Współczynniki korygujące wprowadzono w celu uzyskania podobnych wartości krawędzi dla różnych filtrów. Możliwa liczba lokalnych konfiguracji wartości pikseli jest w praktyce nieograniczona. Przyjmując 256 wartości dla każdego piksela, otrzymamy dla matrycy 3x3 zawrotną liczbę 2⁷² kombinacji. Jeżeli odejmiemy konfiguracje możliwe do uzyskania dzięki lustrzanemu odbiciu lub obrotom, to nadal pozostaje 2⁶⁴ możliwych konfiguracji pikseli. Przeanalizujemy zatem jedynie kilka najprostszych konfiguracji, w których wszystkie piksele obiektu mają taką samą

Tabela 5.1. Wartości krawędzi podawane przez różne filtry do detekcji krawędzi.
Analizowany fragment obrazu o wielkości 3x3 zawiera piksele o wartości 1, zaznaczone jako białe, oraz piksele o wartości 0, zaznaczone jako czarne

filtr						
Roberts	1,41	1	0	1	1,41	1,41
Prewitt	1	1	1	0	0	0
Sobel	1	1	1	0	0	0

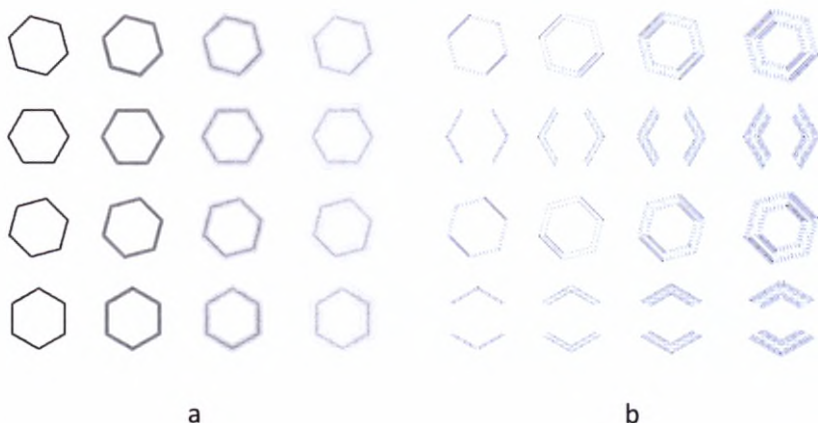
Roberts	1,41	1	0	1	0	1
Prewitt	0,94	0,94	0,94	0,94	0,75	0,75
Sobel	1,06	1,06	1,06	1,06	0,79	0,79

wartość. Piksele tła na potrzeby tej analizy również mają te same wartości. Wartości krawędzi, wyznaczone za pomocą analizowanych filtrów liniowych, zawiera tabela 5.1.

Jedynie w jednym przypadku wszystkie filtry dały takie same wartości. Filtry Prewitt i Sobela dają wartości albo identyczne, albo różniące się nie więcej niż 12%. Potwierdza to zasadność wprowadzonych współczynników korekcyjnych przy obliczaniu wartości gradientów. Natomiast ograniczenie macierzy filtra Robertsa prowadzi w większości przypadków do wyników zdecydowanie różniących się od pozostałych filtrów. W dodatku obrót konfiguracji pikseli, np. w przypadku trzech pierwszych konfiguracji w górnym wierszu tabeli 5.1, prowadzi do uzyskania innych wartości krawędzi. Pozostałe dwa filtry są odporne na tego typu modyfikacje.

W praktyce różnice pomiędzy wynikami detekcji krawędzi za pomocą filtrów Sobela oraz Prewitt są niewielkie i zazwyczaj niezauważalne dla naszego zmysłu wzroku. Potwierdza to porównanie il. 5.22a oraz 5.28a. Weryfikacją naszych wrażeń wzrokowych jest zbadanie względnej różnicy pomiędzy wynikami obu detekcji. Zostało one wyznaczone jako stosunek wartości bezwzględnej różnicy pomiędzy wynikami detekcji filtrami Sobela i Prewitt do wartości krawędzi według filtra Sobela. Negatywny obraz względnych różnic pokazuje il. 5.28b. Jak widać, wartości są identyczne dla krawędzi zorientowanych poziomo lub pionowo. W pozostałych przypadkach różnice są na poziomie 10-20%, co potwierdza wyniki przedstawione w tabeli 5.1. Jedynie w przypadku pojedynczych pikseli w wierzchołkach wielokątów, zdarzały się różnice na poziomie ok. 50%.

Składowe wartości gradientów można wykorzystać do przybliżonego wyznaczenia kąta orientacji krawędzi α . Służy do tego równanie (5.9):

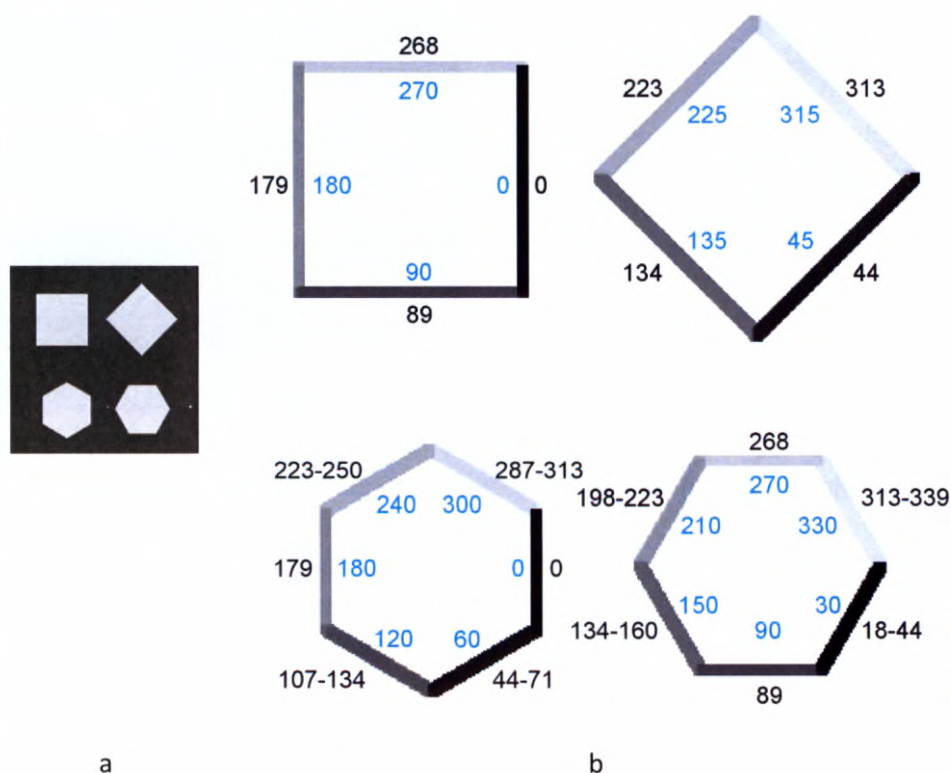


Il. 5.28. Wyniki detekcji krawędzi, uzyskane za pomocą filtra Sobela: a) krawędzie obrazu testowego z il. 5.18a, b) względne różnice z wynikiem detekcji za pomocą filtra Prewitt

$$\alpha = \arctan \frac{G_y}{G_x} \quad (5.9)$$

Na il. 5.29b pokazano orientację krawędzi wykrytych filtrem Prewitt dla testowego obrazu z il. 5.29a. Dla krawędzi zorientowanych poziomo oraz dla wszystkich krawędzi, których kąt orientacji jest wielokrotnością 45° , zanotowano bardzo dobrą zgodność pomiaru z wartościami teoretycznymi. Odstępstwo od wartości teoretycznych nie przekraczało 2° .

W przypadku sześciokątów na il. 5.29b widzimy teoretyczne i zmierzone wartości kątów, będących wielokrotnością kąta 30° . Dla krawędzi poziomych lub pionowych kąty są oczywiście identyczne z kątami dla tak samo zorientowanych krawędzi kwadratów. W przypadku pozostałych kątów nie ma już takiej zgodności i filtr wyznaczył kilka różnych wartości, które powtarzają się cyklicznie wzdłuż krawędzi. Aby zrealizować

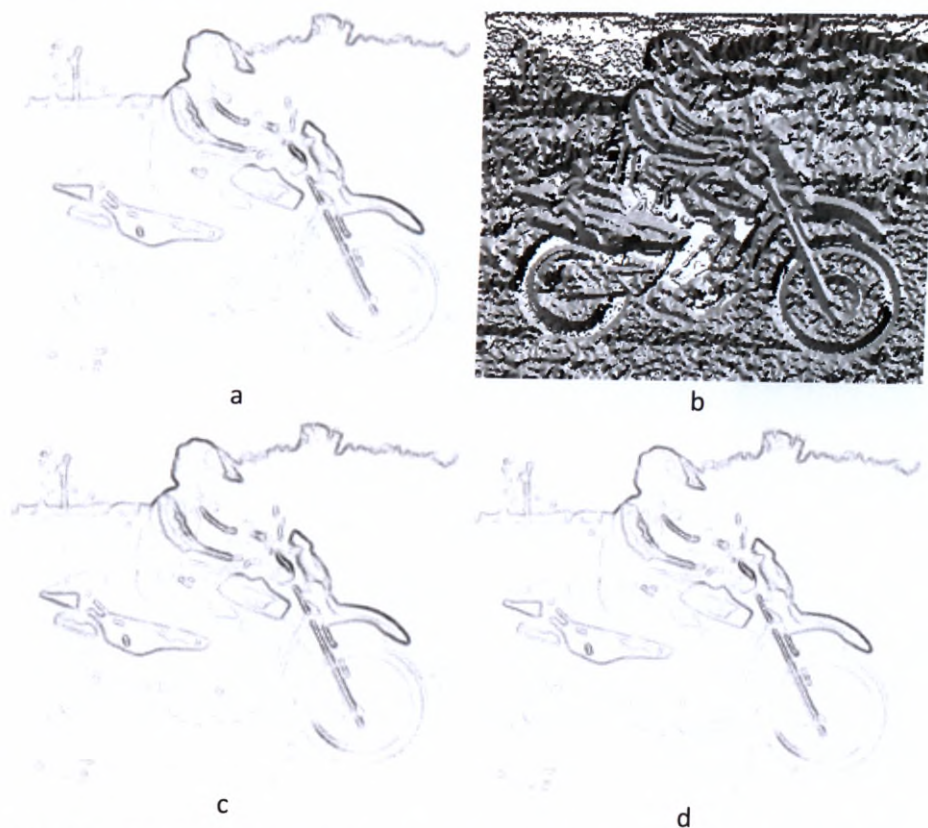


Il. 5.29. Orientacja krawędzi: a) obraz testowy z ostrymi krawędziami zorientowanymi pod różnymi kątami, b) obraz orientacji krawędzi wykrytych filtrem Prewitt; rzeczywiste kąty orientacji krawędzi opisano niebieską czcionką, a kąty wyznaczone za pomocą filtra opisano czarną czcionką

orientację pod kątem 30° , piksele tworzą schodki o długości w cyklicznej sekwencji 122122212221222 piksele, co daje 26 pikseli w poziomie i 15 w pionie. Tangens tak utworzonego kąta daje wartość równą 99,9% wartości tangensa kąta 30° . Przedstawiona powyżej zróżnicowana konfiguracja pikseli powoduje, że lokalnie wyznaczone kąty orientacji są różne i dlatego na il. 5.29b podano ich zakresy, a nie wartości.

Obraz kątów orientacji jest 8-bitowym obrazem szarym i kąty są podawane w zakresie 0-255. Chcąc otrzymać wartości kątów w stopniach, wartości pikseli należy pomnożyć przez 1,4117, czyli wynik dzielenia 360 przez 255.

Odejdziemy teraz na chwilę od obrazów testowych i sprawdzimy, jak omawiane filtry działają na innych obrazach. Wyniki filtracji pokazane na il. 5.30 wydają się być bardzo podobne dla wszystkich filtrów, choć – jak wiemy – w rzeczywistości występują między nimi różnice. Aby sprawdzić, na ile różna jest geometria wykrywanych krawędzi,



Il. 5.30. Wyznaczanie krawędzi dla obrazu motocyklisty: a) krawędzie po filtracji Sobela, b) orientacja krawędzi Sobela, c) krawędzie Prewitt oraz d) krawędzie Robertsa. W celu zwiększenia czytelności obrazy a, c oraz d zostały pokazane jako negatywy oryginalnych obrazów



Il. 5.31. Różnica pomiędzy krawędziami wykrytymi filtrem Roberta i Sobela:

a) wynik odejmowania odpowiednich obrazów krawędzi, b) wartość bezwzględna obrazu (a), dodatkowo nieliniowo wzmocniona w celu poprawy czytelności obrazu

na il. 5.31 przedstawiono różnicę pomiędzy wynikami filtracji Roberta i Sobela. W celu uzyskania obiektywnego wyniku wyjściowe obrazy krawędzi zostały znormalizowane, czyli zakres poziomów szarości zmieniono tak, aby wartość minimalna wynosiła 0, a maksymalna 255. Brak różnicy daje wartość zerową, której odpowiada 50% szarości na il. 5.31a. Jaśniejsze piksele to punkty, gdzie krawędzie Roberta były silniejsze od krawędzi Sobela, a piksele ciemniejsze odpowiadają sytuacji przeciwnej.

Lepsze uwidocznienie różnic w geometrii krawędzi daje wartość bezwzględna il. 5.31a, przedstawiona na il. 5.31b. Jak widać, w wielu miejscach filtry dały nieco inny obraz krawędzi, ale jeśli patrzy się na te obrazy wyświetlone obok siebie, to trudno dostrzec istotne różnice.

Często implementowanym algorytmem detekcji krawędzi jest algorytm Canny–Deriche. Canny zaproponował sposób detekcji krawędzi, wykorzystujący zoptymalizowane filtry oparte na gradientach. Na podstawie teorii sygnałów zaproponował trzy kryteria optymalizujące detekcję krawędzi:

- stosunek sygnału do szumu po detekcji,
- odchylenie standardowe lokalizacji wykrytej krawędzi względem teoretycznie najlepszej,
- średnia odległość pomiędzy krawędziami wykrywanymi z powodu obecności szumu.

Umożliwiło to zbudowanie zoptymalizowanego filtra do detekcji krawędzi. Deriche istotnie ulepszył filtr opracowany przez Canny’ego dzięki rozszerzeniu zakresu filtrów na zbiór funkcji o nieograniczonej dziedzinie. Dzięki temu ograniczono możliwość



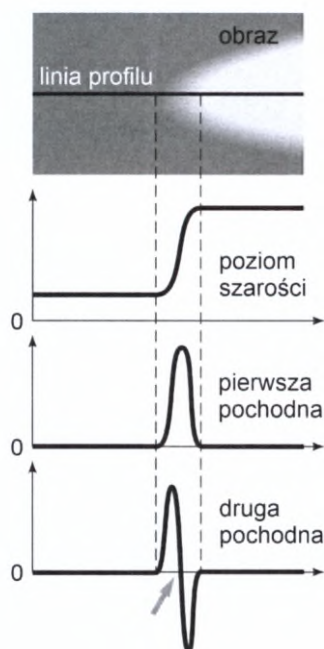
Il. 5.32. Krawędzie otrzymane algorytmem Canny–Deriche (a) oraz negatyw tego obrazu krawędzi (b)

wielokrotnego wykrywania krawędzi. Przykład działania filtra Canny–Deriche pokazuje il. 5.32. Nie będziemy jednak szczegółowo omawiali tego filtra.

Zajmiemy się teraz nieco bliżej matematycznym opisem krawędzi obiektu, którego poziom szarości istotnie różni się od tła (il. 5.33). Jeżeli mamy obiekt istotnie jaśniejszy od tła (patrz obraz u góry il. 5.33), to zwykle przejście od obszaru ciemnego do jasnego jest nieco rozmyte (może to być ograniczone do obszaru o szerokości zaledwie kilku pikseli i wręcz niezauważalne dla człowieka bez dokładnej analizy pojedynczych pikseli). W takim przypadku postuluje się, że krawędź odpowiada połowie zakresu między tłem i obiektem (będzie to miejsce w środku, pomiędzy przerywanymi liniami na il. 5.33). Ten postulat odpowiada naszemu intuicyjnemu wyborowi lokalizacji krawędzi. Zmianę poziomów szarości wzdłuż wybranej linii profilu ilustruje funkcja pokazana schematycznie pod analizowanym obrazem na il. 5.33.

Jeżeli wyznaczy się pierwszą pochodną tej funkcji, to jej maksimum przypadnie dokładnie w miejscu poszukiwanej krawędzi. W przypadku analizy przejścia w przeciwnym kierunku, czyli od jasnego do ciemnego, krawędź przypadnie w miejscu minimum pierwszej pochodnej. Wystarczy jednak wziąć pod uwagę jej wartość bezwzględną, aby maksima tej funkcji odpowiadały miejscom występowania krawędzi.

Opisany powyżej sposób postępowania wykorzystują omówione już wcześniej filtry Sobela, Prewitt oraz Robertsa, które przybliżają pierwszą pochodną. Wspólną wadą tych filtrów jest to, że z powodu dyskretnego charakteru cyfrowych obrazów wartości maksimów pierwszej pochodnej wykazują znaczne rozrzuty oraz wrażliwość na szumy. Jedną z najistotniejszych konsekwencji tego problemu jest brak ciągłości wykrytych krawędzi, co utrudnia lub nawet uniemożliwia skuteczną detekcję obiektów.

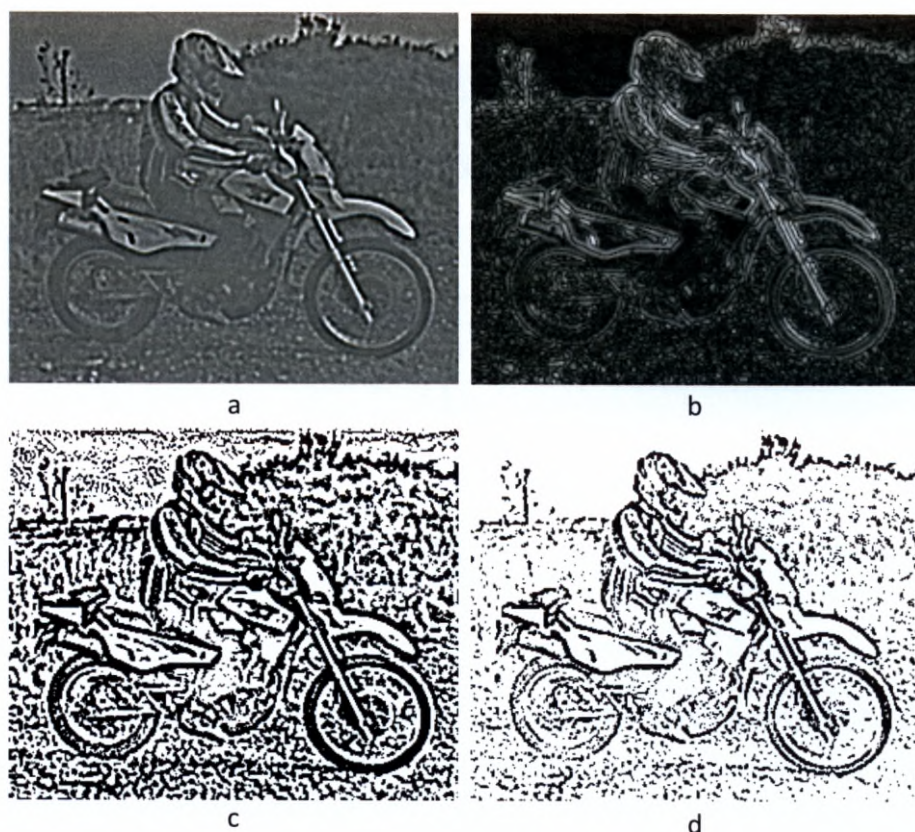


Il. 5.33. Przebiegi funkcji, które można wykorzystać do detekcji krawędzi. Opis w tekście

Właściwym narzędziem do rozwiązania tego problemu wydaje się być wykorzystanie drugiej pochodnej. Miejscem krawędzi powinny być miejsca zerowe drugiej pochodnej przy przejściu pomiędzy wartościami o przeciwnych znakach (strzałka na il. 5.33). Przewagą tego rozwiązania polega na tym, że lokalizacja miejsc zerowych drugiej pochodnej nie zależy od wartości maksimum pierwszej.

Matematycznym operatorem, będącym sumą cząstkowych pochodnych drugiego stopnia, jest laplasjan. Można go w przypadku cyfrowych obrazów wyznaczać w przybliżeniu, wykorzystując wiele różnych macierzy liniowych filtrów. Wspólną ich cechą jest to, że suma współczynników filtra wynosi zawsze zero. Kilka macierzy do wyznaczania laplasjanu ilustruje tabela 5.2 na następnej stronie.

Przykład działania laplasjanu pokazuje il. 5.34a. Zgodnie z modelem pokazanym na il. 5.33 z jednej strony krawędzi mamy wartości dodatnie (najjaśniejsze miejsca na il. 5.34a), które sąsiadują z wartościami ujemnymi (najciemniejsze miejsca na il. 5.34a). Miejsca zerowe drugiej pochodnej powinny być dobrze widoczne na obrazie wartości bezwzględnej laplasjanu jako najciemniejsze miejsca (il. 5.34b). W praktyce, jeżeli chcemy uzyskać możliwie dużo ciągłych krawędzi, musimy wziąć pod uwagę nie tylko wartości zerowe, co prowadzi do powstania grubych krawędzi widocznych na il. 5.34c. Lepszy wynik (il. 5.34d) daje filtr nazywany *zero crossing*,



Il. 5.34. Detekcja krawędzi na obrazie motocyklisty: a) wynik działania laplasjanu, b) wartość bezwzględna laplasjanu, c) bardzo silne skontrastowanie obrazu (a) oraz d) wynik detekcji filtrem *zero crossing*. Opis w tekście

który wyznacza krawędzie w przybliżeniu w miejscach przejścia laplasjanu przez miejsca zerowe. Różnica w porównaniu z il. 5.34c wynika z innego sposobu wyznaczania laplasjanu. W filtrze *zero crossing* wykorzystano metodę DOG (*difference of Gaussians*), w której laplasjan jest przybliżany za pomocą różnicy dwóch obrazów po rozmyciu filtrem Gaussa o różnych rozmiarach.

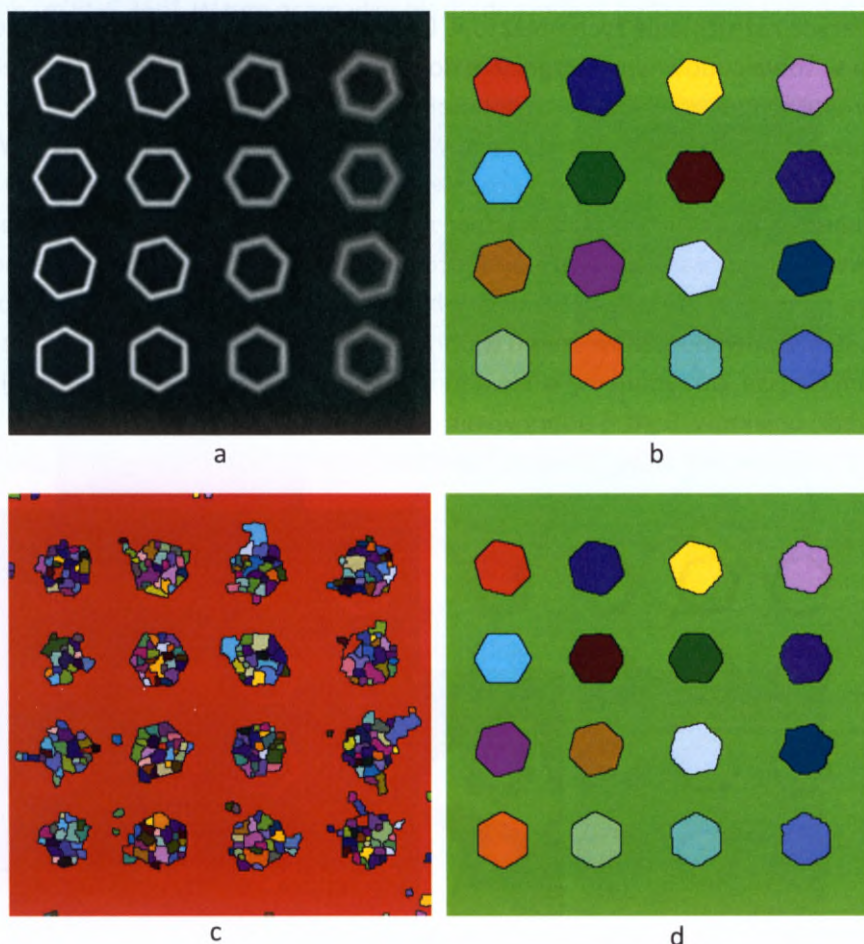
Tabela 5.2. Przykłady macryc filtrów liniowych wyznaczających laplasjan

-1	0	-1
0	4	0
-1	0	-1

-1	-1	-1
-1	8	-1
-1	-1	-1

-1	-1	-1	-1	-1
-1	-1	-1	-1	-1
-1	-1	24	-1	-1
-1	-1	-1	-1	-1
-1	-1	-1	-1	-1

Podsumowując przegląd metod detekcji krawędzi, można stwierdzić, że w przypadku bardzo dobrze widocznych krawędzi wykrywa je poprawnie praktycznie każdy algorytm. Jednak taka sytuacja występuje najczęściej tylko na odpowiednio dobranych i wstępnie przetworzonych obrazach z programów demonstracyjnych. W przypadku rzeczywistych obrazów skuteczna i poprawna detekcja krawędzi wciąż jest bardzo trudna do uzyskania. Na il. 5.35b i 5.35d pokazano dość spektakularne wyniki detekcji krawędzi obiektów z użyciem działów wodnych. To interesujące przekształcenie i możliwości jego wykorzystania będą omawiane w ramach przekształceń morfologii matematycznej. W tym miejscu ograniczymy się do stwierdzenia, że detekcja działów

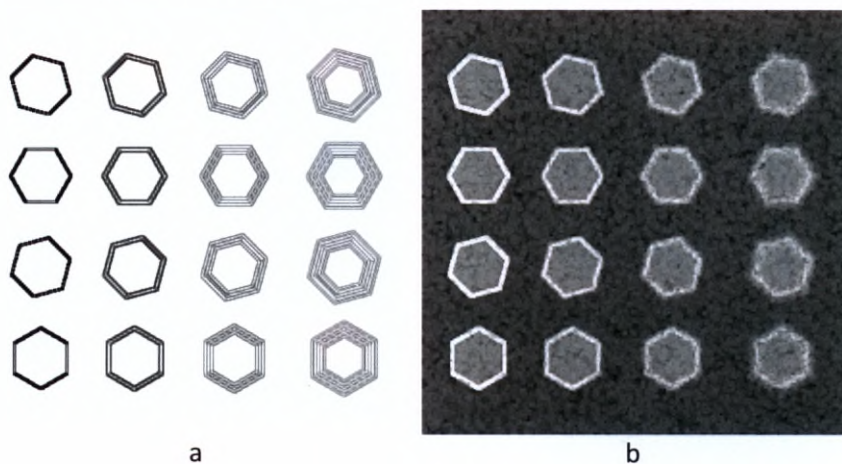


Il. 5.35. Wykorzystanie detekcji działów wodnych do wyznaczenia krawędzi obrazu testowego z il. 5.18: a) wygładzony filtrem Gaussa wynik detekcji filtrem Prewitt, b) detekcja działów wodnych przeprowadzona na obrazie (a), c) analogiczna operacja przeprowadzona na obrazie z szumem, d) detekcja krawędzi na obrazie z szumem, wykorzystująca dodatkowe znaczniki

wodnych to dość złożony proces, w którym bardzo często wyniki działania prostych filtrów do detekcji krawędzi są istotnym elementem całego algorytmu.

Il. 5.35a przedstawia wynik detekcji filtrem Prewitt z późniejszym wygładzeniem filtrem Gaussa. Daje to przebieg zmian poziomów szarości, zbliżony do gładkiej krzywej pierwszej pochodnej (por. il. 5.33). Detekcja działów wodnych to dość złożony algorytm, który wyszukuje lokalne maksima i ogranicza je do grubości pojedynczych pikseli. Dzięki temu uzyskujemy spektakularny wynik (il. 5.35b).

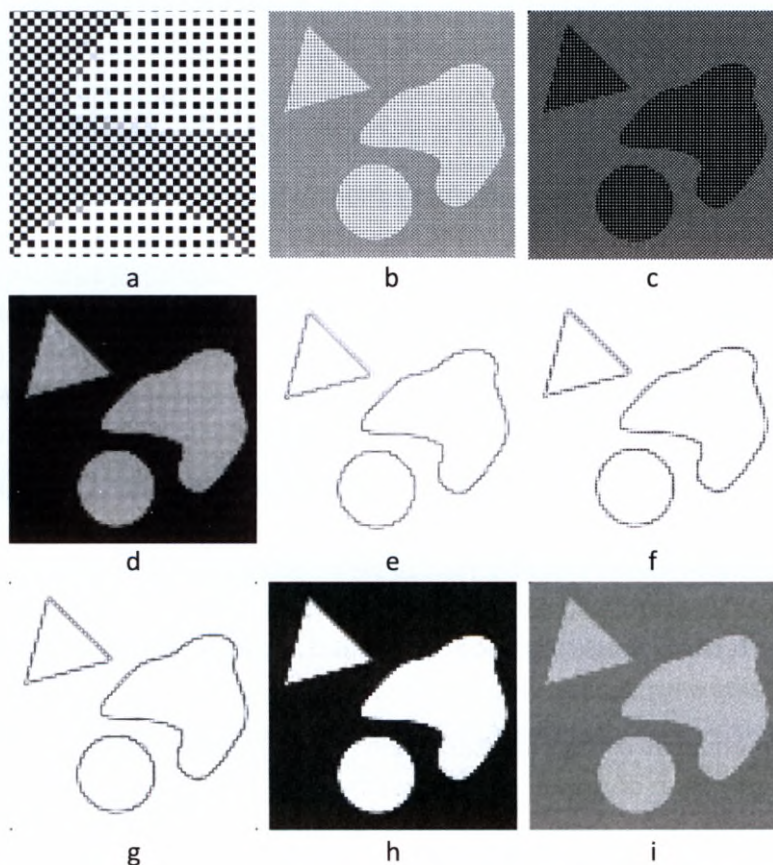
Kolorowe obrazy na il. 5.35b-d są wynikiem zastosowania obrazów etykietowanych, które są wyświetlane w ten sposób, że każdemu obiektowi przypisuje się inny kolor, co ułatwia rozróżnianie tych obiektów. Detekcja działów wodnych nie jest jednak sama w sobie cudownym narzędziem do detekcji granic ziaren. Przekształcenie to po prostu wygenerowało linie o grubości pojedynczego piksela na podstawie obrazu krawędzi uzyskanego innym filtrem. W tym przypadku był to filtr Prewitt. Ważne jest też to, że obraz z il. 5.18a jest pozbawiony szumu. Dzięki temu wygładzony obraz gradientowy (il. 5.35a) ma stałą wartość zerową poza rozmytym obszarem krawędzi obiektów, co pozwala na dobre odwzorowanie krawędzi (il. 5.35b). Próba powtórzenia powyższej metody detekcji na obrazie z szumem kończy się pełną porażką (il. 5.35c). Z powodu śladów szumu w obrazie gradientowym, pomimo jego rozmycia filtrem Gaussa, otrzymujemy wiele nadmiarowych podziałów, czyli nieistniejących krawędzi obiektów. Bardzo dobry wynik detekcji można jednak uzyskać po wykorzystaniu odpowiednio przygotowanych znaczników (il. 5.35d). Trzeba jednak pamiętać,



Il. 5.36. Wykorzystanie filtra morfologicznego do detekcji krawędzi obrazu z il. 5.18:
a) negatyw wykrytych krawędzi obrazu bez szumu, poddany gamma modulacji z $\gamma = 0,19$,
b) wynik detekcji na obrazie z szumem

że generowanie takich znaczników czasami jest bardzo trudne lub wręcz niemożliwe do automatycznego przygotowania. Więcej na temat techniki działów wodnych znajdziemy w rozdziale dotyczącym przekształceń morfologicznych.

Wspomniany wcześniej (por. il. 5.15) filtr morfologiczny okazał się mało skuteczny przy detekcji krawędzi obrazu testowego bez szumu (il. 5.36a). Przede wszystkim podczas detekcji mocno rozmytych krawędzi (prawa kolumna il. 5.36a) filtr ten dał szereg równoległych krawędzi o podobnej wartości. W przypadku zastosowania większej matrycy tego filtru dałby on bardzo grube krawędzie. Takie rozwiązanie okazało się natomiast pomocne podczas próby detekcji krawędzi w obrazie testowym z szumem (il. 5.36b). W tym przypadku niezbędna jednak będzie dalsza obróbka, aby końcowy wynik był akceptowalny.



Il. 5.37. Detekcja krawędzi w obrazie z obiektami różniącymi się teksturą: a) powiększony fragment z widocznymi pojedynczymi pikselami, b) cały obraz testowy, c) wynik filtracji morfologicznej, d) filtracja Roberta, e) filtracja Roberta obrazu (d), f) filtracja Prewitt, g) filtracja Sobela, h) wynik działania filtra rankingowego, i) filtr uśredniający o matrycy 6x6

Na zakończenie omawiania filtrów do detekcji krawędzi zajmiemy się kolejnym obrazem testowym, który posiada obiekty różniące się teksturą (il. 5.37a). W części obrazu piksele tworzą szachownicę, co daje wrażenie 50% szarości, natomiast w pozostałej części jedynie co czwarty piksel jest czarny. Użycie filtra morfologicznego nie daje dobrych wyników, gdyż w obu obszarach wartości po filtracji maksymalnej i minimalnej są identyczne. Można próbować innego wariantu filtracji morfologicznej, polegającego na odjęciu oryginału od wyniku filtracji maksymalnej. Prowadzi to do dość zaskakującego wyniku, którym jest negatyw obrazu wyjściowego (il. 5.37c). Z kolei filtracja Robertsa, ze względu na zgodność wielkości matrycy filtra z wielkością powtarzalnego elementu analizowanego obrazu, daje w wyniku obraz szary (il. 5.37d). W przypadku tak łatwego do detekcji obrazu powtórne zastosowanie filtracji Robertsa daje dobry wynik (il. 5.37e) z artefaktem w postaci czarnych linii na dwóch brzegach obrazu. Podobny efekt daje filtracja Prewitt (il. 5.37f). Również filtracja Sobela daje bardzo podobny wynik (il. 5.37g). Obraz samych krawędzi Sobela zawiera trochę artefaktów, widocznych w skali pojedynczych pikseli, ale można je łatwo zlikwidować. Natomiast nie mamy żadnych linii w pobliżu brzegów obrazu.

Testowy obraz z il. 5.37b jest dobrym przykładem potencjalnego zastosowania filtracji rankingowej. Jeżeli weźmiemy filtr rankingowy o matrycy 2×2 i wybierzemy drugą pozycję w rankingu, to otrzymamy bardzo łatwy do detekcji krawędzi obraz szary o wysokim kontraście (il. 5.37h). Dobry wynik, w postaci obrazu szarego łatwego do detekcji krawędzi, daje też wykorzystanie zwykłego filtra uśredniającego (*box filter*). Wystarczy matryca 2×2 . Wykorzystanie większej matrycy (il. 5.37i) daje wprowadzić rozmycie, ale za to gładkie i łatwe do detekcji krawędzie.

Przeanalizowaliśmy dużą liczbę filtrów do detekcji krawędzi oraz poddaliśmy tej detekcji różne obrazy. Podsumowując tę część rozdziału, można stwierdzić, że każde z analizowanych rozwiązań sprawdza się bardzo dobrze w pewnych warunkach, a jednocześnie okazuje się bezużyteczne w przypadku innych obrazów. Dlatego trzeba wybór techniki detekcji krawędzi dopasować do konkretnego problemu lub zestawu analizowanych obrazów.

Przejdziemy teraz do bardzo ważnego zagadnienia, jakim jest wyostrażanie obrazów. Praktycznie każdy obraz drukowany w kolorowych czasopismach jest wcześniej wyostrażany. Podobnie dzieje się z obrazami rejestrowanymi przez smartfony i aparaty cyfrowe. W tych urządzeniach obraz jest najczęściej wyostrażany bez informowania o tym użytkownika. Poza tym, często sami wyostrażamy obrazy, używając oprogramowania graficznego. Ale czy zastanawiamy się, jak to działa? Poznamy teraz jeden z najprostszych sposobów wykonania takiego przekształcenia. Weźmy pod uwagę filtr liniowy o matrycy zgodnej z tabelą 5.3.

Tabela 5.3. Przykład matrycy filtra wyostrzającego

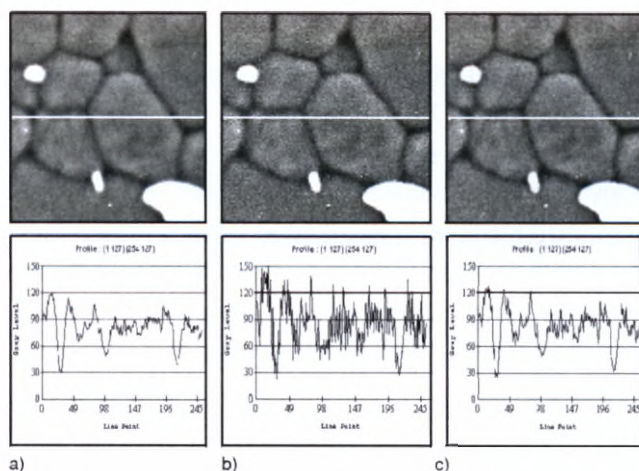
-1	0	-1
0	5	0
-1	0	-1

Konstrukcja tego filtra nie jest przypadkowa. Suma współczynników wynosi 1, co zapewnia brak jakichkolwiek zmian w tych obszarach obrazu, w których wszystkie punkty mają te same wartości, np. x . W wyniku działania naszego filtra w obrazie nie zajdą żadne zmiany, bo $5x - 4x = x$.

Jeżeli jednak jeden, środkowy punkt będzie dwa razy jaśniejszy od otoczenia (czyli o wartości $2x$), to po przeprowadzeniu filtracji jego wartość będzie aż sześciokrotnie większa. Pozostałe punkty będą miały wartości takie same jak przed filtracją, czyli x . Wyjątkiem będą te punkty, których położenie odpowiada narożnym komórkom tabeli (z wartościami -1). Te punkty będą miały wartość zero. Zatem filtr spowoduje lokalnie zwiększenie kontrastu (różnicy wartości sąsiadujących punktów), co nasz wzrok odbierze jako wyostrenie obrazu. Wyjaśnimy jeszcze, dlaczego część współczynników naszego filtra wyostrzającego ma wartości zerowe. Można byłoby zastosować filtr o środkowej wartości równej 9 oraz wartościach -1 dla pozostałych komórek. Taki filtr byłby jednak bardzo podobny do filtra uśredniającego (8 z 9 komórek miałyby tę samą wartość), co osłabiłoby efekt wyostrenia. Dodatkowo, nasz wzrok z łatwością wykrywa pionowe i poziome linie. Dlatego, aby wzmocnić i wyostrzeć krawędzie przebiegające skośnie, w macierzy omawianego filtra wartości -1 przypisano komórkom usytuowanym na przekątnych. Dla porównania, suma współczynników w macierzach liniowych filtrów przeznaczonych do detekcji krawędzi wynosiła 0.

Powyższa analiza filtra wyostrzającego prowadzi do nieco zaskakującego wniosku, że jest on mało przydatny w analizie obrazu, gdyż generuje szum (porównaj il. 5.38), który może utrudnić detekcję poszczególnych obiektów. Należy o tym pamiętać podczas prób konstruowania własnych algorytmów analizy obrazu. Naszym celem nie jest bowiem tworzenie ładnie wyglądających obrazów (zostawmy to grafikom komputerowym), lecz wydobywanie z obrazów informacji. Nieco wcześniej pokazaliśmy, że systemy analizy obrazu potrafią sobie poradzić z rozmytymi krawędziami, o ile nie towarzyszy im szum. Skoro wyostrenie obrazu generuje szum, to może utrudnić automatyczną analizę, choć obraz z punktu widzenia zmysłu wzroku będzie poprawiony.

Warto jednak coś wiedzieć o filtrach wyostrzających, bo czasami generowany przez nie szum może paradoksalnie pomóc w detekcji obiektów. Stanie się tak na przykład wtedy, gdy gęstość punktów szumu będzie istotnie różna w różnych obiektach.



Il. 5.38. Przykłady działania filtrów wyostrzających oraz odpowiadające im rozkłady stopni szarości wzdłuż białej, poziomej linii: a) obraz wyjściowy, b) obraz po filtracji wyostrzającej, której towarzyszy silne zaszumienie obrazu, oraz c) obraz po zaawansowanym wyostrzaniu, minimalizującym powstające szumy. Wykresy wartości pikseli w miejscu poziomej białej linii pokazują nieregularne piki pochodzące od szumu, zwłaszcza w obrazie (b)

Omówimy teraz ciekawy sposób filtracji, który w oprogramowaniu do przetwarzania obrazów często występuje pod nazwą *unsharp mask* lub maska wyostrzająca (to polskie tłumaczenie jest raczej niefortunne; lepszym rozwiązaniem byłoby zapewne określenie „zaawansowane wyostrzanie”).

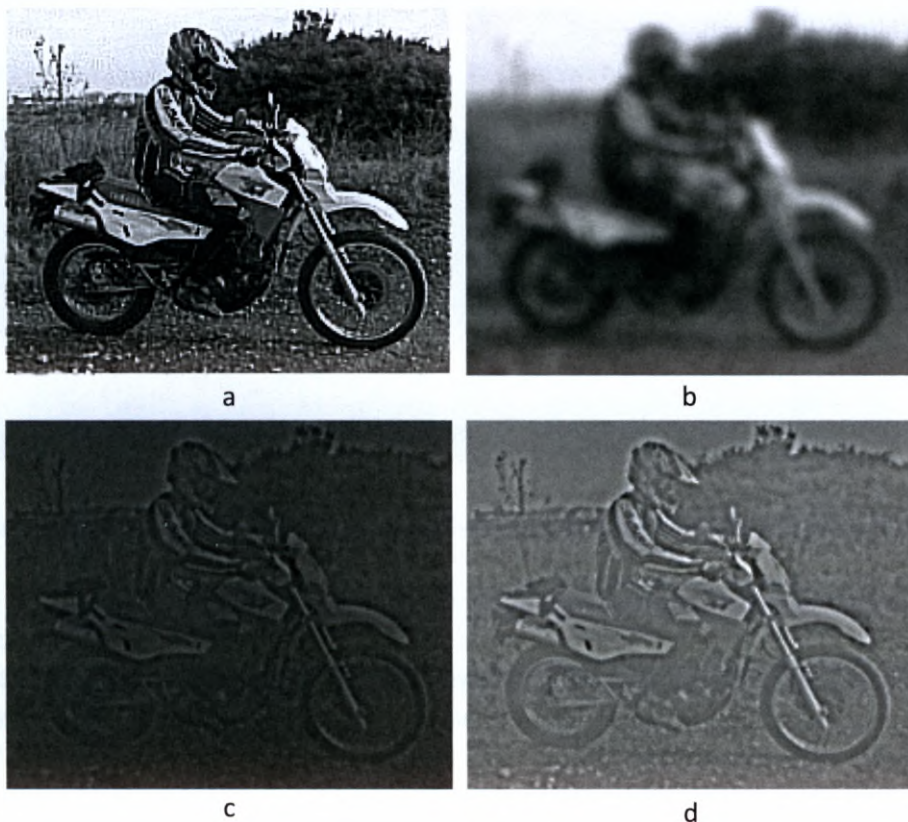
Okazuje się, że można dzięki bardzo sprytnemu, nieco przewrotnemu rozumowaniu opracować taki sposób wyostrzania, który niemal całkowicie eliminuje powstawanie dodatkowego szumu. Wyobraźmy sobie, że mamy do dyspozycji, oprócz naszego wyjściowego obrazu, również obraz idealnie ostry, stanowiący punkt odniesienia dla wyników operacji wyostrzania obrazu. Gdyby udało nam się znaleźć przekształcenie „psujące” obraz, które z idealnie ostrego obrazu tworzy ten, który chcemy skorygować, to można byłoby przekształcenie odwrotne do niego zastosować do naszego obrazu wyjściowego. Można przypuszczać, że takie przekształcenie pozwoliłoby wygenerować różnicę pomiędzy nowym obrazem i obrazem wyjściowym, która byłaby zbliżona do tej pomiędzy obrazem wyjściowym oraz idealnym. Teraz wystarczyłoby dodać tę różnicę do wyjściowego obrazu i otrzymać końcowy obraz, dość bliski ideałowi. Tego typu tok postępowania bywa nazywany metodą wstecznej propagacji. Dla nas ważne jest to, że zastosowany proces wyostrzania praktycznie nie generuje szumu (por. il. 5.38), a końcowy wynik może być rzeczywiście bardzo bliski ostremu oryginałowi.

Pomysł wydaje się być bardzo dobry, ale nie znamy odpowiedniej funkcji. Kluczem do sukcesu jest zatem znalezienie przekształcenia „psującego” obraz. Nie ma tu

jednego, uniwersalnego narzędzia. Najczęściej używa się filtracji rozmywającej obraz (patrz il. 5.39). Poniżej prześledzimy procedurę postępowania.

Najpierw rozmywamy obraz początkowy, przy czym siłę tego rozmycia można regulować przez zmianę wielkości obszaru, z którego wyznaczamy średnią (il. 5.39b). Następnie odejmujemy od siebie obraz początkowy i rozmyty. Wreszcie, na koniec, otrzymaną różnicę dodajemy do obrazu początkowego. Również na tym etapie można regulować siłę przekształcenia. Możemy bowiem dodać całą różnicę lub tylko jej część, na przykład połowę (il. 5.39c i d). Końcowy wynik procesu wyost్రzania pokazuje il. 5.40a.

W oprogramowaniu do przetwarzania obrazu procedura *unsharp masking* ma zwykle dwa parametry: promień i siłę filtra. W naszym przypadku odpowiednikiem promienia jest wielkość matrycy wygładzającej, zaś siłę filtracji regulujemy ułamkiem



Il. 5.39. Wyost్రzanie obrazu: a) obraz z il. 5.9a wyost్రzony filtrem o matrycy zgodnej z tabelą 5.3, b) obraz rozmyty filtrem uśredniającym, c) połowa różnicy obrazu początkowego i rozmytego, d) obraz (c) z zakresem poziomów szarości, rozszerzonym na potrzeby wizualizacji do zakresu 0-255

obrazu pomocniczego, dodawanego do obrazu początkowego. Otrzymany wynik jest bardzo podobny do tego, który oferuje komercyjne oprogramowanie (il. 5.40b). Różnica wyników może być konsekwencją stosowania innych filtrów rozmywających, np. filtracji Gaussa.

Przedstawiona powyżej technika wyost్రzania obrazu, minimalizująca poziom powstających szumów, jest przykładem sprytnego wykorzystania możliwości cyfrowego przetwarzania obrazów. Ma jednak swoje ograniczenia. Jeżeli przesadzi się z doбором parametrów wyost్రzania, to otrzymuje się obraz wyglądający sztucznie, z dodatkowymi krawędziami, które tworzą zakłócające obraz obwódki, nazywane efektem halo.

Liczba możliwych do utworzenia filtrów jest praktycznie nieskończona, choć oczywiście nie wszystkie mają sens i nie wszystkie dają wartościowe wyniki. Można jednak z powodzeniem samemu eksperymentować, co jest łatwe zwłaszcza w przypadku filtrów liniowych.



Il. 5.40. Wyost్రzanie wykorzystujące rozmycie obrazu: a) wynik procedury opisanej w rozdziale, b) wynik z komercyjnego oprogramowania

Przykładem możliwości projektowania dowolnego filtra liniowego może być wspomniane już kilka razy rozmycie gaussowskie. Matryca tego filtra może mieć różne rozmiary i wartości. Przykład pokazany w tabeli 5.4 to tylko jedno z wielu możliwych rozwiązań. Można się domyślić, że ten filtr też jest filtrem uśredniającym, podobnie jak omawiany już *box filter*, ale w przypadku filtra Gaussa punkty położone blisko centrum mają większy wpływ na wynik. Ponieważ suma współczynników matrycy w tabeli 5.4 wynosi 273, to w celu zachowania średniego poziomu szarości filtrowanego obrazu konieczne jest podzielenie ważonej sumy iloczynów współczynników i wartości pikseli przez 273.

Tabela 5.4. Przykład matrycy filtru Gaussa

1/273	1	4	7	4	1
	4	16	26	16	4
	7	26	41	26	7
	4	16	26	16	4
	1	4	7	4	1

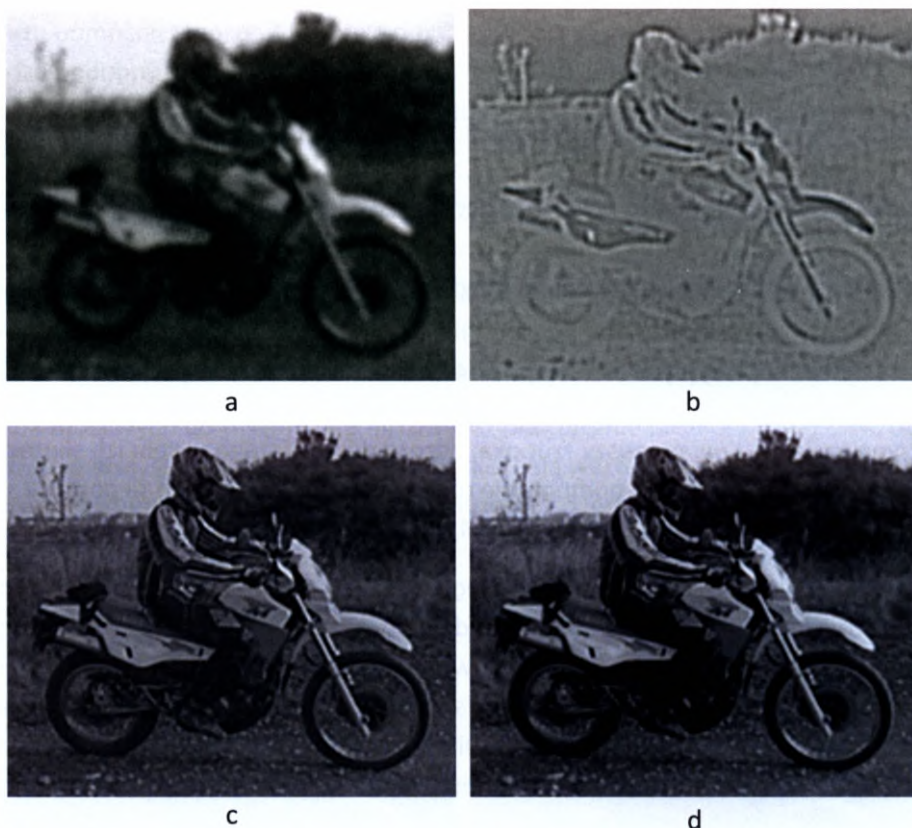
Przykład obrazu po rozmyciu gaussowskim pokazano na il. 5.41a. Różnice w porównaniu z rozmyciem za pomocą średniej arytmetycznej polegają na nieco innym kształcie linii opisujących zmiany poziomów szarości, które w przypadku średniej arytmetycznej mają łagodniejszy przebieg. Różnice pomiędzy tymi filtrami nie są, przynajmniej na tzw. pierwszy rzut oka, bardzo duże. Dlatego w celu ich skutecznej wizualizacji konieczne było pomnożenie różnicy obrazów rozmytych za pomocą średniej arytmetycznej i filtru Gaussa przez 4,5 (il. 5.41b).

Opisywaną już operację *unsharp masking*, czyli zaawansowane wyostrażanie, można prowadzić też w inny sposób. Na przykład do obrazu początkowego dodajemy wynik laplasjanu. Proporcje tych dwóch składowych można zmieniać. Jeżeli zakres wartości laplasjanu skorygujemy tak, aby mieścił się w zakresie od -127 do 128, to uzyskamy wynik pokazany na il. 5.41c. Jeżeli natomiast do wyostrażenia weźmiemy ograniczony do połowy zakres wartości laplasjanu (resztę wartości odrzucamy) i ograniczymy go do 64 stopni szarości, to otrzymamy obraz z il. 5.41d. W porównaniu z il. 5.41c wynikowy obraz wydaje się nieco mniej ostry, ale za to zyskuje na kontraście.

Tabela 5.5. Przykładowa matryca filtru uśredniającego, dająca dobre wyniki w oprogramowaniu do przetwarzania obrazu

1	1	1
1	1	1
1	1	1

Wykonując operacje filtrowania, trzeba pamiętać o zmianie zakresu poziomów szarości. Poszczególne współczynniki w matrycach filtrów są liczbami całkowitymi. Jeżeli filtr będzie miał matrycę taką, jak pokazuje to tabela 5.5, a początkowy zakres poziomów szarości wynosił 0-255, to po filtracji ten zakres zwiększy się do 0-2295 i dla prawidłowego zapisu trzeba będzie użyć 16 bitów zamiast 8, co wystarczyło dla początkowego zakresu. Osoby mające doświadczenie z oprogramowaniem do przetwarzania obrazów mogą być tym zaskoczone, gdyż programy do przetwarzania obrazów przeliczają wartości tak, aby zachować poprzednią głębię bitową, i matryca pokazana w tabeli 5.5 zapewni poprawny wynik. Dla części filtrów, np. laplasjanu lub



Il. 5.41. Filtrowanie obrazów: a) rozmycie gaussowskie, b) 4,5-krotnie wzmocniona różnica rozmycia średnią arytmetyczną i rozmycia Gaussa, c) wyostrenie przez dodanie skompresowanego do 8-bitowego zakresu laplasjanu 7x7 do obrazu początkowego, d) wyostrenie przez dodanie 50% zakresu laplasjanu 7x7, skompresowanego do zakresu 0-64, do obrazu początkowego

wykrywających krawędzie, niezbędne jest umożliwienie zapisu liczb ujemnych. Jeżeli tego nie uwzględnimy, to wyniki mogą być zaskakujące. Gdy wartość liczby wychodzi poza zakres 0-255 w obrazie 8-bitowym, to jest ona dzielona przez 255 i w obrazie zapisywana jest reszta z dzielenia (il. 5.42).

Można eksperymentować z różnymi macierzami filtrów. Niekiedy jest to wręcz niezbędne przy poszukiwaniu algorytmów do rozwiązania złożonych problemów.

Maczyca filtru może być liniowa, jak pokazano to z lewej strony tabeli 5.6. Wykorzystanie takiej maczyca daje obraz niejako poruszony w kierunku poziomym, co daje złudzenie ruchu (il. 5.43a). Z kolei maczyca w postaci dwóch równoległych, oddalonych o siebie odcinków (prawa strona tabeli 5.6) da nam wrażenie nałożenia dwóch obrazów (il. 5.43b).



Il. 5.44. Filtry definiowane przez użytkownika. Opis w tekście, gdzie pokazano matryce o rozmiarach 5x5. Obrazy pokazane na tym rysunku mają rzeczywisty rozmiar 321x265 pikseli, a przekształcenia wykonano filtrem o rozmiarze 9x9

tych wartości: w lewej matrycy znajdują się one na środkach boków, a w prawej zajmują pozycje w narożnikach. Wyniki filtracji takimi filrami, ale o rozmiarze zwiększonym do 9x9, pokazują odpowiednio il. 5.44a oraz 5.44b. Zwiększenie rozmiarów matrycy miało na celu wyraźniejsze pokazanie efektów ich użycia. Filtracje pokazane na il. 5.44 przeprowadzono za pomocą oprogramowania do przetwarzania obrazu. W przypadku użycia programu do analizy obrazu konieczne byłoby podzielenie wyniku przez 4 w celu zachowania dla poziomów szarości zakresu 0-255.

Tabela 5.7. Matryce filtrów prowadzących do wrażenia zwielokrotnienia obrazu

0	0	1	0	0
0	0	0	0	0
1	0	0	0	1
0	0	0	0	0
0	0	1	0	0

1	0	0	0	1
0	0	0	0	0
0	0	0	0	0
0	0	0	0	0
1	0	0	0	1

Różne przykłady przedstawione w tym rozdziale ilustrują prawdziwe bogactwo filtrów oraz możliwości ich projektowania, a także tworzenia złożonych przekształceń, realizujących wielostopniowe filtrowanie. Dobór odpowiedniego filtra do konkretnego problemu wciąż jest zagadnieniem z pogranicza sztuki i techniki. Na początku jest to dość trudne, ale analiza wielu dostępnych rozwiązań demonstracyjnych ułatwia zdobycie odpowiedniego doświadczenia. Szczegółowa analiza sposobu działania poszczególnych filtrów, przedstawiona w ramach tego rozdziału, również miała na celu ułatwienie nauki ich umiętnego stosowania.

Czasami można spotkać się z określeniem filtrów jako dolnoprzepustowe lub górnoprzepustowe. Filtry dolnoprzepustowe to takie, które zostawiają w obrazie elementy wolniej zmieniające się, a usuwają elementy wykazujące częste i szybkie zmiany. Można je różnie definiować, ale ogólnie rzecz biorąc, filtrami dolnoprzepustowymi są filtry uśredniające i wygładzające obraz. Z kolei filtry górnoprzepustowe są ich przeciwieństwem. Pozostawiają lub wręcz wzmacniają w obrazie te elementy, które szybko się zmieniają, czyli sąsiadujące ze sobą piksele o znacznych różnicach wartości. Klasycznym przedstawicielem filtrów górnoprzepustowych jest laplasjan.

Na tym kończymy nasze rozważania dotyczące podstawowych filtrów. Będziemy do nich wielokrotnie wracali, konstruując algorytmy przetwarzania lub analizy obrazów.

6. Korekcja cienia

Korekcja cienia (ang. *shade correction*) nie ma wbrew pozorom wiele wspólnego z cieniem, jaki dają nam drzewa, budynki lub różne przedmioty, oświetlone wiązką światła. Jest to trochę nietrafna nazwa procesu usuwania skutków (czyli korekcji) nierównomiernego oświetlenia.

Aby lepiej zrozumieć źródła tego problemu, musimy poświęcić nieco uwagi sposobowi działania ludzkiego zmysłu wzroku. Nasze gałki oczne są w ciągłym ruchu i skupiamy się na stosunkowo niewielkim obszarze naszego pola widzenia.

Wyobraźmy sobie, że mieszkamy w hotelu w górach i rano, po wyjściu z budynku, zachwyciliśmy się widokiem gór. Chcąc utrwalić tę chwilę, sięgamy po aparat i wykonujemy zdjęcie. Jeżeli wybierzemy domyślne ustawienia aparatu, to prawdopodobnie otrzymamy obraz taki jak z lewej strony il. 6.1 (HDR wyłączony). Efekt nieco rozczarowujący, bo przecież pamiętamy, że w zacienionej części otoczenia hotelu wcale nie było tak ciemno, że prawie nic nie widać. Wyjaśnienie jest bardzo proste. Aparat fotograficzny nie był wcale uszkodzony i prawidłowo zarejestrował obraz. Gdy nasz zmysł wzroku koncentruje się na widoku gór, to pomija analizę informacji z bliskiego otoczenia hotelu i nie przeszkadza mu to, że do oka dociera stamtąd niewiele światła, co zarejestrował aparat fotograficzny. Jeżeli skupimy naszą uwagę na otoczeniu hotelu, to oko szybko dostosowuje się do nowych warunków – zwiększa się otwór źrenicy i dzięki temu możemy dostrzec szczegóły w cieniu. Widoczny kątem oka obraz gór jest pomijany w analizie i nie przeszkadza nam to, że w tym



HDR wyłączony



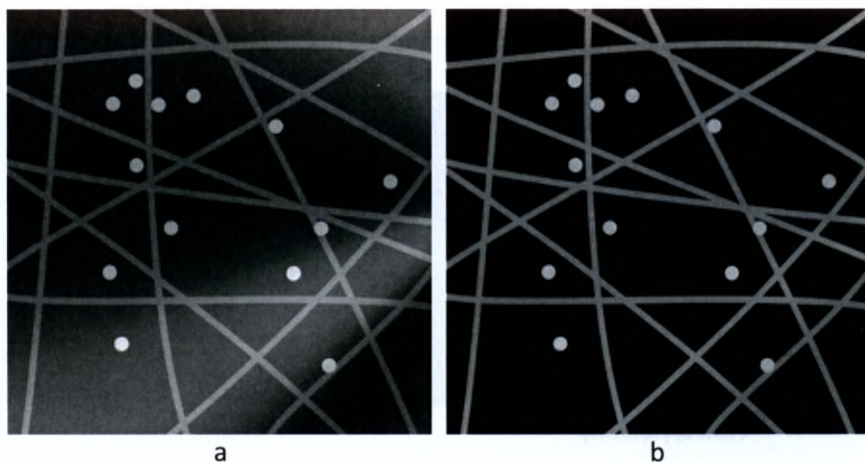
HDR włączony

Il. 6.1. Różnice w rejestracji szczegółów w zacienionym oraz dobrze oświetlonym obszarze zdjęcia. Opis w tekście

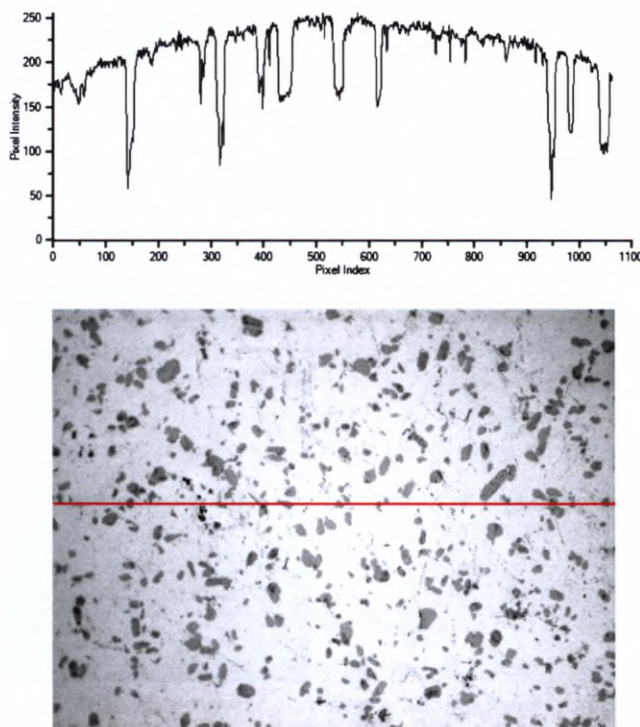
miejscu oko widzi głównie jasną plamę. Te dwa tryby pracy naszego oka dają różne informacje, które są łączone w mózgu i tworzą wirtualny obraz otoczenia, w którym dostrzegamy szczegóły zarówno w miejscach jasnych, jak i ciemnych.

Podobny proces można przeprowadzić również w bardziej zaawansowanych aparatach cyfrowych. Wybierając tryb HDR (*High Dynamic Range*) wykonujemy bezpośrednio po sobie dwa zdjęcia. Jedno ma naświetlone prawidłowo miejsca najjaśniejsze, a drugie miejsca najciemniejsze. Następnie oprogramowanie aparatu składa te dwa obrazy w jeden i otrzymujemy wynik taki, jak pokazuje to prawy obraz na il. 6.1 (HDR włączony). Ten tryb pracy aparatu pozwala zarejestrować obraz odpowiadający temu, co zapamiętał nasz mózg. Oczywiście, mamy tu pewne ograniczenia. Ze względu na potrzebę wykonania dwóch zdjęć nie można w ten sposób rejestrować szybko zmieniających się obiektów, a ponadto mogą pojawić się zmiany w odwzorowaniu kolorów. Na przykład kolor gór i nieba na zdjęciu z włączonym HDR jest mniej efektowny niż na zdjęciu bez HDR. Z drugiej strony jest bliższy temu, co rejestruje nasz zmysł wzroku.

Dzięki takiej lokalnej analizie nasz zmysł wzroku nie ma problemu z wykryciem linii na il. 6.2a, gdyż lokalnie widzi je jako elementy jaśniejsze od otoczenia. Komputer ma z tym problem, bo stosuje globalnie ten sam próg czułości na całym obrazie. Oczywiście możliwe jest wprowadzenie dynamicznie zmieniającego się progu czułości, ale jest to zadanie dość trudne i zasady wyznaczania odpowiedniego progu muszą być różne dla różnych obrazów. Dlatego zazwyczaj lepiej jest przeprowadzić korekcję cienia i analizować obraz skorygowany, taki, jak pokazuje to il. 6.2b.



Il. 6.2. Testowy obraz przed (a) i po (b) korekcji cienia

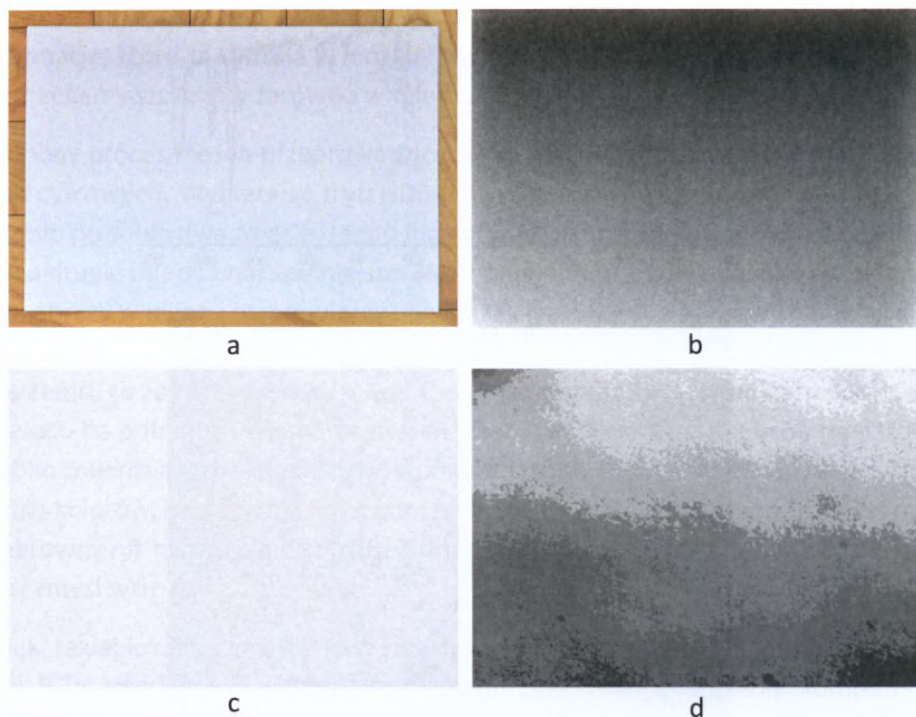


Il. 6.3. Mikroskopowy obraz struktury oraz rozkład poziomów szarości wzdłuż czerwonej linii

Z potrzebą korekcji cienia spotykamy się w praktyce dosyć często. Klasycznym przykładem są obrazy zarejestrowane podczas badań z użyciem mikroskopii świetlnej (il. 6.3), gdzie z powodu różnic w długości drogi przebiegu promieni świetlnych dla centralnej i brzegowej części obrazu występują nieraz znaczne różnice jego jasności. Obraz jest wyraźnie jaśniejszy w centralnej części. Binaryzacja takiego obrazu bez korekcji prowadzi do ewidentnych błędów.

Jeżeli mamy okazję pracować z mikroskopami świetlnymi najnowszej generacji, to ich optyka oraz konstrukcja oświetlaczy sprawiają, że obraz jest niemal pozbawiony nierównomierności oświetlenia. Jednak podczas detekcji bardzo subtelnych zmian w obrazach, również tych pochodzących z najnowszych urządzeń, obrazy te mogą wymagać korekty.

Okazuje się, że w praktyce trudno jest osiągnąć jednolite oświetlenie pochodzenia naturalnego. Aby się o tym przekonać, przeanalizujemy zdjęcia białej kartki papieru. Aby zapewnić dobrą jakość materiału do analiz, wykorzystano aparat fotograficzny Nikon D7200 z profesjonalnym obiektywem AF-S Nikkor 17-55 mm 1:2,8 G ED.



- Il. 6.4. a) kartka papieru sfotografowana w zaciemnionym miejscu w pokoju, z dala od okna,
b) poziomy szarości na zdjęciu kartki papieru z il. (a), rozciągnięte do zakresu 0-255,
c) kartka papieru sfotografowana w zaciemnionym miejscu na wolnej przestrzeni,
d) poziomy szarości obrazu z il. (c) rozciągnięte do zakresu 0-255, przedstawiające geometrię nierównomiernego oświetlenia

Pierwsza próba to wykonanie zdjęcia w pokoju podczas pochmurnego dnia, zapewniającego bardziej rozproszone światło. Kartka papieru formatu A4 (210x297 mm) została umieszczona na podłodze za przeszkodą wysokości 0,5 m, umieszczoną ok. 3 m od okna, co zapewnia oświetlenie tylko za pomocą rozproszonego światła. Odpowiednie zdjęcie przedstawia il. 6.4a.

Okazuje się, że kartka wcale nie jest jednolicie szara. Najjaśniejsze miejsca mają poziom szarości bliski białości (235), ale najciemniejsze miejsca wykazują zaledwie poziom szarości 193. Zatem najjaśniejsze miejsca mają poziom szarości o ponad 20% większy od miejsc najciemniejszych. Po znormalizowaniu obrazu kartki papieru, pokazanej na il. 6.4a, otrzymujemy wizualizację zróżnicowania poziomów szarości, pokazaną na il. 6.4b.

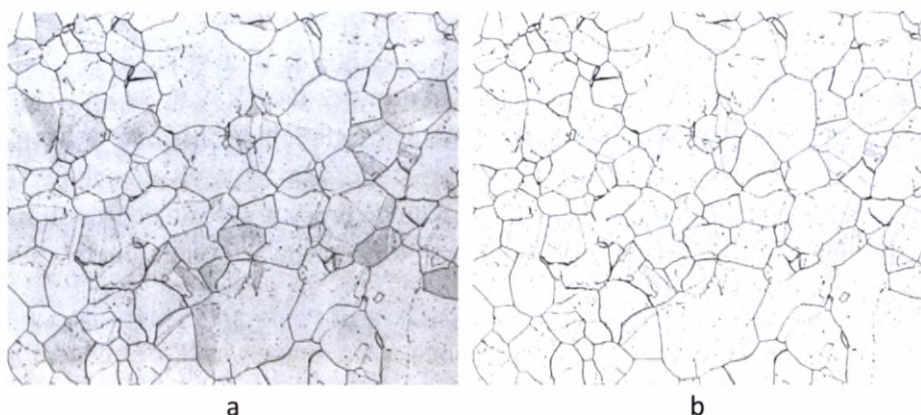
O ile pokój jest mimo wszystko przestrzenią o ograniczonej powierzchni, ze źródłem światła skupionym na stosunkowo niewielkiej powierzchni okna, o tyle zaciemniona przestrzeń za domem powinna dać oświetlenie bardziej równomierne. Odpowiednie

testy przeprowadzono w godzinach popołudniowych, gdy słońce było już schowane za drzewami na wzgórzu za domem. Kartkę papieru umieszczono na otwartej przestrzeni, ok. 2 m od ściany domu, który znajdował się od strony słońca i zapewniał dodatkową ochronę przed bezpośrednim światłem słonecznym. Na il. 6.4c pokazano sfotografowaną w tych warunkach kartkę papieru.

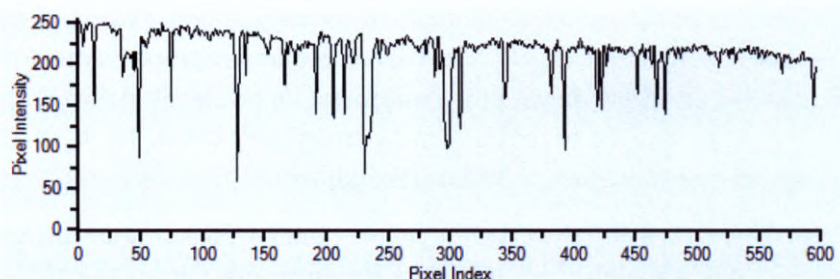
Jeżeli patrzymy na il. 6.4c, to wydaje się, że tym razem uzyskaliśmy już równomierne oświetlenie. Bliższa analiza wykazała jednak zróżnicowanie odcieni szarości w zakresie 209-219. Zatem najjaśniejszy punkt był jaśniejszy od najciemniejszego o blisko 5%. Mapę rozkładu poziomów szarości obrazu z il. 6.4c przedstawia il. 6.4d.

Podsumowując powyższe doświadczenia, widać, że zapewnienie dokładnie równomiernego oświetlenia jest trudne i dlatego korekta cienia wciąż ma rację bytu. W praktyce najprostszym i do tego tanim rozwiązaniem, które zapewnia niemal perfekcyjnie równomierne oświetlenie, jest wykorzystanie skanera. Należy jednak pamiętać, że tylko niektóre obiekty mogą być skanowane.

Najczęściej procedury korekcji cienia są przedstawiane na obrazach z jasnym tłem, zawierających stosunkowo drobne obiekty (jak na il. 6.3) lub cienkie, ciemne linie (jak na il. 6.5a). Korekta może obejmować zarówno globalne (czyli obejmujące cały obraz) nierównomierne tło (il. 6.3 i 6.5a), jak i miejscowe zaciemnienia, np. ziarna, wyraźnie ciemniejsze od otaczających je sąsiadów (porównaj il. 6.5a i 6.5b). Nasz zmysł wzroku zwykle nie dostrzega globalnie nierównomiernego tła, choć takie zjawisko występuje na il. 6.5a. Jeżeli jednak prześledzi się zmiany poziomów szarości wzdłuż przekątnej (od lewego górnego do prawego dolnego rogu), to widoczny jest systematyczny spadek poziomu szarości, który osiąga w skali całego obrazu ok. 20% (il. 6.6).



Il. 6.5. Obraz początkowy (a) i po korekcie cienia (b). Widoczne jest wyrównanie tła również w obrębie pojedynczych ziaren



Il. 6.6. Zmiany poziomów szarości na il. 6.5a wzdłuż przekątnej (od lewego górnego do prawego dolnego rogu)

Możliwe jest również korygowanie rzadziej spotykanych obrazów, w których tło jest ich najciemniejszym elementem. W takim przypadku należy odpowiednio zmodyfikować przedstawione w dalszej części algorytmy. Gdyby stanowiło to trudność, wystarczy przeprowadzić korektę cienia na negatywie obrazu i po zakończeniu korekty wykonać powtórnie negatyw, tym razem wynikowego obrazu.

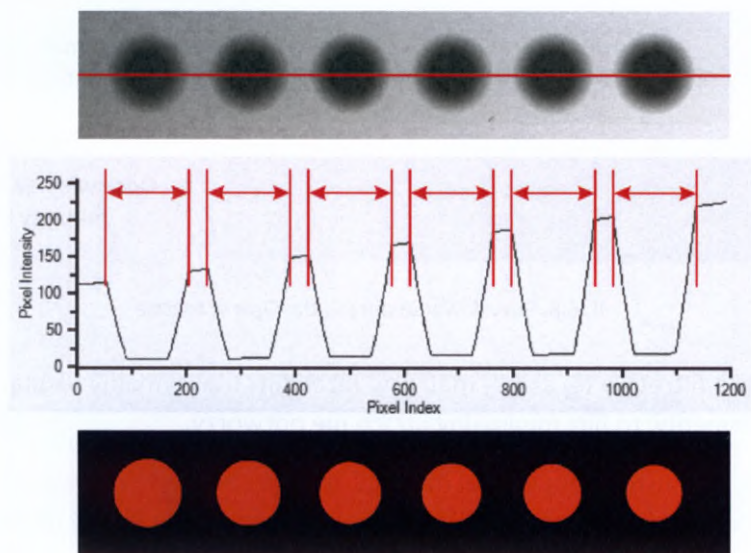
Metody korekty cienia można podzielić na dwie grupy:

- metody wykorzystujące obraz nierównomiernego tła, który może być uzyskany na podstawie przetwarzania korygowanego obrazu lub zostać zarejestrowany oddzielnie,
- metody wykorzystujące analizę częstotliwościową obrazu.

Idea pierwszej grupy metod jest konsekwencją następującego rozumowania (zakładamy, że omawiamy obrazy z jasnym tłem): gdyby obraz był idealnie zarejestrowany, to jego tło byłoby w każdym miejscu białe. A skoro tło nie jest białe, to sprawdźmy, co trzeba byłoby z nim zrobić, żeby stało się białe, i zastosujmy ten sposób postępowania do całego obrazu.

Skąd jednak wziąć tło? W przypadku badań prowadzonych za pomocą mikroskopu świetlnego można uzyskać obraz tła, rejestrując światło przechodzące przez matową płytkę (mikroskop do pracy w świetle przechodzącym) lub światło odbite od wypolerowanej gładkiej powierzchni (mikroskop do pracy w świetle odbitym). Trzeba potem tylko pamiętać, żeby całą serię obrazów rejestrować z tymi samymi parametrami ekspozycji co obraz tła.

Obraz tła można również uzyskać za pomocą przetwarzania analizowanego obrazu. Jest to łatwe zwłaszcza w przypadku tych obrazów, w których tło zajmuje większość powierzchni, a obiekty znajdujące się na obrazie nie mają zbyt dużych rozmiarów (tak jak w przypadku il. 6.3 i 6.5a).



Il. 6.7. Testowy obraz do analizy metod korekty cienia (góra), profil poziomów szarości wzdłuż czerwonej linii (środek) oraz wynik binaryzacji obrazu testowego (dół)

Algorytmy oraz wyniki korekty cienia będziemy śledzili na przykładzie testowego obrazu, pokazanego w górnej części il. 6.7. Wyniki poszczególnych metod będą przedstawiane w analogiczny sposób, tzn. u góry obraz, w środku profil ze zmianami poziomów szarości, a na dole wynik binaryzacji.

W testowym obrazie warto zwrócić uwagę, że niezależnie od poziomu szarości, każdy obiekt ma taką samą szerokość na profilu (strzałki wymiarowe zaznaczono na profilu kolorem czerwonym). Daje to od razu istotną wskazówkę, dotyczącą jakości procedury korekcyjnej – w wyniku detekcji powinniśmy otrzymać zestaw kół o tej samej średnicy, jeżeli próg detekcji będzie zgodny z przyjmowanym za optymalny progiem na poziomie średniej arytmetycznej najjaśniejszego i najciemniejszego punktu binaryzowanego obiektu. W przypadku binaryzacji obrazu testowego (il. 6.7) wykryte koła mają różne średnice, malejące od lewej do prawej strony.

Obraz tła można uzyskać przez złożenie filtracji maksymalnej oraz minimalnej. W przypadku operacji morfologii matematycznej, które będą omawiane w dalszej kolejności, takie połączenie przekształceń nazywamy zamknięciem. Filtracja maksymalna musi być wykonana z użyciem filtra o tak dużej matrycy, żeby zniknęły wszystkie ciemne elementy (i właśnie dlatego ta operacja sprawdza się bardzo dobrze w przypadku niewielkich lub cienkich obiektów). Tak powstały obraz jest jednak wyraźnie jaśniejszy od początkowego tła. Dlatego kolejnym krokiem jest filtracja

Zbyt mała wielkość
matrycy filtraOdpowiednia wielkość
matrycy filtra

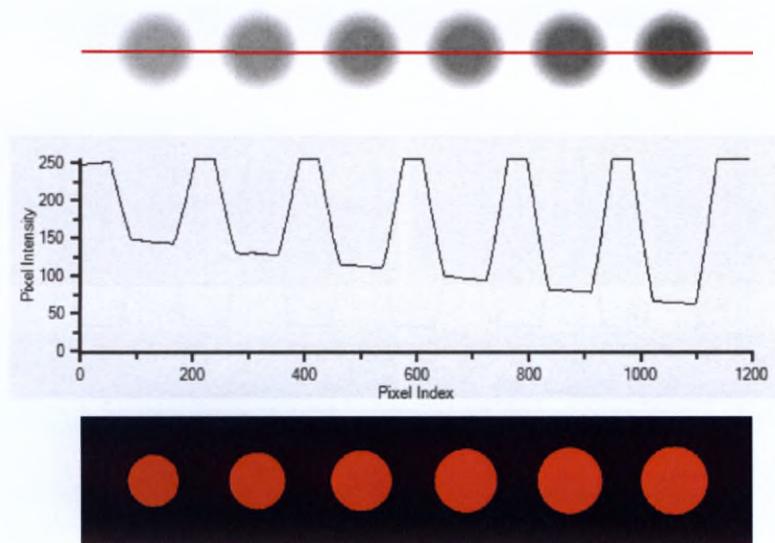
Il. 6.8. Pozyskiwanie obrazu tła. Opis w tekście

minimalna z filtrem o tej samej matrycy. Jeżeli filtr maksymalny usunął wszystkie ciemne elementy, to filtr minimalny już ich nie odtworzy.

Jeżeli użyjemy filtrów o zbyt małej matrycy, to nie odtworzymy tła z dostateczną jakością i pozostaną w nim wyraźne ślady pierwotnego obrazu (górny obraz na il. 6.8). Zwiększenie wielkości lub czasami kształtu matrycy filtra doprowadza w końcu do satysfakcjonującego wyniku (dolny obraz na il. 6.8).

Dobranie odpowiedniej wielkości matrycy filtra nie stanowi problemu w analizie interaktywnej, gdzie możemy to łatwo zrobić metodą prób i błędów. W przypadku potrzeby ustalenia algorytmu w pełni automatycznego należy znaleźć sposób na ustalenie odpowiedniej wielkości matrycy filtra, gdyż ze względu na minimalizację błędów przetwarzania trzeba używać matryc możliwie jak najmniejszych. Jednym ze sposobów jest wykonanie automatycznej binaryzacji. Wielkość matrycy filtra powinna być nieco większa niż największa szerokość obiektu w obrazie binarnym. Można również wykonać szereg filtracji z różnymi wielkościami matryc filtrów i analizować zakres otrzymanych poziomów szarości. Jeżeli znikną niskie wartości, odpowiedzialne za ciemniejsze obiekty, to prawdopodobnie obraz tła będzie już odpowiedni. Oczywiście wszystkie opisane powyżej rozwiązania trzeba najpierw sprawdzić na odpowiednich obrazach testowych. Przez „odpowiednie” należy rozumieć obrazy tak przygotowane, aby zachowywały newralgiczne cechy badanych obrazów. Bardzo dobrym rozwiązaniem jest wykorzystanie do testów kilku obrazów z serii, która ma być następnie automatycznie analizowana.

Jeżeli mamy już dobry obraz tła (dolny obraz na il. 6.8), trzeba ustalić, ile brakuje nam do tego, żeby tło było idealnie białe. Można to zrobić przez stworzenie obrazu o tej samej wielkości co tło, składającego się wyłącznie z pikseli o wartości 255 (odpowiadającej bieli), i odjęcie od niego obrazu tła. Dokładnie taki sam wynik uzyskamy, tworząc negatyw tła, co w przypadku używania systemu analizy obrazu oznacza zmniejszenie liczby operacji i tym samym przyspieszenie całej analizy.



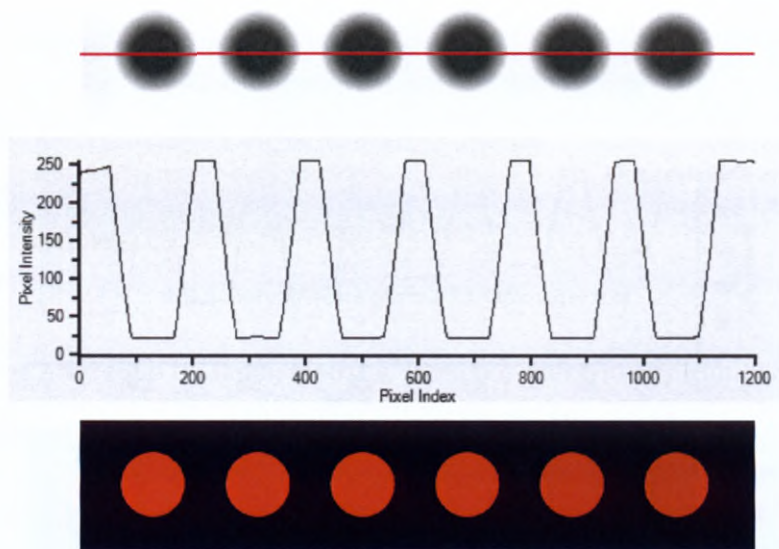
Il. 6.9. Korekcja cienia z wykorzystaniem dodawania negatywu tła

Teraz wystarczy dodać negatyw tła do początkowego obrazu i dostaniemy obraz z wyrównanym tłem (il. 6.9). Opisana metoda daje dobre wyrównanie tła, ale powoduje też, że w miejscach, w których tło było ciemne, rozjaśniają się najciemniejsze elementy obiektów (patrz: lewa strona profilu na il. 6.9). W konsekwencji uzyskany po takiej korekcie obraz binarny będzie nadal wykazywał koła o różnych średnicach. Różnica w porównaniu z obrazem wyjściowym jest taka, że koło o najmniejszej średnicy jest teraz z lewej strony obrazu. Można to zatem podsumować stwierdzeniem, że wpadliśmy z deszczu pod rynnę.

Spróbujmy problem rozwiązać inaczej. Jeżeli wiedzielibyśmy, ile razy tło jest zbyt ciemne w porównaniu z obrazem idealnie białym, to moglibyśmy analizowany obraz pomnożyć przez odpowiednie wartości. Tło stałoby się białe, a ciemne elementy pozostałyby ciemne, bo zamiast dodawać różnicę pomiędzy tłem i bielą, mnożylibyśmy niewielką wartość poziomu szarości ciemnych elementów przez odpowiedni współczynnik (zwykle o wartości nieprzekraczającej ok. 1,5), co w efekcie dalej dawałoby niewielką wartość.

Odpowiedni algorytm jest następujący:

- Wyznaczamy obraz tła.
- Tworzymy obraz idealnie biały o tym samym rozmiarze.
- Biały obraz dzielimy przez obraz tła, przy czym należy pamiętać, że w wynikowym obrazie musimy przewidzieć wartości zmiennoprzecinkowe, a nie liczby całkowite.
- Korygowany obraz mnożymy przez wynik opisanego powyżej dzielenia obrazów.



Il. 6.10. Korekcja cienia z wykorzystaniem dzielenia i mnożenia

Otrzymany wynik (il. 6.10) jest niemal idealny, a w wyniku binaryzacji otrzymujemy zestaw kół o tej samej średnicy. W systemie Aphelion występuje procedura o nazwie *ShadingCorrection*, która jest opisana następująco:

$$\text{Final image} = 255 * (\text{Live image} - \text{Black image}) / (\text{White image} - \text{black image})$$

Obraz *Black image* można pominąć i wtedy otrzymujemy:

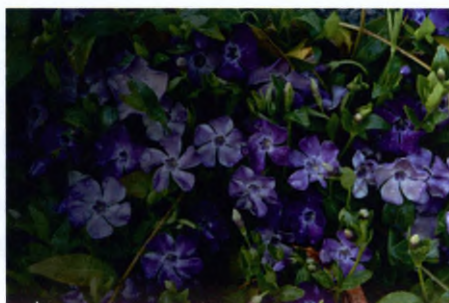
$$\text{Final image} = 255 * \text{Live image} / \text{White image}$$

Niestety, opis ten jest nieco mylący, gdyż *White image* może sugerować obraz biały. Jeśli tak zrobimy, to oczywiście procedura nie wprowadzi żadnych zmian. Trzeba sięgnąć do plików pomocy, aby doczytać, że chodzi o zarejestrowany za pomocą mikroskopu obraz samego tła (o takim rozwiązaniu mówiliśmy nieco wcześniej). Zamiast *White image* lepiej byłoby napisać *Background image*, ewentualnie *Bright background image*. Opisywana procedura daje w wyniku dokładnie to samo, co opisaliśmy jako korekcję za pomocą dzielenia i mnożenia.

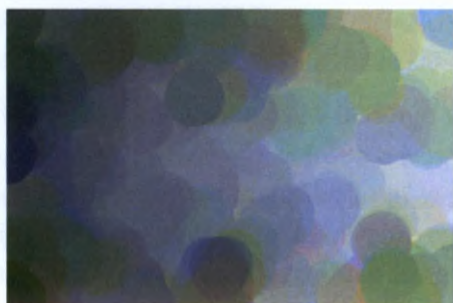
Zobaczmy teraz, czy opisywaną procedurę uda się zastosować do obrazu, który nie ma jednolitego tła, a przedstawia silnie zróżnicowane obiekty (il. 6.11). To, że obraz na il. 6.11a jest obrazem kolorowym, nie stanowi problemu. Jak dowiemy się z rozdziału poświęconego analizie obrazów kolorowych, wystarczy przetworzyć tak samo wszystkie składowe RGB i następnie złożyć z nich wynikowy obraz.



a



b

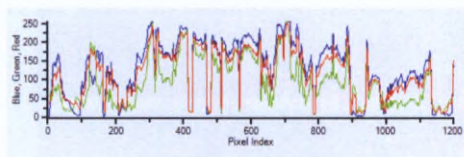


c

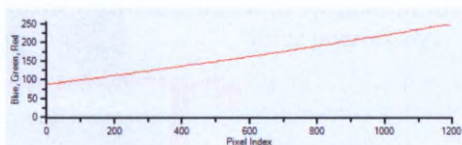


d

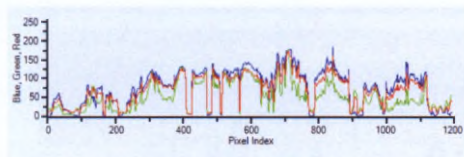
Il. 6.11. Korekta cienia silnie niejednorodnego obrazu. Opis w tekście



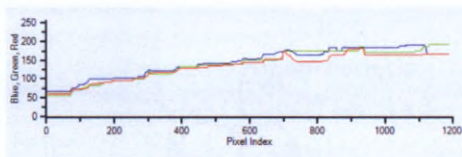
a) Obraz początkowy



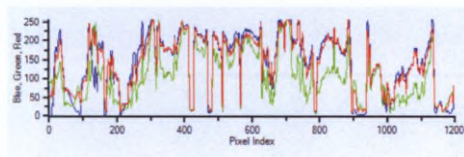
b) Założony przebieg cienia (tła)



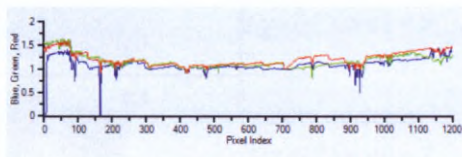
c) Obraz zmodyfikowany cieniem



d) Rekonstrukcja tła



e) Obraz skorygowany



f) Obraz skorygowany/początkowy

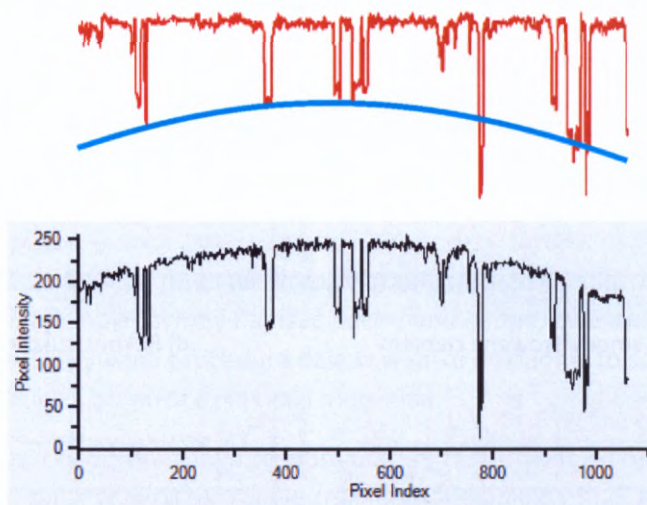
Il. 6.12. Profile wzdłuż poziomej linii przechodzącej przez środek obrazu. Opis w tekście

Wybrany przez nas obraz (il. 6.11a) jest bardzo niejednorodny. Dowodzi tego profil wartości poszczególnych pikseli, wykonany dla poziomej linii przechodzącej przez środek obrazu (il. 6.12a). Trzy linie na profilu, w kolorach czerwonym, zielonym i niebieskim, odpowiadają poszczególnym składowym kolorowego obrazu RGB.

Generujemy teraz obraz zniekształcony przez zmienny poziom jasności – z lewej strony będzie przyciemniony, następnie liniowo będzie stawał się coraz jaśniejszy, aż wreszcie z prawej strony pozostaną wartości z obrazu początkowego. W tym celu mnożymy obraz początkowy przez przygotowany obraz oczekiwanego cienia, czyli maksymalne wartości tła (profil na il. 6.12b), i całość dzielimy przez 255, aby pozostać w zakresie 24 bitów. Uzyskany zmodyfikowany obraz pokazuje il. 6.11b, a jego profil il. 6.12c.

Teraz staramy się odtworzyć obraz tła. W tym celu wykonujemy filtrację maksymalną, a następnie minimalną za pomocą filtrów o dużej matrycy. Aby uzyskać bardziej naturalny efekt, wykorzystujemy matrycę w kształcie koła, a nie kwadratu. Obraz tak wyszukiwanego tła pokazuje il. 6.11c, a jego profil il. 6.12d. Jeżeli porównamy il. 6.12b i 6.12d, to zauważymy, że wygenerowane przez nas tło jest ciemniejsze od założonego w skrajnych obszarach, czyli z lewej i z prawej strony. Można się zatem spodziewać, że w tych rejonach obraz skorygowany będzie jaśniejszy od początkowego.

Końcowy wynik naszych działań, czyli obraz po korekcji, przedstawia il. 6.11d, a jego profil pokazuje il. 6.12e. Zarówno wizualne porównanie obrazów (il. 6.11a i 6.11d),



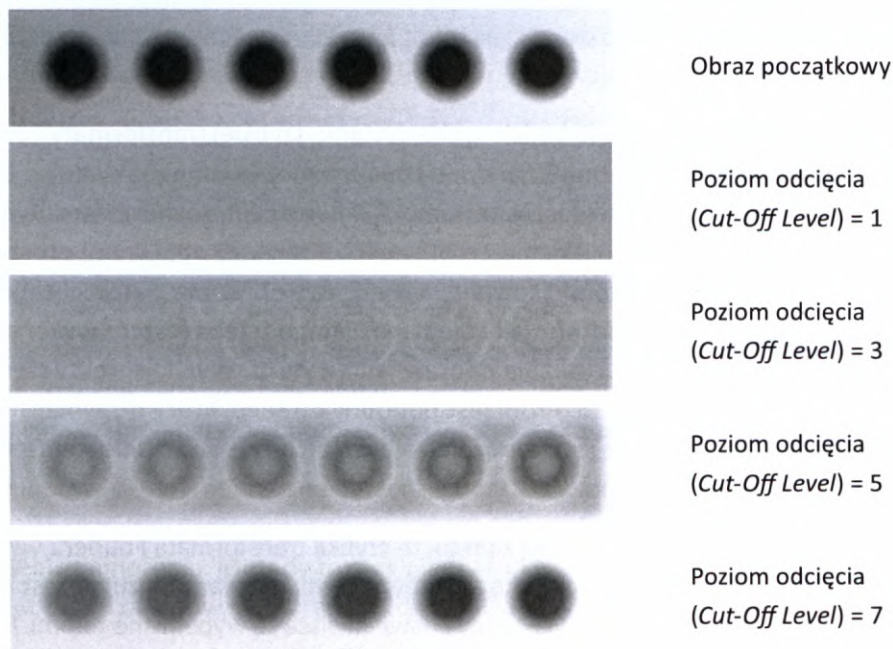
Il. 6.13. Idea interpretacji obrazu jako złożenia funkcji. Tło opisuje funkcja o niskich częstotliwościach (niebieska linia), a zmiany różnych obiektów – funkcja o wysokich częstotliwościach (czerwona linia)

jak i profili (il. 6.12a i 6.12e) pokazuje, że obraz po korekcji jest jaśniejszy od początkowego zarówno z lewej, jak i z prawej strony. Potwierdza to profil obrazu, będącego wynikiem dzielenia obrazu skorygowanego przez początkowy (il. 6.12f).

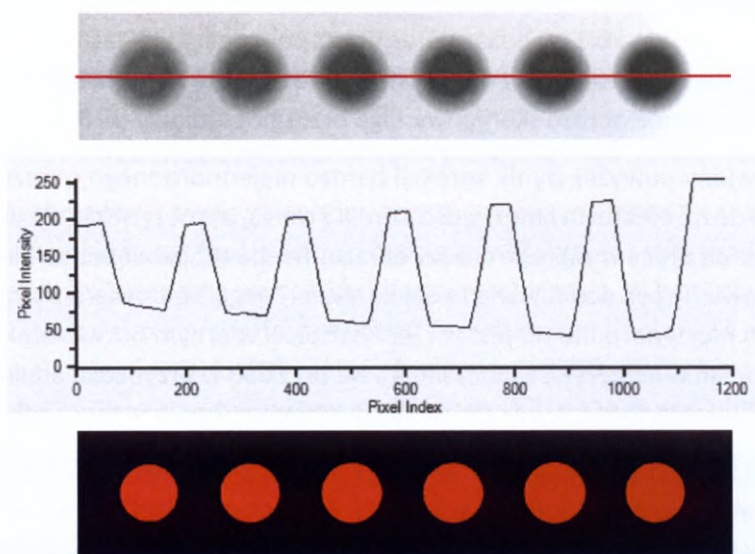
Przedstawiony powyżej wynik korekcji bardzo niejednorodnego obrazu pokazuje w sposób dosyć efektowny możliwości korekty cienia, ale oczywiście nie jest to właściwy obraz do pracy w zakresie analizy obrazu. Trzeba też pamiętać, że nie wszystkie obrazy powinny być poddawane takiej korekcji. Mogą się bowiem zdarzyć obrazy, w których niejednorodne tło jest w rzeczywistości wiernym odtworzeniem rzeczywistych zmian w kolorystyce lub jasności. Na przykład w przypadku analizy struktur materiałów może to być wynik zmian składu chemicznego, a w przypadku dzieł malarskich – efekt zmian grubości lub składników kolorystycznych farby.

Przejdziemy teraz do metod korekcji cienia, wykorzystujących analizę częstotliwościową obrazu. Opierają się one na spostrzeżeniu, że zmiany poziomu szarości można opisać za pomocą funkcji okresowych, a konkretnie sinusa i kosinusa. Ideę takiej interpretacji pokazuje il. 6.13.

W systemie analizy obrazu Aphelion znajduje się filtr o nazwie LowStop, który odcina z obrazu niskie częstotliwości, przy czym można zmieniać poziom odcięcia. Wyniki



Il. 6.14. Wyniki filtracji odcinającej niskie częstotliwości. Opis w tekście

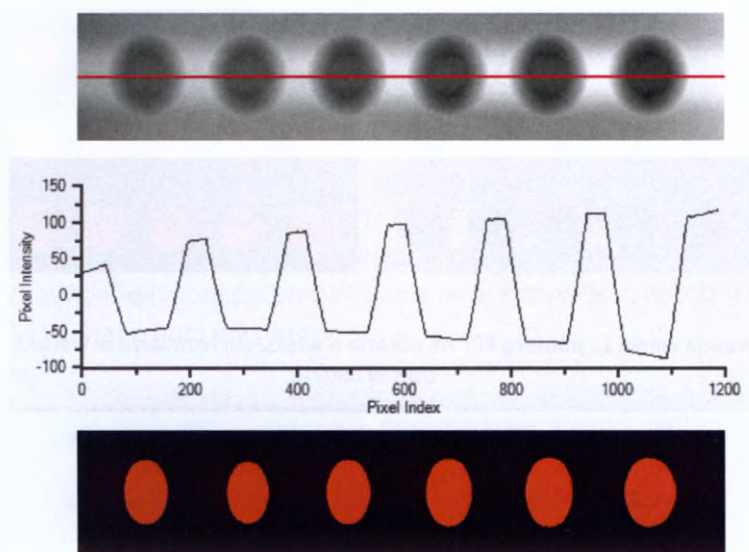


Il. 6.15. Wyniki korekcji cienia za pomocą filtra LowStop. Opis w tekście

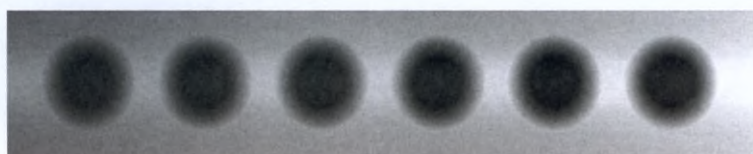
działania tego filtra pokazuje il. 6.14. Nie działa on perfekcyjnie, zwłaszcza w porównaniu z metodą opartą na dzieleniu i mnożeniu. Daje pewne wyrównanie tła, ale nie jest ono całkowicie równomierne, podobnie jak najniższe wartości dla poszczególnych obiektów. Wartości na profilu są jak gdyby rozchwiane, ale wyniki detekcji okazują się zaskakująco dobre (il. 6.15).

Ostatnią analizowaną metodą będzie wykorzystanie szybkiej transformaty Fouriera (FFT – *Fast Fourier Transform*). W widmie Fouriera można usunąć składowe odpowiadające za niskie częstotliwości i za pomocą odwrotnej transformaty odtworzyć obraz, ale już pozbawiony niskich częstotliwości. Powyższy opis brzmi prosto, ale cała operacja jest dosyć skomplikowana, a wynikowy obraz ma wartości zmiennie-przecinkowe, zarówno dodatnie, jak i ujemne. Dlatego trzeba jeszcze wykonać dodatkowe operacje w celu uzyskania poprawnego wyniku. Wyniki korekty cienia za pomocą FFT pokazuje il. 6.16. Profil jest podobny do tego, który otrzymaliśmy po filtracji LowStop, ale obraz jest wyraźnie zniekształcony – wykryte obiekty nie są kołami, tylko elipsami.

Błędy widoczne na il. 6.16 wynikają z faktu, że szybka transformata Fouriera wymaga obrazów o wymiarach będących potęgą liczby 2. Jeżeli rozmiar obrazu nie jest odpowiedni, to jest on wirtualnie powiększany i nowe miejsca są wypełniane zerami. Ponieważ nasz testowy obraz ma niewielkie rozmiary w kierunku pionowym, sprawia to, że nie ma odpowiedniej liczby danych do wyznaczenia funkcji o niskich częstotliwościach.



Il. 6.16. Wyniki korekcji za pomocą transformaty Fouriera. Opis w tekście



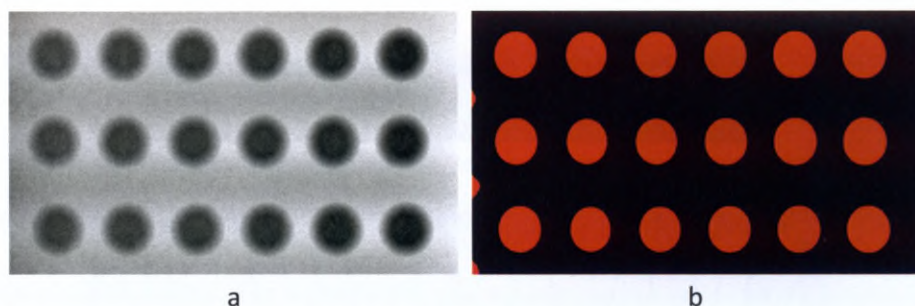
Il. 6.17. Korekcja cienia za pomocą FFT z mniejszą liczbą usuwanych częstotliwości. Opis w tekście

Można starać się pokonać to ograniczenie przez usunięcie mniejszej liczby niskich częstotliwości. Wynik pokazuje il. 6.17. Kształt obiektów jest już dość dobrze zachowany, ale z powodu usunięcia zbyt małej liczby częstotliwości efekt cienia nie został właściwie skorygowany.

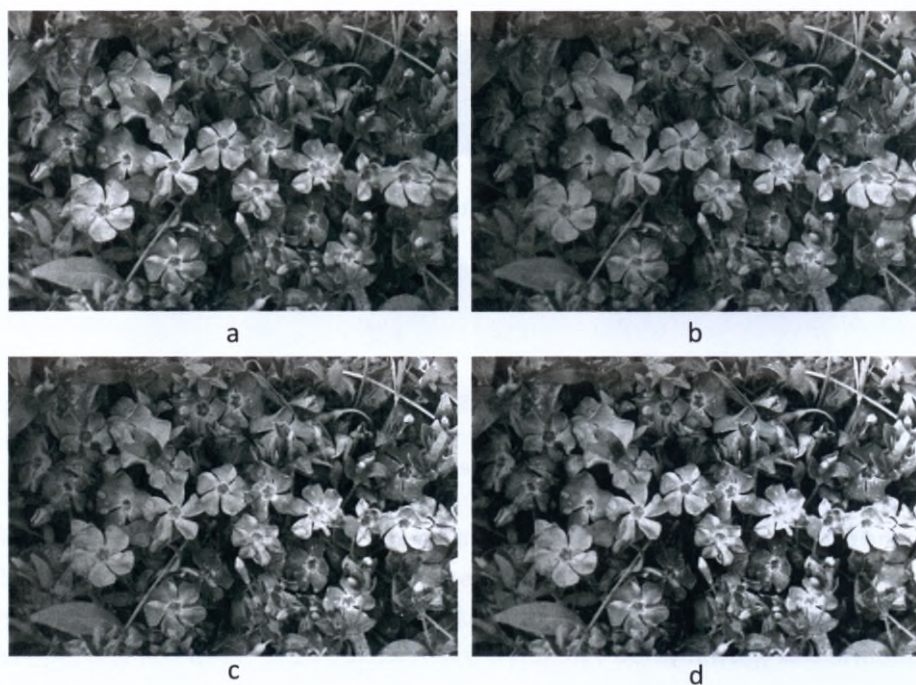
Jeżeli natomiast zwiększylibyśmy wysokość obrazu testowego przez jego multiplikację, to wynik filtracji Fouriera dałby całkiem zadowalające wyniki (il. 6.18a), choć rozmiary kół po binaryzacji pozostałyby jeszcze lekko zróżnicowane.

Podsumowując, wyrafinowane narzędzia oparte na analizie częstotliwościowej nie dały w przypadku analizowanych obrazów wyników lepszych niż metoda oparta na dzieleniu i mnożeniu. Metody te mogą jednak okazać się bardziej skuteczne w przypadku takich obrazów, w których trudno jest wyekstrahować obraz tła.

Wykonamy odpowiednią próbę na monochromatycznej wersji obrazu z il. 6.11a, którą przedstawia il. 6.19a. Obraz z nierównomiernym tłem (il. 6.19b) poddaliśmy



Il. 6.18. Korekcja cienia za pomocą FFT na obrazie o większym rozmiarze w kierunku pionowym. Opis w tekście



Il. 6.19. Filtracja Fouriera na niejednorodnym obrazie. Opis w tekście

filtracji fourierowskiej i w wyniku odwrotnej transformaty Fouriera otrzymaliśmy obraz z il. 6.19c. Po dodatkowej liniowej poprawie kontrastu dostajemy końcowy obraz (il. 6.19d), który jednak nie oddaje w pełni obrazu wyjściowego (il. 6.19a). Zatem filtracje częstotliwościowe nie są najlepszym narzędziem do korekcji cienia. Sprawdzają się natomiast w takich problemach, jak redukcja szumu, wykrywanie elementów wykazujących okresowość lub określoną orientację itp. Więcej informacji na ten temat Czytelnik znajdzie w rozdziale poświęconym zaawansowanym metodom analizy.

7. Binaryzacja

Najczęściej w analizie obrazu spotykamy jeden z trzech rodzajów obrazów, które różnią się liczbą bitów przypadających na jeden piksel (co nazywamy również głębią bitową), wyglądem podczas wyświetlania oraz, oczywiście, ilością niesionej informacji. Są to znane nam już obrazy:

- binarne,
- szare (nazywane też wieloodcieniowymi),
- kolorowe.



Il. 7.1. Przykłady zastosowania obrazów binarnych. Opis w tekście

Obrazy binarne swoją nazwę zawdzięczają temu, że każdy piksel może przyjmować tylko dwie wartości: zero lub jeden. Ilość informacji niesionej przez pojedynczy piksel jest niewielka, ale dzięki temu obrazy binarne zajmują mało miejsca. Są wykorzystywane na cztery podstawowe sposoby:

- Stanowią podstawę wszelkich pomiarów. Spójne zbiory pikseli, najczęściej tych o wartości 1, definiują obiekty, na których wykonywane są pomiary. Gdyby nie było obrazów binarnych, system analizy obrazu nie wiedziałby, gdzie w obrazie są granice analizowanych obiektów (il. 7.1a). Można wręcz stwierdzić, że są one bardzo często końcowym etapem przetwarzania obrazów w procesie analizy obrazu.

- Drugim bardzo ważnym zastosowaniem obrazów binarnych jest dostarczanie informacji niezbędnej w wielu złożonych algorytmach, między innymi w morfologii matematycznej. Stanowią wręcz niezbędny element pośredni lub pomocniczy w tych przekształceniach.
- Obrazy binarne są również pożądanym źródłem informacji w systemach OCR, czyli rozpoznawania pisma. Podobnie obrazy binarne wykorzystuje się do zapisu kodów kreskowych oraz kodów QR (il. 7.1b).
- Nadają się również, choć ostatnio jest to rzadko stosowane, do rejestrowania obrazów, które zawierają tylko białe i czarne elementy (il. 7.1c). Obecnie takie ilustracje są najczęściej zapisywane w postaci obrazów szarych.

Powyższe zestawienie pokazuje, jak ważną rolę w analizie obrazu pełnią obrazy binarne. Najczęściej punktem wyjścia do ich tworzenia jest obraz szary, rzadziej kolorowy. Procedurę zamiany na obraz binarny nazywamy binaryzacją. W tym rozdziale ograniczymy się do binaryzacji obrazów szarych.

Porównanie obrazów: kolorowego, szarego oraz binarnego przedstawia il. 7.2.



Il. 7.2. Porównanie obrazu kolorowego, szarego i binarnego

Sposób sterowania procedurą binaryzacji obrazów kolorowych może być różnie zorganizowany w poszczególnych systemach analizy obrazu, ale nie będziemy obecnie tego zagadnienia analizowali szerzej, gdyż binaryzacja obrazu kolorowego sprowadza się do binaryzacji jego składowych, które są obrazami szarymi. W ogólności



Il. 7.3. Dwa sposoby wyświetlania obrazów binarnych. Opis w tekście

każda składowa może być przetwarzana inaczej. Ostateczny wynik binaryzacji obrazu kolorowego jest najczęściej wynikiem operacji logicznych na binarnych formach poszczególnych składowych. Dodatkowe informacje na temat modeli kolorów oraz binaryzacji obrazów kolorowych znajdują się w rozdziale poświęconym ich analizie.

Systemy graficzne do przetwarzania obrazu zwykle wyświetlają obrazy binarne w ten sposób, że piksele o wartości 0 pozostają czarne, a piksele o wartości 1 stają się białe (il. 7.3a). Ponieważ identycznie byłby wyświetlany obraz szary, mający tylko dwie wartości pikseli (np. 8-bitowy obraz szary z pikselami o wartości wyłącznie 0 lub 255), w systemach analizy obrazu moglibyśmy popełniać błędy polegające na złym zakwalifikowaniu obrazu. Dlatego obrazy binarne wyświetla się w nich zazwyczaj tak, że piksele o wartości 1 są pokazywane w kolorze czerwonym (il. 7.3b). W dalszej części tej książki obrazy binarne będą pokazywane z wykorzystaniem obu omówionych sposobów, przy czym kryterium będzie stanowić czytelność ilustracji.

Z matematycznego punktu widzenia binaryzacja jest funkcją przekształcającą wartości poszczególnych pikseli. Załóżmy, że binaryzowany obraz składa się z punktów, których wartość minimalna wynosi $w_{\min}$, a maksymalna $w_{\max}$. Binaryzację opisuje wtedy funkcja $f(p_d, p_g)$, której argumenty p_d i p_g nazywamy odpowiednio progiem dolnym i progiem górnym. Skrajne wartości punktów oraz progi powinny spełniać nierówność:

$$w_{\min} \leq p_d \leq p_g \leq w_{\max} \quad (7.1)$$

Funkcja binaryzacji dla dowolnego punktu obrazu x przybiera postać:

$$f(x) = \begin{cases} 1 & \text{dla } p_d \leq x \leq p_g \\ 0 & \text{dla } x < p_d \text{ lub } x > p_g \end{cases} \quad (7.2)$$

Jeżeli $p_g = w_{\max}$, to binaryzację nazywamy binaryzacją z progiem dolnym. Jeżeli $p_d = w_{\min}$, to mówimy o binaryzacji z progiem górnym. Opisując funkcję binaryzacji

słowami, stwierdzamy, że wszystkie punkty o wartościach mniejszych od progu dolnego i punkty o wartościach większych od progu górnego przyjmują wartość zero, a pozostałe przyjmują wartość 1. Obecność pojęcia progów binaryzacji sprawia, że operacja ta bywa nazywana również progowaniem (ang. *thresholding*).

W najczęściej spotykanych obrazach szarych informację zapisujemy za pomocą 8 bitów, a więc są to wartości całkowite z zakresu 0-255. Ten zakres w zupełności wystarczy do celów grafiki, gdyż ludzkie oko rozróżnia tylko ok. 40-50 poziomów szarości. Systemy analizy obrazu nie skupiają się jednak na zagadnieniach wizualizacji, tylko na przetwarzaniu i wyodrębnianiu informacji. Dlatego występują obrazy szare, zawierające wartości 16-bitowe (niezbędne, gdy chcemy pomnożyć przez siebie dwa obrazy), wartości ujemne (niezbędne przy odejmowaniu obrazów) oraz wartości zmiennooprzecinkowe (niezbędne m.in. do operacji dzielenia oraz transformaty Fouriera). Każda z wymienionych wyżej odmian obrazów szarych może być poddana binaryzacji.

Aby ułatwić manualną, czyli interaktywną binaryzację, systemy analizy obrazu pokazują na bieżąco, jak będzie wyglądał początkowy obraz (il. 7.4a) po binaryzacji. W tym celu na początkowy obraz nakładana jest półprzezroczysta czerwona maska, pokazująca, którym pikselom zostanie przypisana wartość 1 (il. 7.4b). Zmieniając suwakiem lub wpisując z klawiatury wartości progów detekcji, na bieżąco widzimy zmiany obszarów poszczególnych wartości w przyszłym obrazie binarnym. Ostateczny wynik binaryzacji pokazują il. 7.3a i 7.3b.

Kolejnym narzędziem, które ułatwia dobór progów binaryzacji, jest analiza profilu, czyli wykresu zmian stopni szarości. Jego wykorzystanie omówimy na przykładzie il. 7.5a, przedstawiającej strukturę spiekanego proszku, w której można wyróżnić trzy podstawowe elementy (tzw. składniki strukturalne). Są to: pory, które widzimy



a

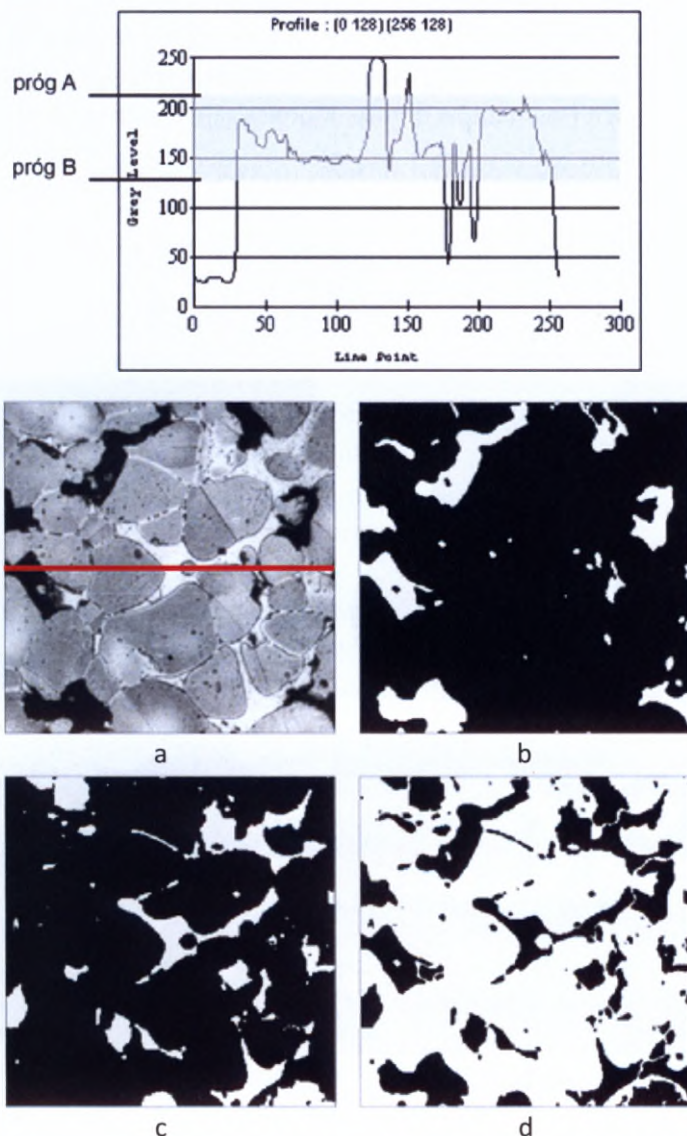


b

Il. 7.4. Wygląd obrazu podczas procesu binaryzacji: a) obraz początkowy, b) obraz początkowy z półprzezroczystą maską, pokazującą elementy, którym zostanie przypisana wartość 1

w postaci ciemnych plamek; ziarna spieku, widoczne w postaci szarych obszarów; oraz dodatkowe cząstki (wydzielenia), stanowiące białe plamy. Spróbujemy teraz rozdzielić te elementy za pomocą binaryzacji.

Spójrzmy na profil, widoczny u samej góry il. 7.5. Przedstawia on zmiany stopni szarości wzdłuż poziomej czerwonej linii, biegnącej przez środek il. 7.5a. Jak łatwo dostrzec, białe wydzielenia odpowiadają miejscom, w których punkty obrazu mają wartości

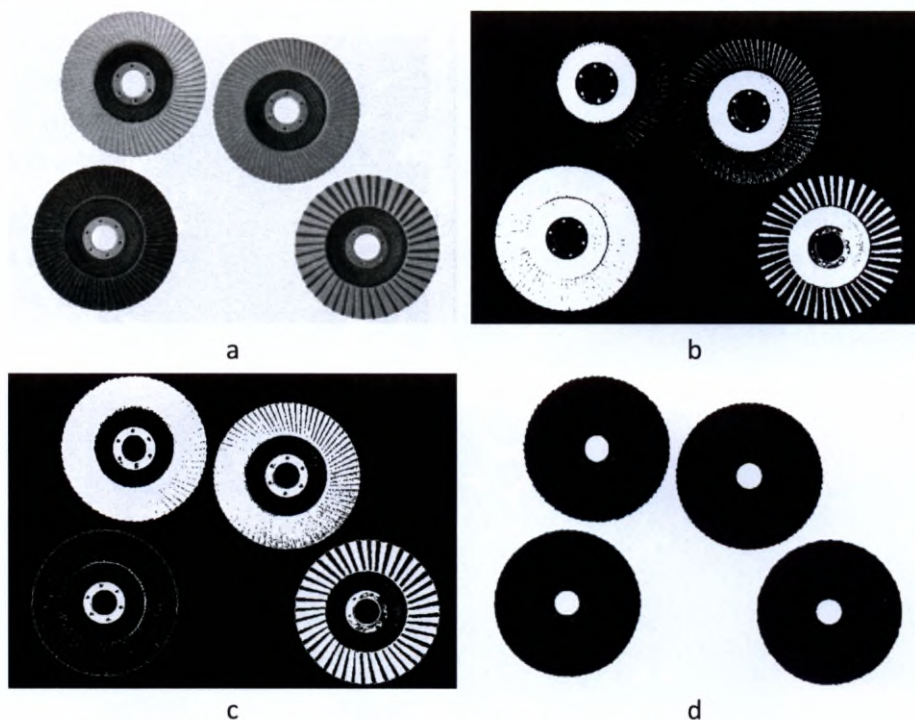


Il. 7.5. Wykorzystanie profilu do ustalania wartości progów binaryzacji. Opis w tekście

wyższe od progu A, natomiast obszar zajęty przez pory odpowiada punktom o wartościach mniejszych od progu B. Jeżeli wykonamy polecenie: „wybierz wszystkie punkty, które mają wartości mniejsze od progu B”, to wykonamy binaryzację z górnym progiem (zakres wartości jest ograniczony przez nas od góry) i otrzymamy obraz obszarów zajętych przez pory (il. 7.5b).

Jeżeli wykonamy polecenie: „wybierz wszystkie punkty, które mają wartości większe od progu A”, to wykonamy binaryzację z dolnym progiem (zakres wartości jest ograniczony przez nas od dołu) i otrzymamy obraz obszarów występowania wydzieliń (il. 7.5c). Możemy również wykonać binaryzację z dwoma progami, górnym i dolnym, co doprowadzi do wybrania tych fragmentów obrazu, które są zajęte przez ziarna spieku (il. 7.5d).

W zależności od przyjętych progów możemy uzyskać zdecydowanie odmienne obrazy, co ukazuje il. 7.6. Jeżeli wyjściowy obraz ściernic listkowych, zapisany jako 8-bitowy (il. 7.6a), zostanie poddany binaryzacji z progami 0 oraz 100, to wykryjemy najciemniejsze elementy (il. 7.6b). Jest to jednocześnie binaryzacja z progiem górnym, gdyż wartość progu dolnego odpowiada dolnemu kresowi poziomów szarości w analizowanym



Il. 7.6. Wpływ progów detekcji na wynik binaryzacji: a) obraz początkowy, b) binaryzacja z progami 0 oraz 100, c) binaryzacja z progami 100 oraz 210, d) binaryzacja z progami 210 oraz 255

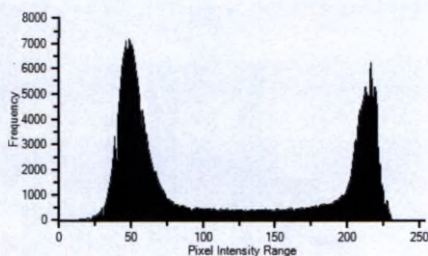
obrazie. Jeżeli wybierzemy progi 100 oraz 210, to otrzymamy binarny obraz jaśniejszych elementów ściernic (il. 7.6c), ale bez tła, które jest znacznie jaśniejsze. Wreszcie, jeżeli przeprowadzimy binaryzację w zakresie 210-255, czyli binaryzację z progiem dolnym 210, to otrzymamy obraz samego tła (il. 7.6d).

Przedstawione powyżej przykłady pokazują, że wynik binaryzacji bardzo silnie zależy od doboru wartości progowych. To wyjaśnia, dlaczego niekiedy na początku całego procesu analizy obrazu wykonuje się normalizację lub wyrównanie histogramu – często umożliwia to przyjęcie tego samego progu czułości dla całej serii obrazów. Nie zawsze jednak udaje się to zrobić. W takich przypadkach próbuje się wykorzystać różne metody automatycznej binaryzacji, które wyznaczają wartości progowe na podstawie analizy wybranych cech obrazu. Odpowiednio zdefiniowane metody automatyczne są względnie mało czułe na pewne zmiany w charakterystyce obrazu, zwłaszcza dotyczące jego jasności i kontrastu, co daje szansę na uzyskiwanie powtarzalnych wyników. Trzeba jednak pamiętać, że każda metoda ma swoje ograniczenia i nie dysponujemy uniwersalnym automatycznym narzędziem, które mogłoby zostać wykorzystane do binaryzacji wszystkich obrazów.

Wyjątkowo wdzięcznym obiektem do prób automatycznej binaryzacji są obrazy zawierające jasne obiekty na ciemnym tle lub ciemne obiekty na jasnym tle (przykład takiego obrazu pokazuje il. 7.7a). Często staramy się tak właśnie przygotować obrazy do analizy. Obraz na il. 7.7a nie jest obrazem binarnym, ale wartości poszczególnych pikseli grupują się w okolicach dwóch pików na histogramie (il. 7.7b): odpowiadającego ciemnym elementom (z lewej strony) oraz odpowiadającego jasnym elementom (z prawej strony). Wydaje się logiczne, że w celu oddzielenia części jasnej od ciemnej próg detekcji powinien odpowiadać minimalnej wartości między pikami



a



b

Il. 7.7. Obraz (a) oraz histogram jego poziomów szarości (b)

histogramu. Takie właśnie progi detekcji są zwykle wybierane podczas manualnej (interaktywnej) binaryzacji. Obserwację przebiegu histogramu można wykorzystać do wybrania progu detekcji.

Jednym ze sposobów automatycznej realizacji tego postulatu jest minimalizacja wariancji odchyłeń wartości pikseli od wartości progowej. W praktyce wystarczy zminimalizować sumę kwadratów odchyłeń.

Algorytm postępowania będzie następujący: założymy, że w obrazie mamy P poziomów szarości, oznaczonych odpowiednio: $p_1, p_2, p_3, p_4, \dots, p_P$, które spełniają warunek $p_1 < p_2 < p_3 < p_4 \dots < p_P$. Obraz zawiera N pikseli, przy czym $N = np_1 + np_2 + np_3 + np_4 + \dots + np_P$, gdzie np_i oznacza liczbę pikseli o poziomie szarości p_i . Wybieramy teraz dowolną wartość progową px , która jest równa jednemu z poziomów szarości. Nie ma przy tym znaczenia, czy jakkolwiek piksel obrazu ma taką wartość, czy też nie. Obliczamy teraz wartość sumy $S(px)$ kwadratów odchyłeń wartości wszystkich pikseli od wartości progowej:

$$S(px) = \sum_{i=1}^P np_i \cdot (px - p_i)^2 \quad (7.3)$$

Ta wartość px , dla której suma $S(px)$, opisana wzorem (7.3), osiągnie wartość minimalną, stanowi poszukiwaną przez nas wartość progu detekcji. Wynik takich obliczeń pokazano dla symulowanych danych na il. 7.8. Jak widać, wyznaczona wartość progowa dobrze wyznacza minimum pomiędzy dwoma pikami histogramu. Opisany sposób wyznaczania wartości progowej odpowiada tzw. metodzie Otsu.



Il. 7.8. Rozkład poziomów szarości oraz próg detekcji zaznaczony czerwonym kolorem

Drugim bardzo popularnym sposobem automatycznej binaryzacji jest metoda maksimum entropii (ang. *entropy threshold*). Algorytm postępowania w tej metodzie jest następujący: założmy, podobnie jak w przypadku binaryzacji metodą minimalizacji wariancji, że w obrazie mamy P poziomów szarości, oznaczonych odpowiednio: $p_1, p_2, p_3, p_4, \dots, p_P$, które spełniają warunek $p_1 < p_2 < p_3 < p_4 \dots < p_P$. Obraz zawiera N pikseli, przy czym $N = np_1 + np_2 + np_3 + np_4 + \dots + np_P$, gdzie npi oznacza liczbę pikseli o poziomie szarości pi . Wybieramy teraz dowolną wartość progową px , która jest równa jednemu z poziomów szarości. Wartość ta dzieli wartości wszystkich pikseli na dwie grupy, które w pierwszej grupie mają poziom szarości mniejszy lub równy px , a w drugiej grupie większy od px . Liczebności tych grup, które oznaczymy odpowiednio A i B , są następujące:

$$A = \sum_{i=1}^{px} npi \quad B = \sum_{i=px+1}^P npi \quad (7.4)$$

Następnie wyznaczamy entropię $E(px)$ wszystkich punktów względem wartości px :

$$E(px) = \sum_{i=1}^{px} -\frac{npi}{A} \log\left(\frac{npi}{A}\right) + \sum_{i=px+1}^P -\frac{npi}{B} \log\left(\frac{npi}{B}\right) \quad (7.5)$$

Ta wartość px , dla której entropia $E(px)$ osiągnie najwyższą wartość, stanowi poszukiwany próg do binaryzacji. Na il. 7.9 pokazano próg detekcji dla tego samego histogramu, który pokazano na il. 7.8. Jak widać, wartości progu detekcji pokrywają się.



Il. 7.9. Rozkład poziomów szarości oraz próg detekcji zaznaczony czerwonym kolorem

Rozkład wartości obu pików na il. 7.8 i 7.9 był symetryczny. Zobaczmy teraz, jak zachowują się obie metody w sytuacji, gdy jeden pik jest zdecydowanie większy od drugiego (il. 7.10). W tym przypadku obie metody dały identyczny wynik i dlatego pokazujemy wykres jedynie dla metody maksimum entropii.

Czasami, np. w celu poprawy czytelności lub ujednolicenia parametrów analizy, przeprowadza się zmiany jasności i kontrastu. Nie mają one praktycznie większego wpływu na wyniki automatycznej detekcji omawianymi metodami. Jest jednak pewien drastyczny sposób modyfikacji obrazu – wyrównanie histogramu. Teraz sprawdzimy, jak taka modyfikacja histogramu z il. 7.10 wpłynie na wyniki automatycznej binaryzacji.



Il. 7.10. Rozkład poziomów szarości oraz próg detekcji zaznaczony czerwonym kolorem.
Wykres dla minimum wariancji jest identyczny

Wyrównanie histogramu polega na takim jego przetworzeniu, aby wynikowa dystrybuanta była możliwie bliska liniowej dystrybuanty, którą wykazuje rozkład doskonale równomierny. Ze względu na dyskretny (nieciągły) charakter cyfrowych danych takie idealne dopasowanie nie jest możliwe, ale można się do niego znacznie przybliżyć, co ukazuje il. 7.11. Jeżeli histogram widoczny na il. 7.11 podzielimy na pięć klas i w każdej zsumujemy cztery wartości występujące w tej klasie, to otrzymamy kolejno: 16, 13, 16, 12 i 17, a więc wartości te są zdecydowanie równiejsze, niż było to przed operacją wyrównania histogramu. Wtedy bowiem dla pięciu klas, liczonych jak wyżej, otrzymalibyśmy następujące liczebności: 8, 13, 5, 31 i 17.

Wyniki symulacji wyboru progu detekcji po wyrównaniu histogramu dla obu analizowanych metod pokazano na il. 7.11. W porównaniu z histogramem przed



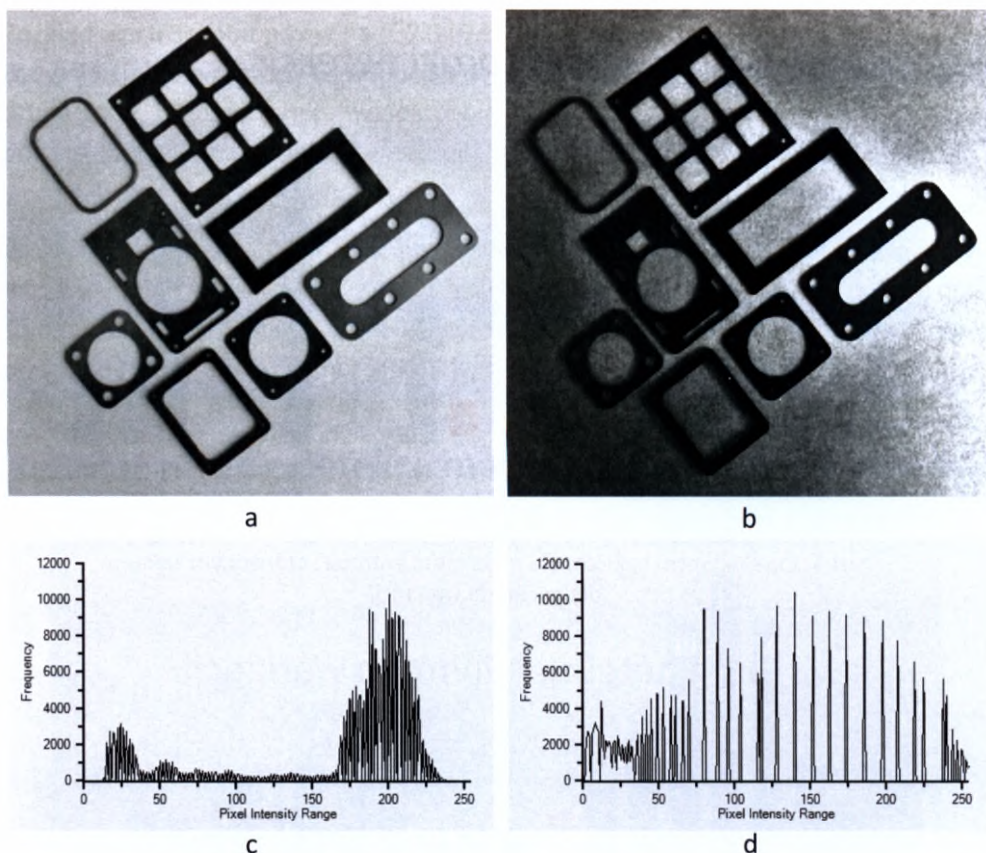
Il. 7.11. Rozkład poziomów szarości dla wyrównanego histogramu oraz progi detekcji zaznaczone zielonym kolorem dla maksimum entropii i czerwonym kolorem dla minimum wariancji



Il. 7.12. Automatyczny wybór progu detekcji dla metody minimum wariancji w przypadku histogramu wykazującego tylko jedno maksimum. Metoda maksimum entropii daje podobny wynik

wyrównaniem metoda minimum wariancji dała wzrost liczby wartości mniejszych lub równych progowi detekcji o 34,6%, a metoda maksimum entropii tylko o 11,6%. Potwierdza to powszechny pogląd, że metoda entropii jest stosunkowo mało czuła na wybrane modyfikacje badanego obrazu.

Kolejny przykład (il. 7.12) przedstawia wyniki symulacji wyboru progu detekcji dla metody minimum wariancji dla histogramu wykazującego tylko jedno maksimum.

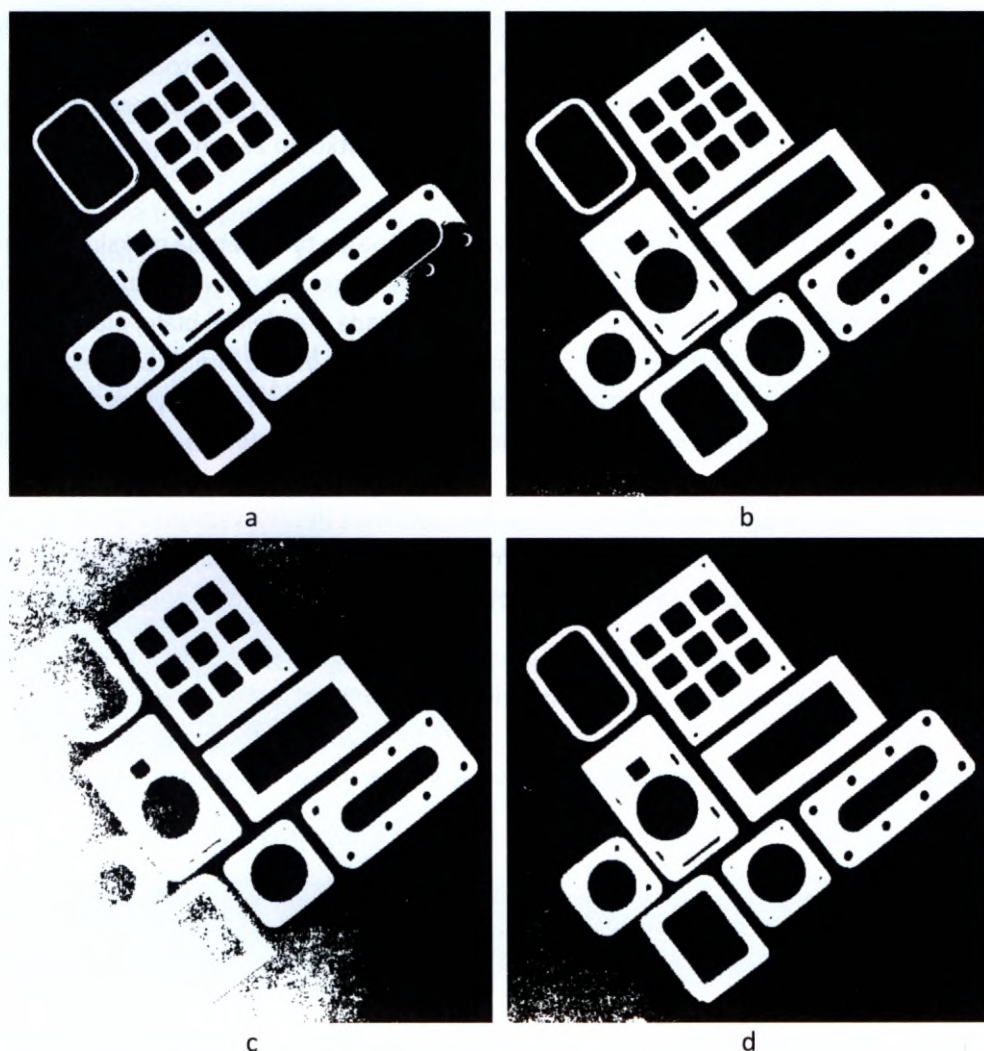


Il. 7.13. Obraz przed i po wyrównaniu histogramu. Opis w tekście

Gdyby dotyczyło to rzeczywistego obrazu o takim właśnie rozkładzie poziomów szarości, to wynik mógłby być dla nas zaskakujący. Binarystacja metodą maksimum entropii dałaby w tym przypadku podobny wynik.

Sprawdźmy teraz, jak obie metody automatycznej detekcji poradzą sobie z rzeczywistymi obrazami. Na il. 7.13a widać kontrastowy obraz uszczelki, który wykazuje silnie niesymetryczny histogram, zamieszczony poniżej zdjęcia. Po wyrównaniu histogramu (patrz il. 7.13b) otrzymujemy obraz, który wygląda dla naszych oczu zdecydowanie gorzej, ale zachowuje jedną cechę obrazu początkowego: jeżeli jeden piksel był jaśniejszy od drugiego w obrazie początkowym, to będzie jaśniejszy również po wyrównaniu histogramu. To właśnie ta cecha daje szansę na poprawną detekcję również po wyrównaniu histogramu.

Przed wyrównaniem histogramu obie metody dały dość dobre wyniki. Metoda Otsu zgubiła niestety część uszczelki z prawej strony obrazu (il. 7.14a). Metoda maksimum



Il. 7.14. Wyniki detekcji: a) metoda Otsu, b) metoda maksimum entropii. U dołu wyniki detekcji po wyrównaniu histogramu: c) metoda Otsu, d) metoda maksimum entropii

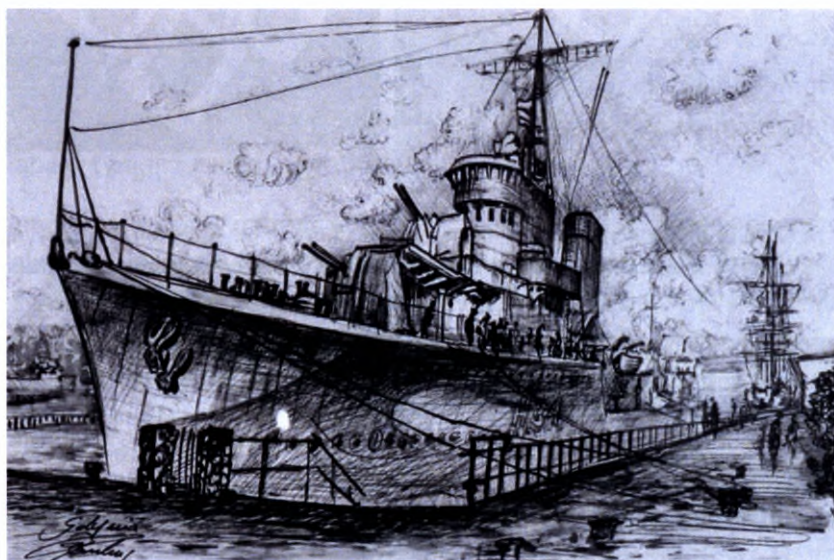
entropii nie zrobiła tego błędu, ale stało się to za cenę niepotrzebnego zwiększenia wymiarów uszczerek z lewej strony obrazu (il. 7.14b).

Sytuacja zmienia się drastycznie po wyrównaniu histogramu. Maksimum entropii wciąż daje dobre wyniki detekcji (il. 7.14d), chociaż u dołu z lewej strony pojawiły się niepotrzebne artefakty. Natomiast wynik metody Otsu to prawdziwa katastrofa (il. 7.14c). Okazuje się jednak, że można uzyskać poprawny wynik binaryzacji, jeżeli wykorzysta się binaryzację z histerezą, która jest opisana nieco dalej.

Ostatni przykład, pokazany na il. 7.13 i 7.14, przekonuje do częstszego stosowania metody maksimum entropii. Okazuje się jednak, że użycie tej metody może przynieść sporą niespodziankę. Próba automatycznej binaryzacji rysunku, przedstawiającego ORP „Błyskawica” (il. 7.15), prowadzi do wyjątkowo zaskakującego wyniku, pokazanego na il. 7.16b.

O ile metoda Otsu dała całkiem dobry wynik (il. 7.16a), o tyle metoda maksimum entropii zupełnie nie poradziła sobie z problemem i dała jedynie niewielki czarny obszar (il. 7.16b). Okazuje się, że ten czarny obszar odpowiada niewielkiej białej plamce, widocznej na il. 7.15, a będącej w rzeczywistości odbiciem błysku lampy, zarejestrowanym podczas reprodukcji obrazu metodami fotograficznymi. Po skorygowaniu tego błędu obie metody dają niemal identyczne wyniki (różnice występują na poziomie pojedynczych pikseli, zajmujących łącznie ok. 3% pola powierzchni obrazu). Z tych analiz płynie następujący wniosek: trzeba bardzo uważać z bezkrytycznym stosowaniem metod automatycznej binaryzacji. Czasami drobny element, nie mający znaczenia dla naszego zmysłu wzroku, może spowodować drastyczne zmiany w wyniku detekcji.

Jedną z cech binaryzacji, o której często zapominają autorzy podręczników, jest to, że bardzo rzadko występuje ona samodzielnie. Najczęściej jest fragmentem jakiegoś algorytmu, który albo modyfikuje obraz przed binaryzacją, albo koryguje jej wynik. Czasami skuteczne narzędzia są dość proste, ale nie mając żadnego doświadczenia, trudno na nie wpaść.



Il. 7.15. ORP „Błyskawica” (grafikę udostępnił autor, Tomasz Poruba)

Przykładem takiego nietypowego problemu może być binaryzacja testowego obrazu, pokazanego na il. 7.17a, gdzie powinniśmy wykryć włókna oraz małe kółeczka – najlepiej oddzielnie, jako dwie klasy obiektów.

Wykorzystanie metody Otsu daje wynik jedynie częściowo poprawny, bo w miejscu rozjaśnienia tła pojawiła się duża plama (il. 7.17b). Maksimum entropii z kolei dało dobre wyniki jedynie tam, gdzie całkowicie zawiodła metoda Otsu (il. 7.18a). Pozostaje jeszcze detekcja manualna (il. 7.18b), która daje wynik podobny do metody Otsu.

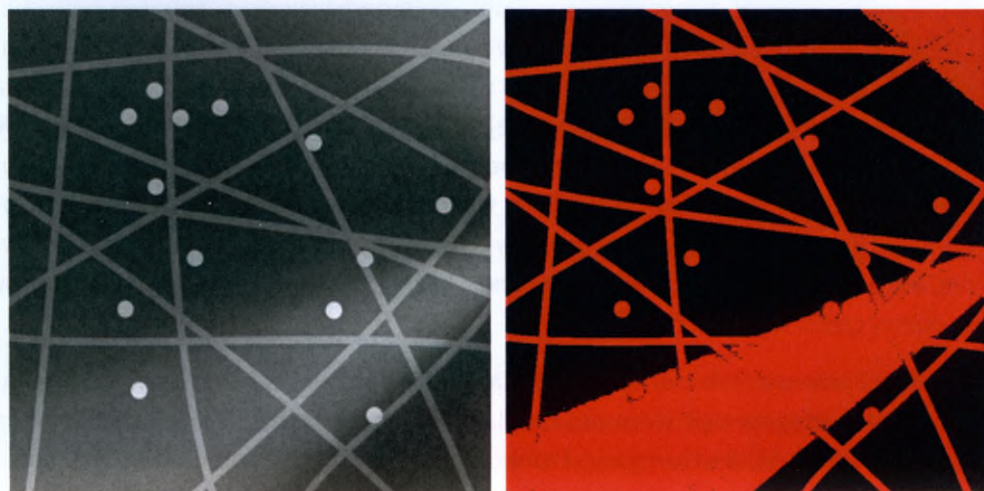
Do rozwiązania problemu poprawnej detekcji włókien w obrazie z il. 7.17a można próbować wykorzystać wyniki binaryzacji metodami automatycznymi. Wyznaczenie



a

b

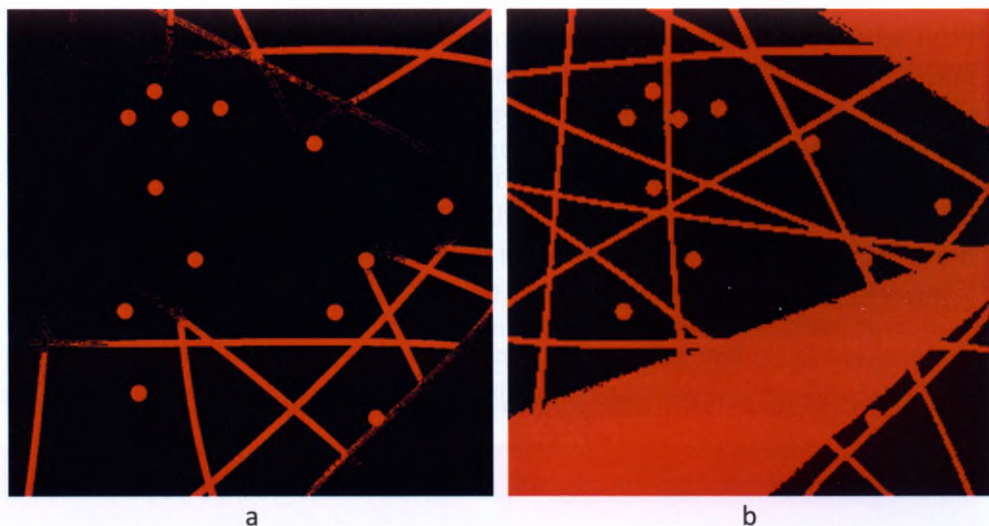
Il. 7.16. Automatyczna binaryzacja obrazu z il. 7.15: a) metoda Otsu, b) maksimum entropii



a

b

Il. 7.17. Przykład problemu z binaryzacją: a) początkowy obraz, na którym należy wykryć włókna i kółka, b) wynik binaryzacji metodą Otsu

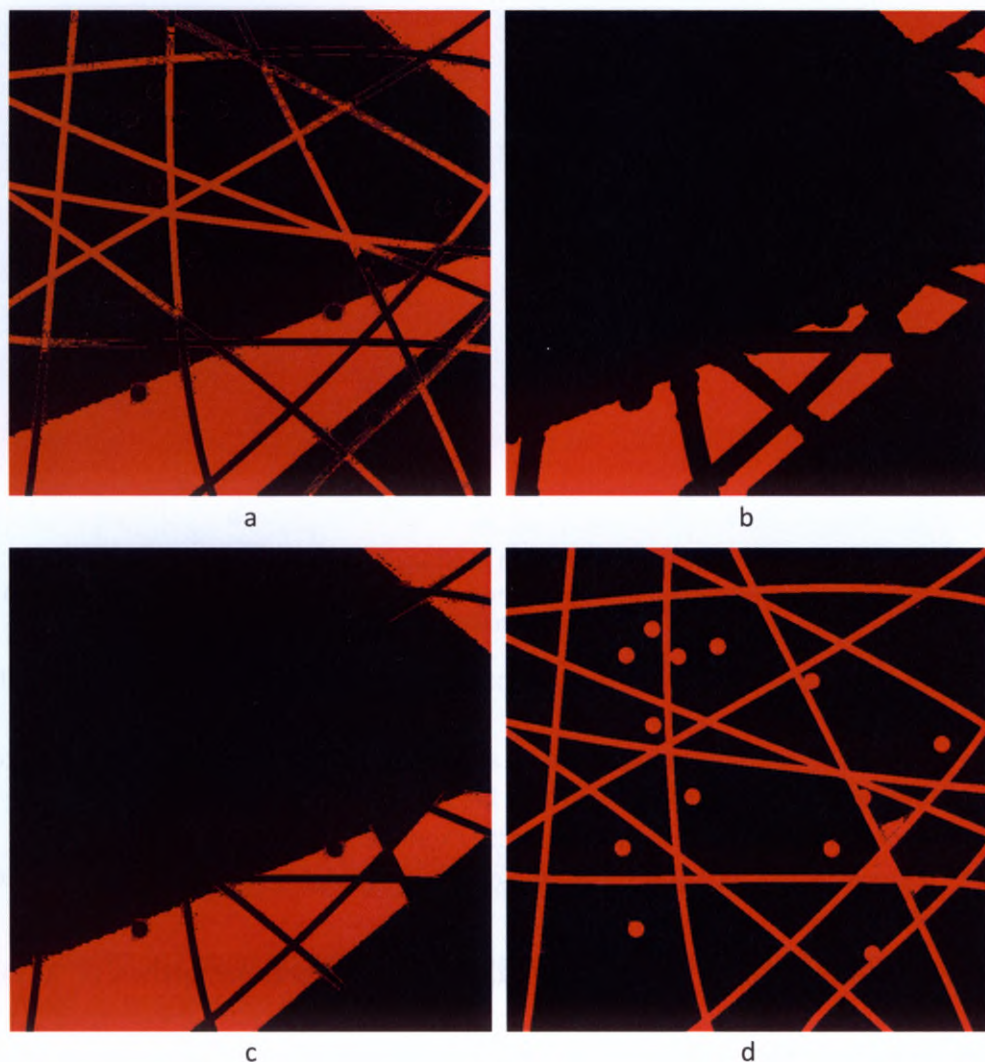


Il. 7.18. Detekcja obiektów w obrazie z il. 7.17a: a) wynik binaryzacji metodą maksimum entropii, b) wynik binaryzacji interaktywnej

różnic pomiędzy tymi wynikami dzięki operacji logicznej XOR daje nam, jak gdyby, poszatkowanie dużych plam obszarów błędnie wykrytych metodą Otsu (il. 7.19a). Kolejnym krokiem jest filtr minimalny o tak dużym rozmiarze, że usuwa włókna i zostawia tylko fragmenty niepotrzebnych struktur (il. 7.19b). Niestety, część tych niepotrzebnych struktur również zostaje usunięta, ale na to nie mamy wpływu.

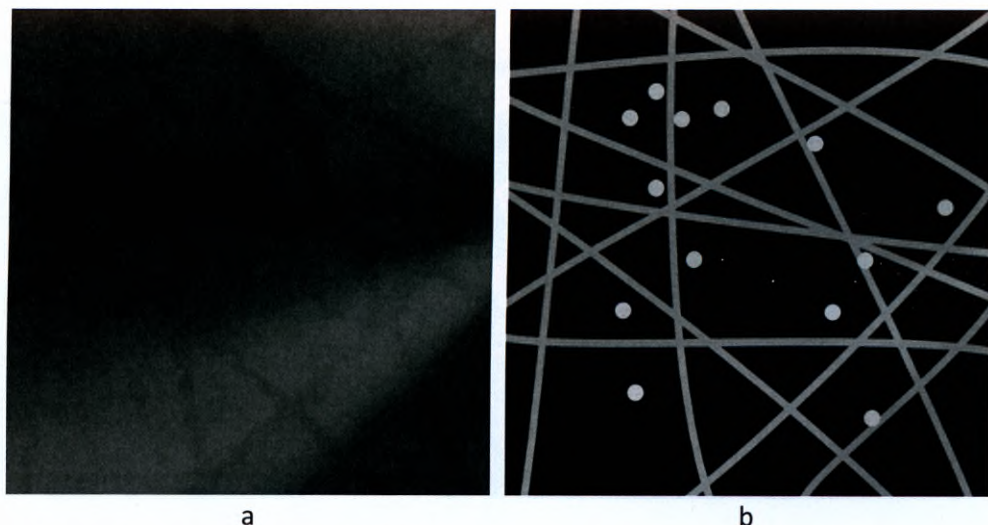
Kolejnym krokiem jest rekonstrukcja, czyli odtworzenie, tych elementów il. 7.19a, których fragmenty (nazywane znacznikami) pozostały po filtracji minimalnej (il. 7.19b). Wynik rekonstrukcji przedstawia il. 7.19c. Wspomniana tu operacja rekonstrukcji należy do tak zwanych operacji morfologicznych i zostanie przedstawiona podczas omawiania tych operacji w kolejnym rozdziale. Obraz widoczny na il. 7.19c odejmujemy teraz od wyniku detekcji metodą Otsu (il. 7.17b). Otrzymany obraz pośredni dodajemy (suma logiczna OR) do wyniku detekcji metodą maksimum entropii (il. 7.18a). Końcowy wynik przedstawia il. 7.19d. Wynik można ocenić pozytywnie, ale pozostały drobne błędy, widoczne na skrzyżowaniu włókien z prawej strony w dolnej części il. 7.19d.

Ostatnie i jednocześnie skuteczne rozwiązanie problemu detekcji elementów obrazu z il. 7.17a okazuje się bardzo proste. Najpierw wykonujemy filtrację minimalną za pomocą filtra o dużym rozmiarze, która usunie obraz włókien oraz kółek. Następnie przeprowadzamy filtrację maksymalną za pomocą filtra o tej samej matrycy, co dość dokładnie przywraca obraz nierównomiernego tła (il. 7.20a). Następnie tak uzyskany obraz po prostu odejmujemy od początkowego obrazu. Uzyskuje się w ten sposób



Il. 7.19. Detekcja obiektów w obrazie z il. 7.17a: a) suma rozłączna (logiczne XOR) obrazów uzyskanych metodami Otsu oraz maksymalnej entropii, b) filtr minimalny obrazu (a), c) rekonstrukcja obrazu (a) na podstawie znaczników z obrazu (b), d) końcowy wynik, opisany w tekście

obraz bardzo łatwy do binaryzacji (il. 7.20b). Przedstawione rozwiązanie to nic innego, jak omawiana w poprzednim rozdziale korekcja cienia. Różnica polega jedynie na tym, że omawiając korekcję cienia, skupialiśmy się na przykładach obrazów z jasnym tłem, a tutaj mamy do czynienia z ciemnym tłem. Choć przedstawione rozwiązanie jest faktycznie proste i skuteczne, to w praktyce często trudno wpaść na pomysł jego wykorzystania. Na obrazie z il. 7.20b bez trudu można wykryć osobno kółka i włókna.



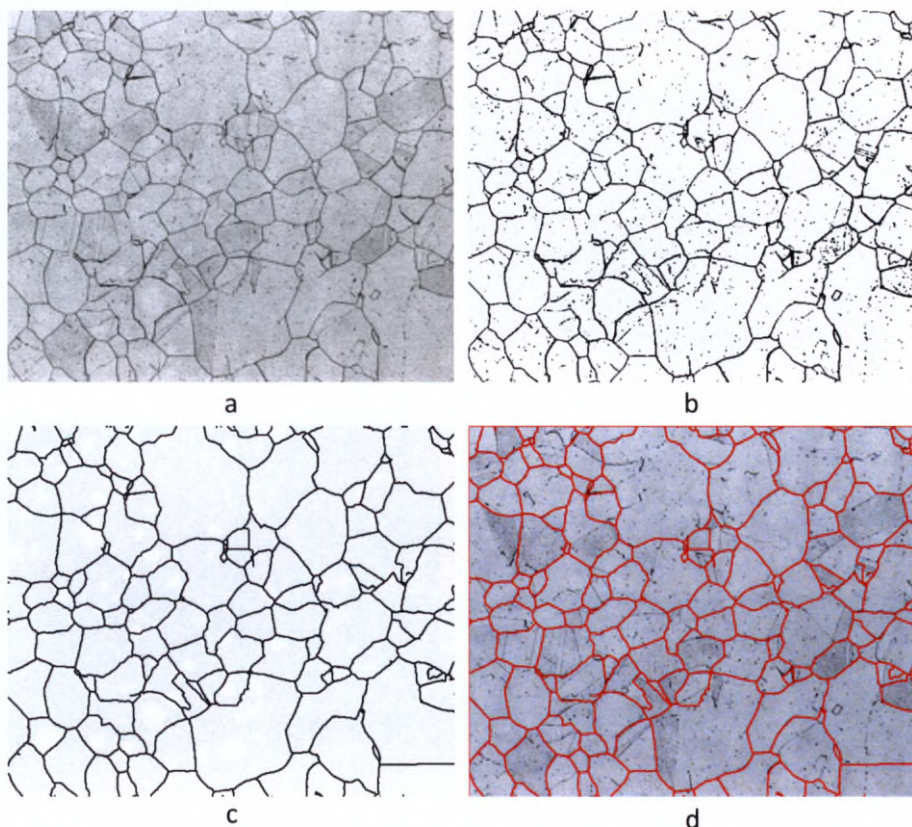
Il. 7.20. Detekcja obiektów w obrazie z il. 7.17a: a) wynik zastosowanych kolejno na obrazie początkowym dwóch filtrów: minimalnego i maksymalnego o bardzo dużej matrycy, b) różnica obrazu początkowego i obrazu (a)

Przedstawione powyżej przykłady pozwalają na wysunięcie trzech istotnych wniosków:

- Należy z dużą ostrożnością stosować automatyczne metody binaryzacji, bo mogą być one źródłem bardzo poważnych błędów.
- Sama binaryzacja często nie wystarczy i konieczne są dodatkowe operacje.
- Zazwyczaj można znaleźć kilka różnych rozwiązań, które mogą być podobnie skuteczne.

Inny problem związany z binaryzacją ukazuje il. 7.21. Obraz początkowy (il. 7.21a) jest łatwy do binaryzacji, ale jej wynik ma liczne wady, z których najważniejszymi są brak ciągłości niektórych linii oraz szum (il. 7.21b). Wbrew pozorom nie jest to sytuacja bardzo trudna, gdyż większość granic pomiędzy ziarnami została już poprawnie wykryta. Otrzymany obraz binarny można skutecznie skorygować (il. 7.21c). Na il. 7.21d widać, że te korekty były rzeczywiście skuteczne. Wykorzystane zostały tutaj przede wszystkim operacje morfologii matematycznej, które zostaną omówione w następnym rozdziale. Najistotniejsze w tym momencie jest to, że binaryzacja dała akceptowalnej jakości obraz binarny.

Czasami dobrym rozwiązaniem jest eliminacja szumu przed binaryzacją. Przykład takiej operacji pokazuje il. 7.22. Binaryzacja obrazu z szumem (il. 7.22a) prowadzi do binarnego obrazu z artefaktami w postaci drobnych, odizolowanych kropek (il. 7.22b). Można je oczywiście usunąć za pomocą np. filtracji medianowej, ale można też

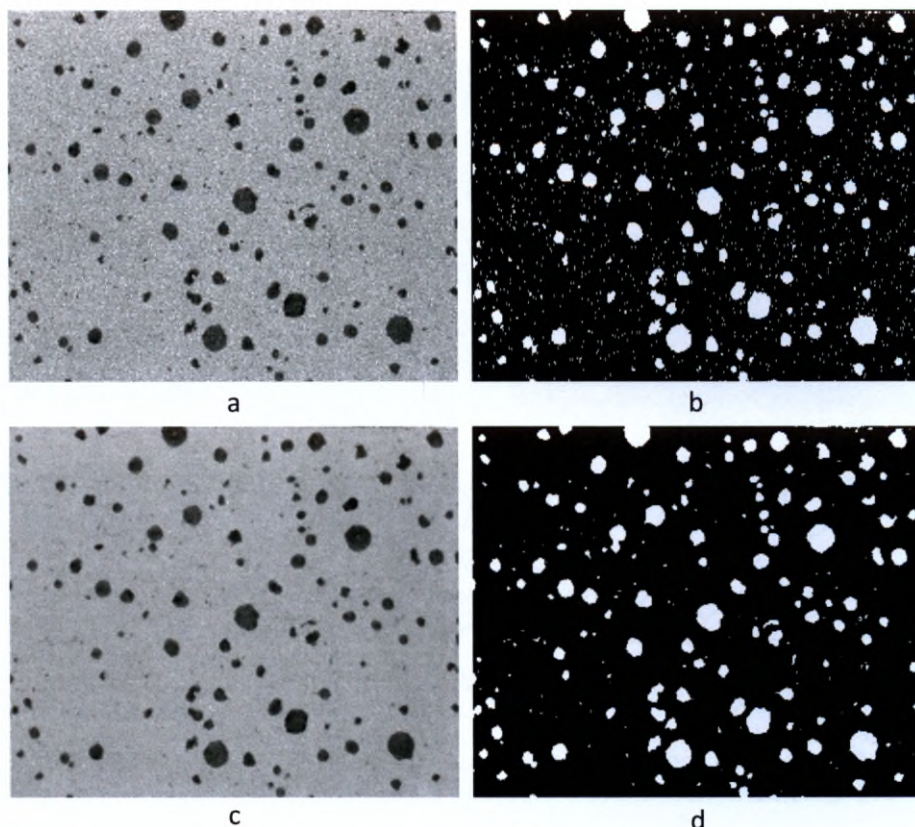


Il. 7.21. Detekcja ziaren: a) obraz początkowy, b) obraz binarny, c) skorygowany obraz binarny (końcowy wynik przekształceń), d) obraz (c) nałożony na obraz (a)

poddać filtracji obraz początkowy, co ukazuje il. 7.22c. Binaryzacja tego obrazu prowadzi do poprawnego wyniku (il. 7.22d).

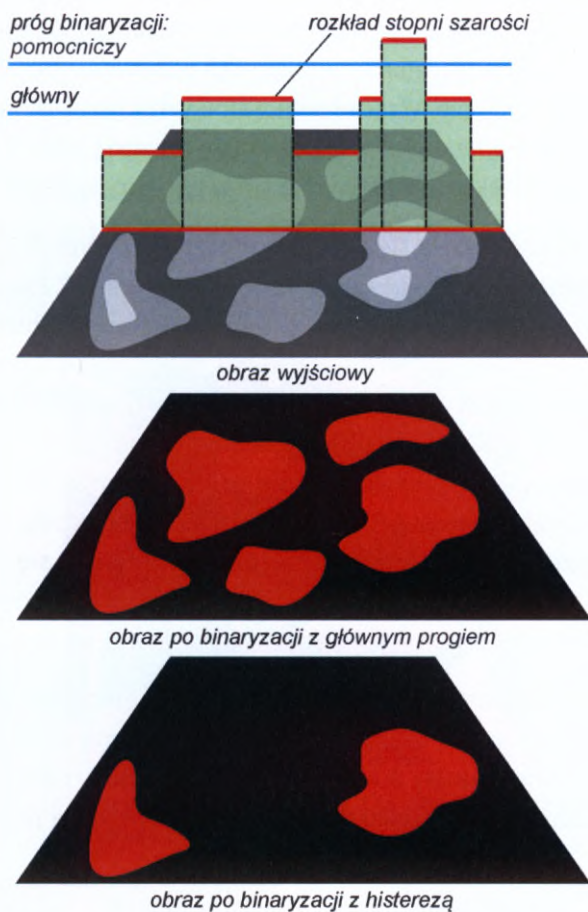
Dyskusja, które rozwiązanie jest lepsze – korekta obrazu przed binaryzacją czy też korekta obrazu binarnego – jest w istocie skazana na niepowodzenie, gdyż jest podobna do dyskusji na temat tego, co było pierwsze: jajko czy kura? Wszystko zależy od analizowanego obrazu i tego, co ma być wynikiem detekcji. Należy jednak podchodzić do filtracji przed binaryzacją z dużą ostrożnością, gdyż możemy usunąć elementy istotne dla późniejszej poprawnej binaryzacji. Z drugiej strony trzeba pamiętać, że nadmiar artefaktów w początkowym obrazie może wręcz uniemożliwić poprawną detekcję.

Interesującą odmianą binaryzacji jest binaryzacja z histerezą. Polega ona na zastosowaniu dwóch zakresów binaryzacji. Nazwijmy je głównym i pomocniczym. Wynikiem końcowym procesu jest zaliczenie do obrazu binarnego tych elementów wykrytych

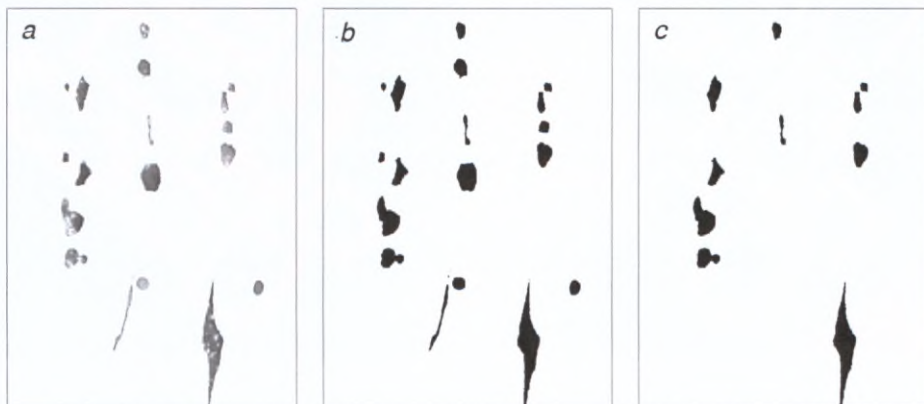


Il. 7.22. Binarystacja obrazu z szumem: a) obraz początkowy, b) obraz binarny – widoczne ślady szumu w postaci białych kropek, c) obraz początkowy po filtracji medianowej, d) obraz binarny uzyskany w wyniku binaryzacji obrazu (c)

przez główny zakres, które zawierają elementy wykryte również przez zakres pomocniczy. Elementy niezawierające żadnych fragmentów wykrytych przez pomocniczy zakres zostają odrzucone. Opisywane przekształcenie można byłoby wykorzystać np. do wyznaczenia sylwetki jednego z dwóch samochodów, różniących się tym, że jeden ma zapalone światła, natomiast drugi – nie. Schemat binaryzacji z histerezą pokazuje il. 7.23, natomiast przykład takiej binaryzacji ukazuje il. 7.24. Na il. 7.23 w obrazie wyjściowym pięć obiektów, z których dwa zawierają jasne plamy. Jeżeli wykonamy normalną binaryzację z dolnym progiem (oznaczonym jako próg główny), to otrzymamy pięć obiektów, jak pokazuje to obraz po detekcji z progiem głównym na il. 7.23. Jeżeli natomiast dołożymy pomocniczy próg detekcji, to wykryjemy trzy niewielkie obiekty, zaznaczone jako najjaśniejsze w obrazie wyjściowym na il. 7.23. Tylko dwa obiekty zawierają takie elementy i dlatego też one stanowią wynik binaryzacji z histerezą (obraz po binaryzacji z histerezą na il. 7.23).



Il. 7.23. Binaryzacja z histerezą. Opis w tekście

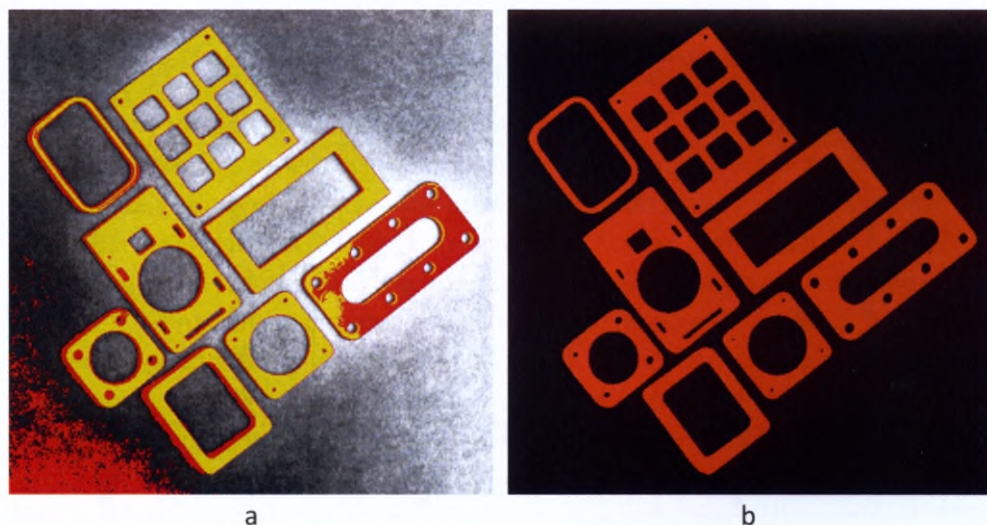


Il. 7.24. Binaryzacja z histerezą: a) obraz wyjściowy, b) obraz po klasycznej binaryzacji, c) obraz po binaryzacji z histerezą – wykryte zostały tylko te elementy obrazu (a), które zawierają jasne plamy

Przykład działania binaryzacji z histerezą na rzeczywistych obiektach pokazuje il. 7.24, przedstawiająca obrazy wtrąceń niemetalicznych w pewnej stali. Dzięki tej operacji możliwe stało się wydzielenie wtrąceń wielofazowych, charakteryzujących się obecnością jasnych plam. Dodatkowo, przez operację logiczną XOR, możemy otrzymać obraz samych tylko cząstek jednofazowych, czyli w tym przypadku jednolicie ciemnych.

Binaryzacja z histerezą okazała się bardzo skuteczna w przypadku obrazu uszczelkek po wyrównaniu histogramu (il. 7.13b). Obraz pokazany na il. 7.25a jest obrazem wyświetlanym przez system analizy obrazu bezpośrednio przed binaryzacją. Wykryte przez system analizy obrazu zostają te elementy, które są pokazane na il. 7.25a jako czerwone (detekcja z wykorzystaniem progu głównego), ale pod warunkiem, że w obszarze czerwonym będą występowały elementy żółte (wykrywane za pomocą progu pomocniczego). Dzięki temu unika się detekcji ciemnych obszarów w lewym dolnym rogu, co miało miejsce np. podczas binaryzacji metodą maksimum entropii (il. 7.14d). Najczęściej binaryzację z histerezą wykorzystuje się w sposób interaktywny, ale można też wykonać ją w pełni automatycznie. Końcowy wynik detekcji uszczelkek pokazuje il. 7.25b.

Nieco zbliżoną, przynajmniej na pierwszy rzut oka, metodą binaryzacji jest wykorzystanie rozrostu obszarów. W tej metodzie poszczególne obiekty są wyznaczane w ten sposób, że kolejne punkty dodaje się do obiektu wtedy i tylko wtedy, kiedy różnica poziomów szarości w stosunku do pierwszego punktu nie przekracza założonej



Il. 7.25. Binaryzacja z histerezą: a) przygotowanie do binaryzacji obrazu uszczelkek, b) końcowy wynik detekcji

wartości. Zwykle powstaje dość dużo obiektów i dlatego są one zapisywane w obrazie etykietowanym. Każdy kolejny region otrzymuje swoją etykietę, czyli kolejny numer. Wszystkie punkty należące do jednego obiektu mają tę samą wartość, równą etykietcie. Zazwyczaj, dla lepszej czytelności, kolejne obiekty są wyświetlane w różnych kolorach (il. 7.26). Ze względu na powstawanie dużej liczby obiektów zwykle mówi się, że rozrost obszarów jest metodą segmentacji (podziału), a nie binaryzacji.



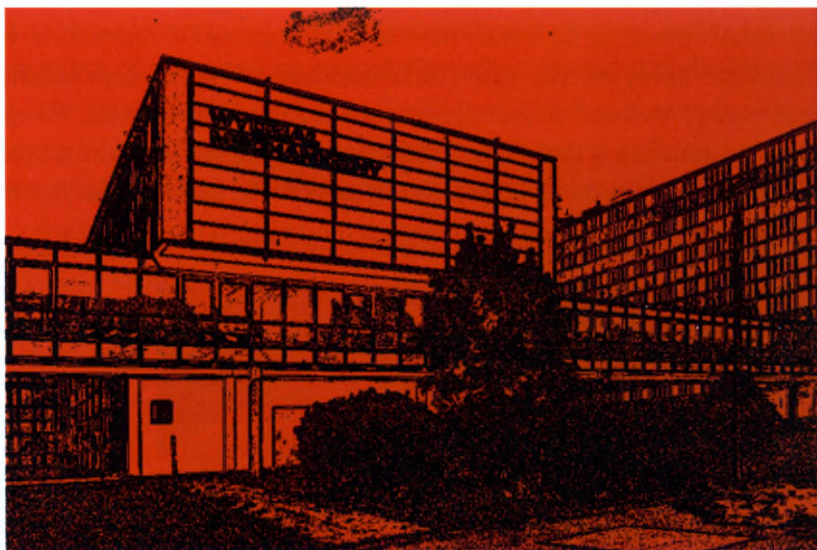
Il. 7.26. Etykietowany obraz segmentacji obrazu z il.7.17a, otrzymany metodą rozrostu obszarów przy progu dopuszczalnej różnicy, wynoszącym 28

Obrazy etykietowane są w istocie odmianą obrazów binarnych, gdyż opisują kolejne obiekty i mogą być wykorzystywane jako podstawa do pomiarów.

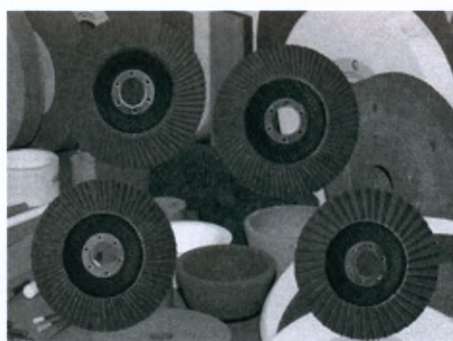
Połączenie filtracji laplasjanem z binaryzacją w okolicach wartości zerowej daje obraz obszarów, w których wartości poziomów szarości zmieniały się powoli. Dobrym przykładem jest obszar nieba na il. 7.27. Jedyne zakłócenie występuje u góry w środkowej części, gdzie pojawiła się chmura. Reszta stanowi jedną całość, choć niebo nie miało w każdym miejscu takiego samego koloru.

Występują jeszcze inne rodzaje binaryzacji, na przykład wykrywające różne obszary na podstawie kilku progów detekcji lub dobierające próg detekcji tak, żeby wykryte obiekty zajmowały określony ułamek powierzchni (czyli np. wykrywamy najjaśniejsze punkty, które łącznie tworzą połowę powierzchni obrazu).

Można próbować budować też algorytmy z dynamicznie zmiennym progiem czułości, co stanowiłoby zbliżenie się do sposobu działania ludzkiego zmysłu wzroku, który nie ma trudności z rozpoznanem włókien i kółek na il. 7.17a. Z drugiej strony równie



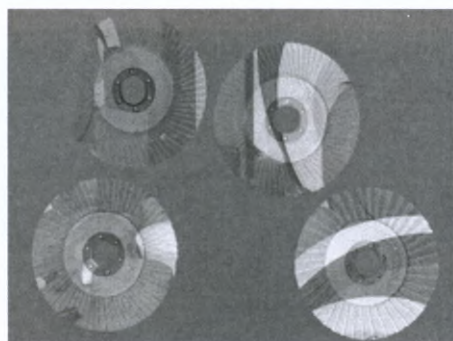
Il. 7.27. Połączenie filtracji laplasjanem z następną binaryzacją laplasjanu



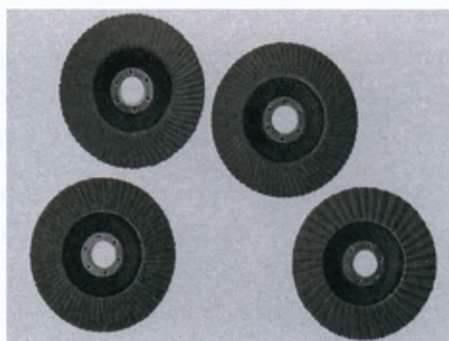
a



b



c



d

Il. 7.28. Detekcja obiektów z wykorzystaniem wcześniej zarejestrowanego tła: a) trudne do wykrycia obiekty, b) tło, c) różnica obrazu początkowego i tła, d) wykryte obiekty

skuteczne, jak dynamiczny próg detekcji, jest odpowiednie przetworzenie obrazu, co pokazuje il. 7.20b. Zmiana poziomu szarości tła pełni tutaj taką samą rolę jak lokalna zmiana progu detekcji.

Również omówione już sposoby binaryzacji można jeszcze modyfikować. Na przykład rozrost obszarów można prowadzić, rozpoczynając od predefiniowanych lokalizacji punktów startowych. Nie będziemy jednak tych zagadnień szczegółowo omawiać, gdyż metody przedstawione dotychczas w ramach tego rozdziału są wystarczające dla zdecydowanej większości problemów.

Na zakończenie omówimy jeszcze pewien trik, używany niekiedy w analizie obrazu, np. do detekcji poruszających się obiektów. Jeżeli mamy jakieś obiekty na bardzo zróżnicowanym tle (np. ściernice listkowe na tle innych ściernic – il. 7.28a), to nie ma praktycznie żadnej możliwości, aby je wykryć w sposób automatyczny. Jeżeli jednak posiadamy obraz samego tła (il. 7.28b), to możemy odjąć od siebie te dwa obrazy (il. 7.28c). Wynik będzie różny od zera tylko w miejscu, gdzie znajdują się interesujące nas obiekty, co można wykorzystać do ich wydzielenia i umieszczenia na dowolnym innym tle. Na il. 7.28d jest to jednolite jasnoszare tło.

Opisaną powyżej metodę można wykorzystać w detekcji poruszających się obiektów (ludzie, samochody, zwierzęta), jak również do rejestracji różnych obiektów, które chcemy później analizować, a chwilowo nie dysponujemy odpowiednim lub jednolitym tłem. Ważne jest jednak, że zarówno warunki ekspozycji, jak i geometria układu rejestracji obrazów (kamera i tło) muszą pozostawać stałe.

8. Operacje morfologiczne

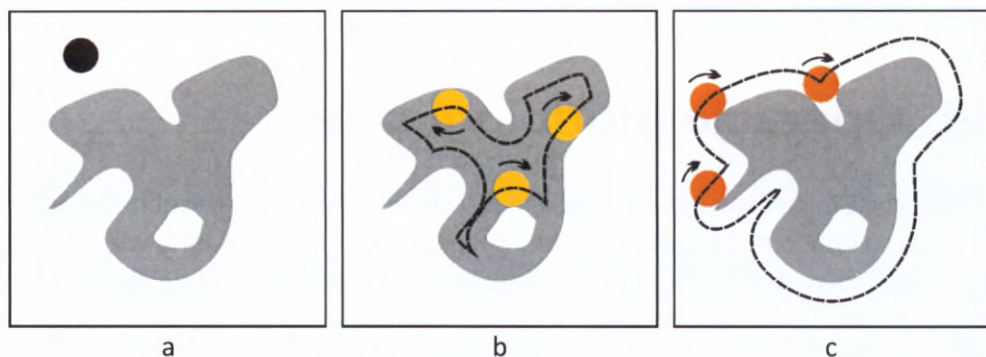
Morfologia matematyczna jest częścią teorii cyfrowych przekształceń obrazów. Wprowadza do nich formalizm matematyczny z odpowiednią symboliką, bada własności poszczególnych przekształceń oraz opracowuje algorytmy ich implementacji. Jednak w celu praktycznego korzystania z operacji morfologii matematycznej wystarczająca okazuje się znajomość podstawowych własności poszczególnych operacji, bez konieczności wglębiania się w teorię oraz formalizm matematyczny. Dlatego też w ramach niniejszego rozdziału nie będziemy zajmowali się ściśle matematyczną stroną tych przekształceń.

Przekształcenia morfologiczne są powszechnie używane jako bardzo skuteczne oraz uniwersalne narzędzie przetwarzania obrazu. Mogą być stosowane do obrazów binarnych, szarych i kolorowych, zarówno 2D, jak i 3D. Choć podstawowa idea jest dość prosta, zrozumienie działania tych operacji wymaga sporego wysiłku. Aby to ułatwić, ograniczymy się na wstępie do operacji morfologicznych w odniesieniu do obrazów binarnych. Dopiero po omówieniu przekształceń obrazów binarnych zajmiemy się rozszerzeniem operacji morfologicznych na obrazy szare.

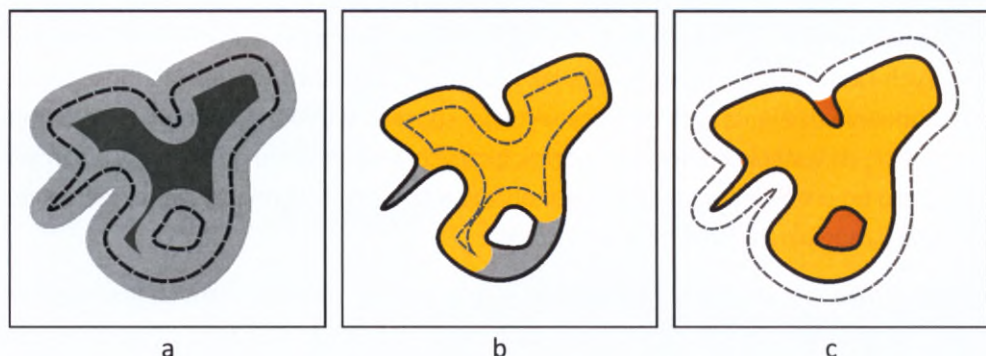
Najważniejszym wstępnym pojęciem w operacjach morfologicznych jest element strukturalny (ang. *structuring element*), za pomocą którego definiuje się poszczególne przekształcenia. Element strukturalny można traktować jako wzorec, szablon lub model fragmentu obrazu. Jest on przykładany do wszystkich pikseli obrazu i w zależności od tego, czy lokalna konfiguracja pikseli zgadza się, czy też nie, z elementem strukturalnym, podejmowane są odpowiednie działania.

Za twórców podstaw morfologii matematycznej uważa się dwóch francuskich uczonych. Byli to Georges Matheron i Jean Serra. O znaczeniu morfologii matematycznej może świadczyć to, że prestiżowa francuska uczelnia, École de Mines de Paris, utworzyła w Fontainebleau silną jednostkę badawczą o nazwie Centre de Morphologie Mathématique, którą przez wiele lat kierował prof. Jean Serra. Teoretyczne podstawy morfologii matematycznej przedstawił w wydanej w roku 1980 przez Academic Press monografii *Image Analysis and Mathematical Morphology*. Znajdziemy tam m.in. definicje podstawowych przekształceń. Założono przy tym, że przetwarzamy obraz binarny, a element strukturalny jest kołem (il. 8.1a).

Możemy teraz łatwo zdefiniować dwa podstawowe przekształcenia morfologii matematycznej: erozję (ang. *erosion*) oraz dylatację (ang. *dilation*). Erozja polega na tym,



Il. 8.1. Podstawowe przekształcenia morfologii matematycznej: a) kołowy element strukturalny oraz figura, którą będziemy poddawali przekształceniom, b) erozja oraz c) dylatacja kołowym elementem strukturalnym



Il. 8.2. Figura wyjściowa po erozji oraz dylatacji (a), po otwarciu (b) i po zamknięciu kołowym elementem strukturalnym (c). Opis w tekście

że kołowy element strukturalny przetaczamy po wewnętrznej stronie brzegu figury (il. 8.2b). Kolejne położenia środka koła, stanowiącego element strukturalny, wyznaczają zerodowaną figurę, zaznaczoną na il. 8.2b za pomocą przerywanej linii. Jak widać na załączonym rysunku, erozja zmniejsza figurę, a ponadto pozbawia ją wąskich fragmentów, zwłaszcza o charakterze wystających wypustek.

Dylatacja jest przekształceniem przeciwnym do erozji. Element strukturalny przetaczamy po zewnętrznej stronie brzegu figury, a granice figury po dylatacji wyznaczają kolejne położenia środka elementu strukturalnego, zaznaczone na il. 8.1c również linią przerywaną. Dylatacja powiększa figurę, zalewa wąskie zatoki oraz niewielkie otwory.

Do erozji i dylatacji będziemy jeszcze niejednokrotnie wracali. Obecnie zajmijmy się złożeniem tych przekształceń. O ile zmiany kształtu, wprowadzane przez erozję i dylatację, mogą być często pożądane, o tyle towarzysząca temu zmiana wielkości

figury może stanowić istotną wadę. Ponieważ wpływ erozji i dylatacji na wielkość obiektu jest dokładnie przeciwny, ich złożenie powinno zachować wielkość figury. Oczywiście spodziewamy się pewnej modyfikacji jej kształtu.

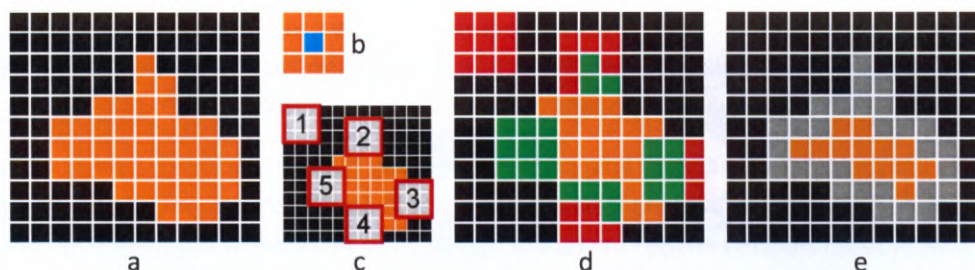
Na il. 8.2a widać zaznaczony linią przerywaną brzeg figury wyjściowej, a także ciemną figurę po erozji oraz jasną (zawierającą w sobie figurę zerodowaną) figurę po dylatacji. Jeżeli złożymy erozję i następnie dylatację, to otrzymamy otwarcie (ang. *opening*) figury. Wynik zaznaczono na il. 8.2b kolorem żółtym. Jak widać, przekształcenie to powoduje otwarcie niektórych zamkniętych obszarów wokół otworów oraz usunięcie wąskich wypustek. Zasadnicza wielkość figury pozostaje jednak niezmienną.

Jeżeli przeprowadzimy najpierw dylatację, a następnie erozję, to otrzymamy figurę (il. 8.2c) składającą się z obszaru żółtego, odpowiadającego wyjściowej figurze, oraz – dodatkowo – zalanych lub zamkniętych obszarów zaznaczonych na pomarańczowo. To przekształcenie nazywamy zamknięciem (ang. *closing*). Zestawienie figur po wszystkich omówionych przekształceniach pokazuje il. 8.3. Widać, że te cztery podstawowe przekształcenia potrafią sporo pozmieniać w kształcie i wielkości figury.



Il. 8.3. Szara figura wyjściowa oraz kolejno od lewej do prawej: figura po erozji, dylatacji, otwarciu i zamknięciu

Idea omówionych przekształceń jest więc dość prosta, ale problemy pojawiają się, gdy chcemy przeprowadzić te operacje na rzeczywistych obrazach. Kwadratowa siatka punktów w cyfrowych obrazach słabo nadaje się do modelowania koła. O wiele lepszym niż kwadrat modelem koła jest sześciokąt foremny. Takie rozwiązanie przyjęła w 1974 roku firma Leitz, wypuszczając na rynek pierwszy komercyjny analizator, wykorzystujący operacje morfologii matematycznej, nazwany TAS (*Texture Analysis System*). Analizator pracował na obrazach binarnych z heksagonalną siatką pikseli, zaś algorytmy opracowało Centre de Morphologie Mathématique. Analizatory z heksagonalną matrycą nie wytrzymały próby czasu. Kłopotliwa była zarówno produkcja odpowiednich monitorów, jak i problemy z odpowiednią konwersją obrazów z kamer. Zostawiamy więc siatkę heksagonalną jako ciekawostkę historyczną



Il. 8.4. Erozyja na obrazie z kwadratową siatką pikseli: a) obraz wyjściowy, b) element strukturalny, c) lokalizacja elementu na obrazie, d) ocena zgodności elementu z obrazem, e) obraz po erozji

i przejdziemy do analizy erozji i dylatacji w obrazach z kwadratową matrycą pikseli. Operację erozji omówimy, wykorzystując il. 8.4.

Poddamy erozji obraz pokazany na il. 8.4a. Czarne kwadraty to piksele o wartości 0, pomarańczowe – piksele o wartości 1. Elementem strukturalnym, który jest najczęściej używany w praktyce, będzie siatka pikseli o wartości 1 i rozmiarze 3x3. Punkt centralny został wyróżniony kolorem niebieskim (il. 8.4b). Algorytm erozji zdefiniujemy teraz następująco:

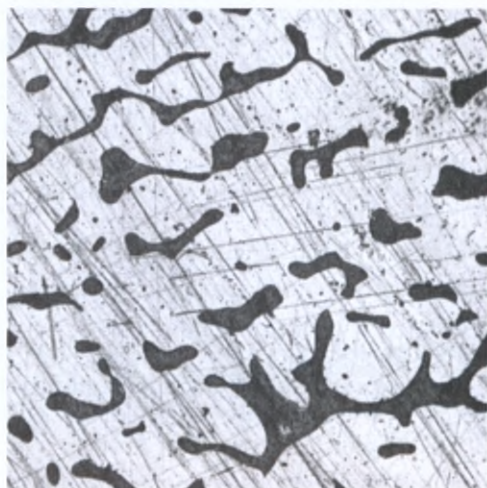
- Przykładamy element strukturalny kolejno do wszystkich pikseli w ten sposób, że centralny punkt elementu strukturalnego pokrywa się z aktualnie analizowanym pikselem.
- Jeżeli otoczenie analizowanego piksela pokrywa się z elementem strukturalnym, to piksel nie zmienia swojej wartości, niezależnie od tego, czy miał wartość 1, czy 0.
- Jeżeli otoczenie analizowanego piksela nie pasuje do elementu strukturalnego, to piksel przyjmuje wartość 0, niezależnie od tego, czy miał wartość 1, czy 0.
- Jeżeli analizowany piksel leży na krawędzi obrazu i brakuje pikseli do porównania z elementem strukturalnym, to przyjmujemy, że brakujące punkty mają wartość 1.

Przeanalizujemy teraz, co się dzieje podczas erozji z obrazem z il. 8.4a na podstawie pięciu różnych położeń elementu strukturalnego, zaznaczonych na il. 8.4c:

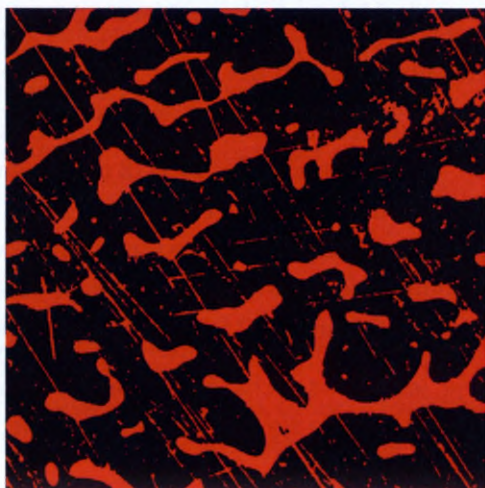
- Położenie 1 – element nie pasuje do otoczenia (jest zaznaczony na czerwono), zatem analizowany piksel przyjmuje wartość 0 (w tym przypadku nie ulegnie ona zmianie).
- Położenia 2 oraz 3 – element tylko częściowo pasuje do otoczenia (pasujące piksele są zaznaczone na zielono, a piksele niepasujące na czerwono), zatem analizowany piksel przyjmuje wartość 0 (w tym przypadku ulegnie ona zmianie, bo piksele odpowiadające centralnemu punktowi elementu strukturalnego miały wartość 1).

- Położenie 4 – element tylko częściowo pasuje do otoczenia (pasujące piksele są zaznaczone na zielono, a piksele niepasujące na czerwono), zatem analizowany piksel przyjmuje wartość 0 (w tym przypadku nie ulegnie ona zmianie).
- Położenie 5 – element pasuje do otoczenia (odpowiednie piksele są zaznaczone na zielono), zatem analizowany piksel zachowuje wartość 1.

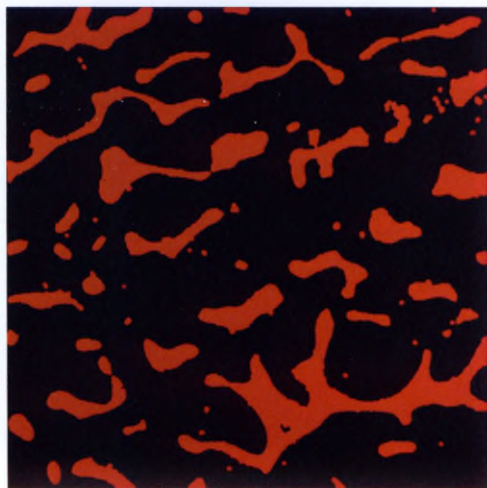
Mamy zatem adaptację wcześniejszej definicji do warunków kwadratowej siatki pikseli. Wydaje się, że szczegółowe analizowanie realizacji dylatacji, otwarcia i zamknięcia w kwadratowej siatce pikseli nie jest w tym momencie niezbędne, a szczegółowe omówienie



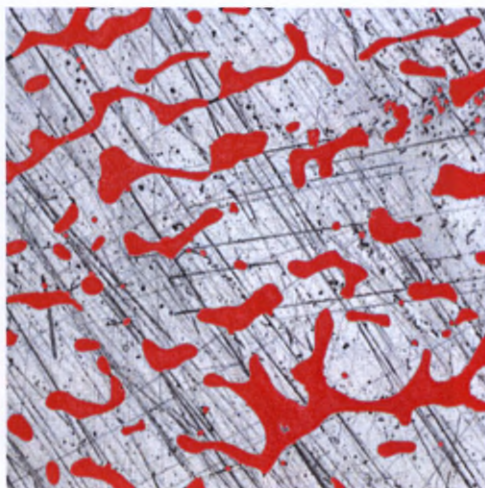
a



b



c



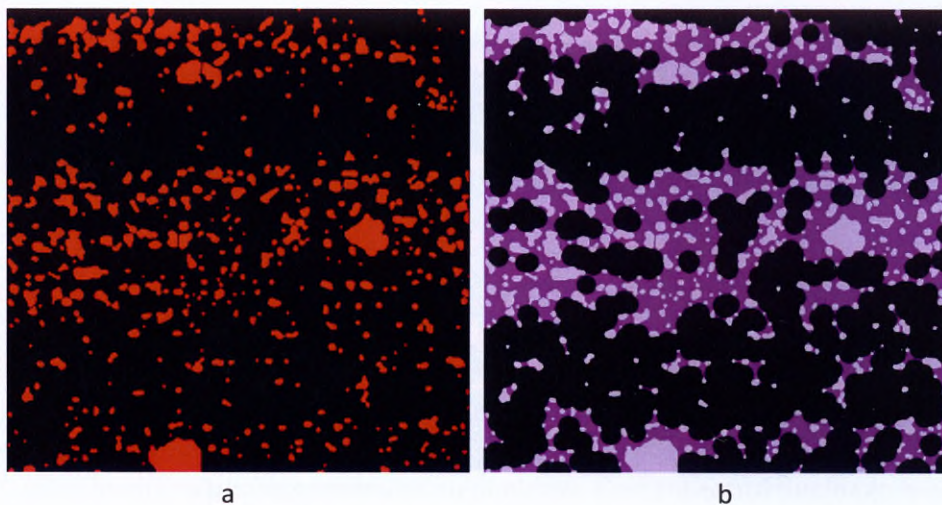
d

erozji powinno być wystarczające. Omawiane przekształcenia mają pewne cechy, które są istotne z praktycznego punktu widzenia i dlatego zostaną tutaj wymienione:

- Erozja jest przeciwieństwem dylatacji. Dylatację można zrealizować za pomocą erozji przez sekwencję trzech poleceń: negatyw – erozja – negatyw.
- Erozja i dylatacja są addytywne. Oznacza to, że erozja elementem strukturalnym o wielkości 2 jest równoważna dwukrotnej erozji o wielkości 1. Ma to istotne znaczenie z punktu widzenia efektywności programu i złożoności odpowiednich algorytmów. Skoro erozja jest addytywna, to wystarczy ją zdefiniować i zaprogramować dla elementu o wartości 1. Nie ma potrzeby programowania przekształcenia dla większych elementów strukturalnych. Dodatkowo, dzięki addytywności przyspieszamy działanie programu. Jeżeli w przykładzie z il. 8.4 każdorazowo analizowaliśmy 9 pikseli, to po zwiększeniu elementu strukturalnego do rozmiaru 5×5 musielibyśmy analizować $5 \times 5 = 25$ pikseli. Tymczasem dzięki addytywności wystarczy przeanalizować $2 \times 9 = 18$ pikseli. Im większy element strukturalny, tym korzyści z redukcji liczby, a tym samym czasu obliczeń, stają się większe.
- Otwarcie i zamknięcie nie są addytywne. Powtórzenie ich z elementem o tej samej wielkości nie wprowadzi do obrazu żadnych zmian.

Warto zapamiętać, że możliwość korekty kształtu, oferowana przez erozję i dylatację, jest cenną cechą, ale wiąże się ze zmianą wielkości obiektu. Dlatego w praktyce często stosuje się otwarcie lub zamknięcie.

Il. 8.5a przedstawia słabej jakości preparat metalograficzny, zawierający ciemne wydzielenia, których obraz jest zakłócony przez dużą liczbę rys. Są one oczywiście



Il. 8.6. Wykorzystanie zamknięcia do detekcji skupisk

widoczne na obrazie binarnym (il. 8.5b), co skutecznie utrudnia wykonywanie jakichkolwiek pomiarów. Ponieważ jednak wady obrazu są przede wszystkim cienkimi liniami, możemy je łatwo usunąć za pomocą otwarcia (il. 8.5c). Końcowy wynik, nałożony na obraz początkowy, pokazuje il. 8.5d. Oczywiście usunięcie rys zostało okupione pewną utratą szczegółów oraz zmianami w kształcie wydzieli. Zmiany te są jednak na tyle niewielkie, że można je pominąć, a efekt usunięcia rys jest bardzo dobry.

Il. 8.6a przedstawia zbiór wydzieli sprawiających wrażenie gromadzenia się w skupiskach, które można uważać za pasma. Jeżeli wykona się na tym obrazie zamknięcie o odpowiedniej wielkości, to najbliższe sobie leżące wydzielienia zostaną skleione i uzyskamy obraz skupisk. Wynik tej operacji, nałożony na obraz początkowy, pokazuje il. 8.6b. Po omówieniu bardziej złożonych operacji wrócimy do tego przykładu i dodatkowo go zmodyfikujemy. Ten prosty sposób detekcji skupisk można wykorzystać np. do wyznaczania na zdjęciach satelitarnych granic obszarów zurbanizowanych lub występowania pewnych upraw.

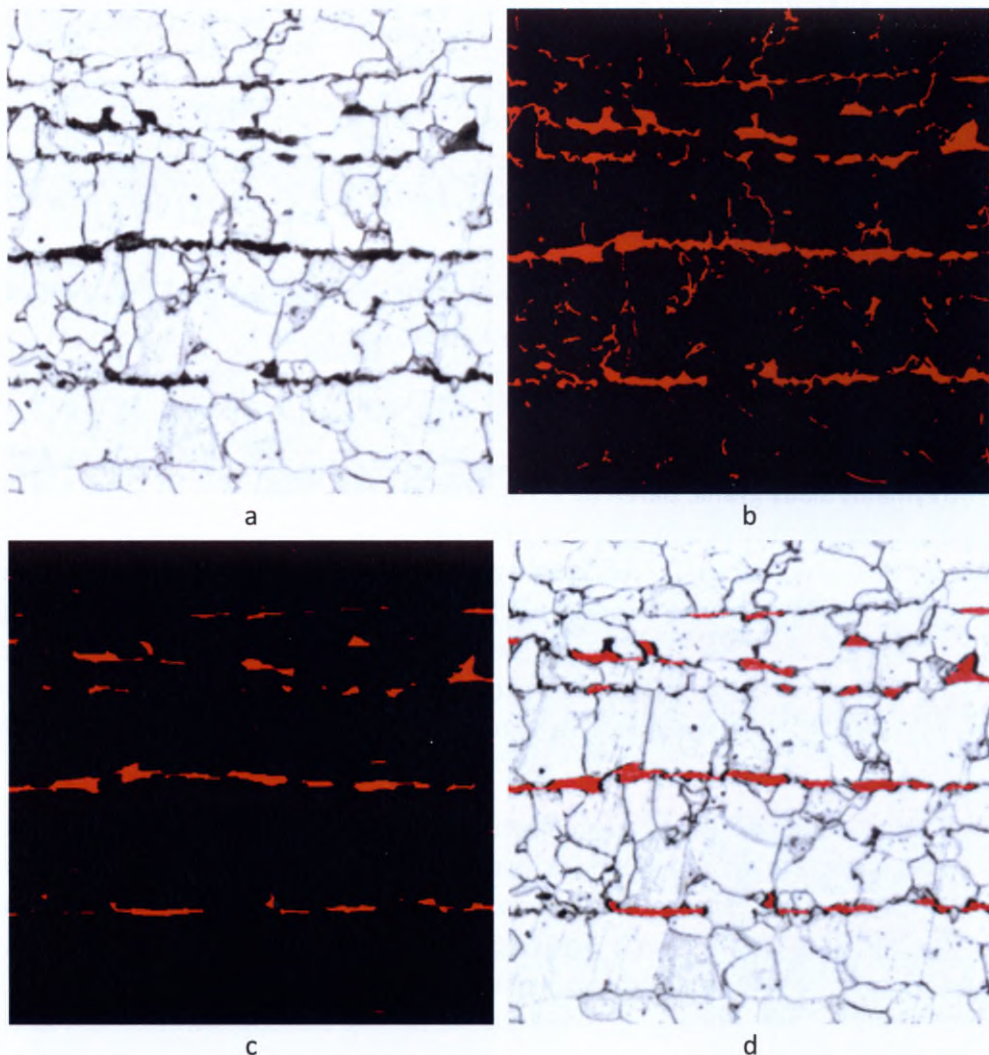
Elementy strukturalne używane podczas erozji nie muszą być symetryczne. Często używa się odcinków linii prostej i wtedy taką erozję nazywamy liniową (ang. *linear erosion*). Do czego można wykorzystać takie przekształcenie? Czasami pojawiają się problemy z elementami liniowymi w obrazie, jak na przykład wtrącenia niemetaliczne, widoczne na il. 8.7a. Jeżeli dokona się binaryzacji tego obrazu, to oprócz wtrąceń otrzymamy ślady granic ziaren (il. 8.7b). Można wtedy wykorzystać erozję z liniowym elementem strukturalnym. Jeżeli ten element będzie zorientowany poziomo, to erozja usunie wszystkie cienkie linie zorientowane pod innymi kątami, a długie elementy poziome trochę skróci (il. 8.7c). Jeżeli zrobimy jeszcze drobny *fine tuning* w postaci usunięcia najmniejszych obiektów na il. 8.7c oraz niewielkiej dylatacji, to otrzymamy całkiem dobry wynik detekcji wydłużonych wtrąceń. Można go zobaczyć na il. 8.7d. Liniowe elementy strukturalne zwykle są definiowane dla czterech kątów: 0° , 45° , 90° oraz 135° .

Zajmiemy się teraz bliżej elementem strukturalnym. Najczęściej jest on definiowany jako kwadratowa matryca o wielkości 3×3 , z centralnym elementem w środku (il. 8.8). W praktyce nie ma potrzeby definiowania środkowego piksela, bo o tym, jaką ma przyjąć wartość, decyduje przepis danego przekształcenia. Element strukturalny definiuje się tylko za pomocą trzech symboli: 1, 0 oraz X:

- 1 oznacza, że element pasuje do obrazu, jeżeli w danym położeniu obraz zawiera piksel o wartości również 1,
- 0 oznacza, że element pasuje do obrazu, jeżeli w danym położeniu obraz zawiera piksel o wartości również 0,

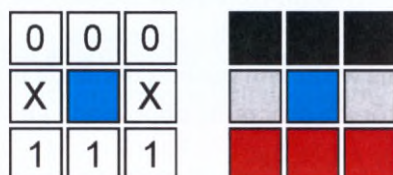
- X oznacza, że element pasuje do obrazu, niezależnie od tego, jaka jest wartość piksela obrazu w danym miejscu. Inaczej mówiąc, piksele te nie są brane pod uwagę podczas operacji. Symbol X pozwala na projektowanie elementów strukturalnych o różnych kształtach, np. koła lub rombu.

Czasami wygodnie jest przedstawić element strukturalny w formie graficznej. Wtedy miejsca z symbolem 1 zaznaczamy kolorem czerwonym, miejsca z symbolem 0 zaznaczamy kolorem czarnym, a odpowiadające symbolowi X – kolorem szarym.

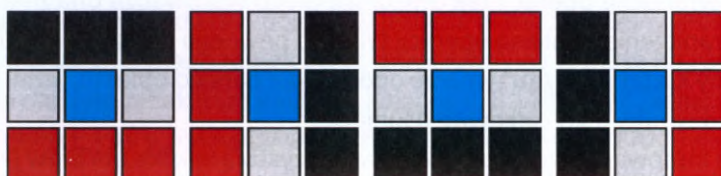


Il. 8.7. Próba detekcji wydłużonych, zorientowanych poziomo wydzieleni: a) obraz początkowy, b) wynik binaryzacji ze zbyt dużymi elementami granic ziaren, c) wynik liniowej erozji obrazu z il. (b), c) końcowy wynik detekcji nałożony na obraz początkowy

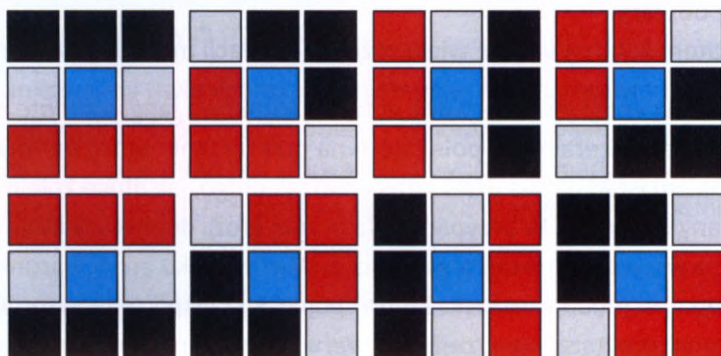
Element strukturalny, pokazany na il. 8.4b, którego użyliśmy do erozji, jest symetryczny. Natomiast element strukturalny z il. 8.8 jest niesymetryczny. Łatwo sobie wyobrazić, że podczas przekształceń dobrze byłoby mieć możliwość korzystania z takiego samego elementu, ale obróconego. Rozwiązano to w ten sposób, że w niektórych przekształceniach wykorzystuje się tzw. wirujący element strukturalny (ang. *rotating structuring element*). Przekształcenie wykonuje się z danym elementem, a następnie powtarza z elementem odwróconym. Mamy w takim przypadku dwa rodzaje sąsiedztwa pikseli: sąsiedztwo czterospójne oraz ośmiozwojne. W czterospójnym najbliższymi sąsiadami są piksele położone na lewo, na prawo, z góry i z dołu w stosunku do analizowanego piksela. W sąsiedztwie ośmiozwojnym za najbliższych sąsiadów uważa się również te piksele, które stykają się z analizowanym pikselem narożnikami (jeżeli piksel modelujemy za pomocą małego kwadratu). W przypadku sąsiedztwa czterospójnego wirujący element strukturalny ma cztery



Il. 8.8. Element strukturalny. Opis w tekście



Il. 8.9. Kolejne położenia wirującego elementu strukturalnego przy czterospójnym sąsiedztwie



Il. 8.10. Kolejne położenia wirującego elementu strukturalnego przy ośmiozwojnym sąsiedztwie

różne konfiguracje (il. 8.9), a w przypadku sąsiedztwa ośmiospójnego wirujący element strukturalny ma osiem różnych konfiguracji (il. 8.10).

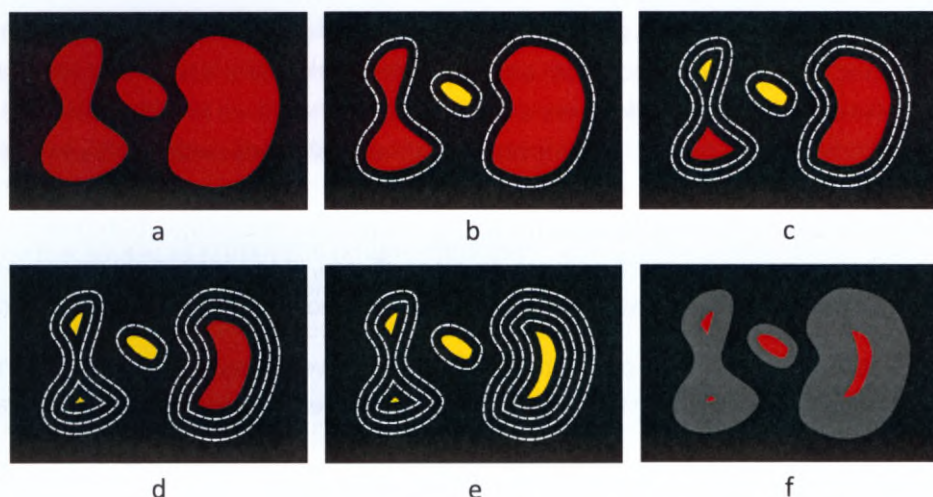
Przekształcenia z wirującym elementem można też przeprowadzić nieco inaczej. Zamiast obracać element po przekształceniu całego obrazu, można go obracać w każdym analizowanym punkcie tak długo, aż będzie pasował do układu pikseli albo do momentu wyczerpania wszystkich konfiguracji elementu strukturalnego.

W tym momencie widoczne stają się pewne podobieństwa do klasycznych filtrów, które zostały zdefiniowane jako wieloargumentowe funkcje przekształcające obraz. Poznane dotąd przekształcenia morfologiczne odpowiadają tej definicji. Mało tego, omówiona przez nas erozja to jednocześnie po prostu filtr minimalny, zaś dylatacja jest filtrem maksymalnym. Można to łatwo sprawdzić, wykorzystując il. 8.4. W przypadku innych przekształceń morfologicznych takiej zgodności z klasycznymi filtrami już nie będzie. Przekształcenia morfologiczne są bardziej selektywne, a także bardziej elastyczne niż klasyczne filtry z kilku powodów:

- Klasyczne filtry działają w każdym punkcie obrazu, a przekształcenia morfologiczne jedynie w tych, których otoczenie odpowiada elementowi strukturalnemu.
- Element strukturalny może być elementem wirującym, co odpowiada 4 lub 8 osobnym filtrom.
- W przekształceniach morfologicznych można definiować dodatkowe warunki, np. powtarzanie operacji tak długo, jak długo wprowadza ona zmiany w obrazie; ograniczenie działania operacji tylko do pewnego obszaru obrazu (może mieć dowolny kształt) lub zatrzymanie operacji dla danego obiektu, jeżeli kolejny krok spowodowałby jego całkowite usunięcie.
- Można budować złożone operacje, wykorzystujące kilka obrazów jednocześnie. Na przykład przekształcamy obraz, wykorzystując element strukturalny oraz inny obraz, który definiuje punkty początkowe operacji lub wyklucza z przekształceń pewne obszary.

Możemy zatem spodziewać się wielu różnych operacji morfologicznych.

Pewną odmianą erozji jest tak zwana erozja warunkowa (ang. *ultimate erosion*). Jest to przekształcenie iteracyjne, polegające na powtarzaniu erozji ograniczonej pewnym warunkiem (stąd nazwa przekształcenia) tak długo, jak długo kolejny krok wprowadza zmiany w obrazie. W przypadku klasycznej erozji doprowadziłoby to po prostu do usunięcia wszystkich pikseli o wartości 1. W przypadku erozji warunkowej zabezpiecza przed tym jeden warunek: żaden spójny zestaw pikseli nie może całkowicie zniknąć. Zatem powtarza się erozję pod warunkiem, że nie spowoduje ona całkowitego zniknięcia jakiegokolwiek obiektu.



Il. 8.11. Erozja warunkowa. Opis w tekście

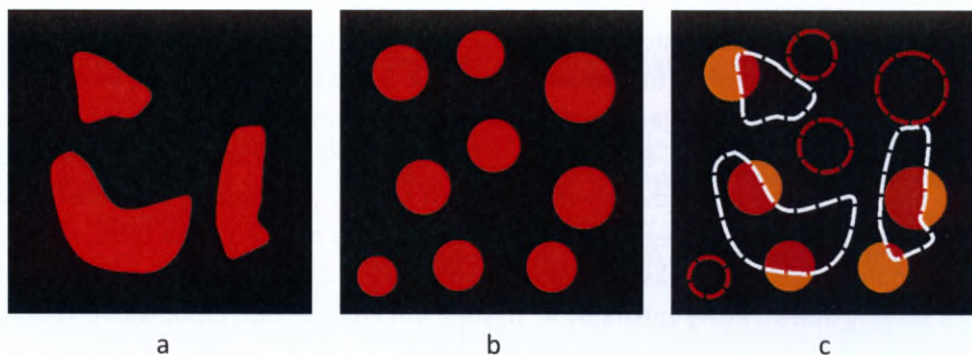
Ideę erozji warunkowej pokazuje il. 8.11. Zaczynamy od obrazu zawierającego trzy obiekty (il. 8.11a). Po pierwszym kroku erozji dostajemy trzy zmniejszone obiekty (il. 8.11b). Najmniejszy z nich jest tak mały, że kolejny krok erozji usunąłby go całkowicie. Dlatego tego obiektu, zaznaczonego na il. 8.11b kolorem żółtym, w dalszych krokach już nie modyfikujemy. Kolejny krok erozji powoduje rozpad obiektu z lewej strony na dwa obiekty (il. 8.11c), co jest skutkiem silnego przewężenia pierwotnego obiektu. Jeden z tych dwóch obiektów jest tak mały, że kolejny krok erozji usunąłby go całkowicie. Dlatego ten element, zaznaczony na il. 8.11c na żółto, również nie podlega dalszym przekształceniom. Z kolei na il. 8.11d widać zaznaczony na żółto kolejny obiekt, który zniknąłby w następnym kroku erozji. Erodowanie kończymy w momencie pokazanym na il. 8.11e, gdyż nie ma już tak dużych obiektów, żeby mogły być nadal erodowane bez ryzyka całkowitego ich usunięcia. Il. 8.11f pokazuje końcowy wynik erozji warunkowej (ang. *ultimate eroded set*), nałożony na ciemnoszare obszary początkowych obiektów.

Erozja warunkowa ma dwie charakterystyczne cechy: pozostawia z każdego obiektu pewien minimalny zbiór pikseli i obiekty wykazujące lokalne przewężenia dzieli na kilka zbiorów. Dzięki temu jest używana do tworzenia markerów (początkowych obiektów) przy operacjach rozdzielania sklejonych części lub uzupełniania brakujących fragmentów granic pomiędzy obiektami.

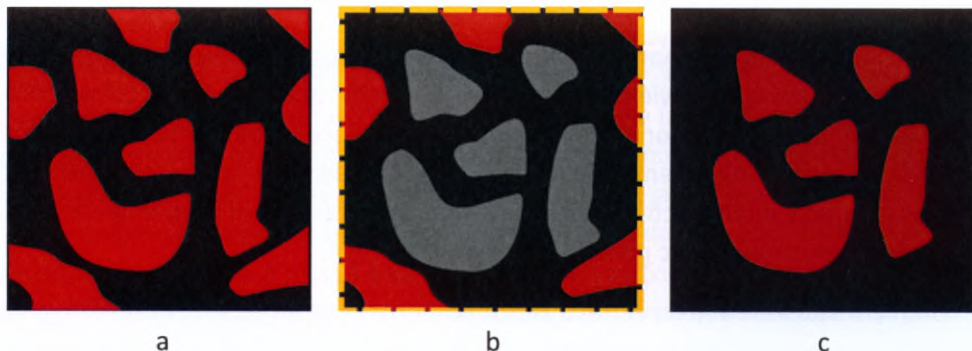
Kolejnym sposobem poszerzenia możliwości przekształceń morfologicznych są operacje geodezyjne, które można określić jako operacje z dodatkowym warunkiem ograniczającym. Nazwa tej grupy operacji nawiązuje wprost do geodezji, czyli

w uproszczeniu do nauki o geometrycznych własnościach powierzchni Ziemi. Istnieje pojęcie „drogi geodezyjnej”, którą można określić jako najkrótsza droga omijająca przeszkody. Takie właśnie drogi wyszukują programy nawigacji satelitarnej. Jeżeli za przeszkody uznamy opłaty za niektóre drogi lub spowolnienia jazdy i korki, to możemy drogę geodezyjną wybierać na kilka sposobów, np. najkrótsza, najszybsza, bez odcinków płatnych, omijająca promy itp. W przypadku morfologii matematycznej przeszkody wprowadzamy za pomocą drugiego obrazu, który nazywamy maską (ang. *mask*). Maską definiuje dozwolony obszar, w którym mogą zachodzić przekształcenia.

Prześledźmy ten proces na przykładzie il. 8.12. Chcemy wykonać dylatację geodezyjną il. 8.12a przy obecności maski z il. 8.12b. W celu ułatwienia zrozumienia całej operacji na il. 8.12c mamy zaznaczone zarysy obiektów (biała przerywana linia) i maski (czerwona przerywana linia). Pierwszym krokiem w omawianym przekształceniu jest wyznaczenie wspólnej części (logiczne AND) obiektów i maski. Otrzymujemy wtedy obiekty zaznaczone na il. 8.12c kolorem czerwonym. Możemy teraz wykonywać dylatację tych czerwonych obiektów, ale nigdy nie przekroczymy granic maski. Po prostu, po każdym



Il. 8.12. Dylatacja geodezyjna. Opis w tekście



Il. 8.13. *Border kill*. Opis w tekście

kroku dylatacji wyznaczamy wspólną część z maską. Jeżeli wykonamy dostatecznie dużo kroków dylatacji, to otrzymamy pięć kółek, widocznych na il. 8.12c. Czerwonym kolorem oznaczone są elementy początkowe, a pomarańczowym największy obszar, który może powstać w wyniku dylatacji geodezyjnej.

Wykorzystując operacje geodezyjne, możemy zdefiniować kolejne przekształcenia, które w istocie są krótkimi makropoleceniami:

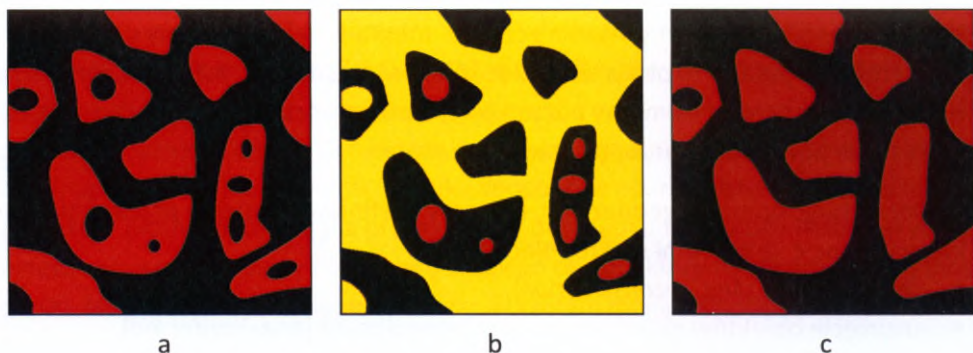
- rekonstrukcję (ang. *reconstruction*),
- usunięcie obiektów przeciętych przez brzeg obrazu (ang. *border kill*),
- wypełnienie (zalanie) otworów (ang. *hole fill*).

Rekonstrukcja to dylatacja geodezyjna, powtarzana tak długo, aż kolejny krok nie przyniesie zmian w przetwarzanym obrazie. Pełne kółka na il. 8.12c to właśnie wynik rekonstrukcji. Zbiór kółek z il. 8.12b został zrekonstruowany (częściowo odtworzony) w ten sposób, że odtworzono te kółka, które miały jakieś wspólne piksele z obrazem początkowym. Mówimy też, że obraz na il. 8.12a zawiera markery do rekonstrukcji, bo od jego zawartości zależy, które obiekty zostaną zrekonstruowane, a które nie.

Rekonstrukcję wykorzystuje się często np. do usuwania z obrazu drobnych obiektów. Oczywiście, do usunięcia takich obiektów można wykorzystać też poznane już otwarcie. Erozja usunie małe obiekty, a późniejsza dylatacja przywróci pozostałym ich początkową wielkość. Niestety, ubocznym skutkiem takiej operacji jest zmiana kształtu obiektów. Natomiast, jeżeli obraz po erozji wykorzysta się jako zbiór markerów, a początkowy obraz jako maskę, to dzięki rekonstrukcji otrzymamy obiekty w takim kształcie, w jakim znajdowały się na obrazie wyjściowym.

Rekonstrukcję można wykorzystać do zbudowania operacji usuwającej obiekty przecięte przez brzeg obrazu (ang. *border kill*). Wyjściowy obraz stanowi maskę (il. 8.13a), natomiast markerem jest zbiór pikseli leżących na brzegu obrazu (żółta przerywana linia na il. 8.13b). W wyniku rekonstrukcji otrzymamy wyłącznie obiekty mające jakieś punkty położone na brzegu obrazu (czerwone obiekty na il. 8.13b). Odejmując te obiekty od początkowego obrazu, otrzymujemy obraz pozbawiony obiektów przeciętych przez brzeg (il. 8.13c).

Nieco podobny przebieg ma operacja wypełniania otworów (ang. *hole fill*). Z wyjściowego obrazu (il. 8.14a) wykonujemy negatyw, a następnie przeprowadzamy opisaną powyżej operację *border kill*. W ten sposób usuwamy wszystkie obiekty zaznaczone żółtym kolorem na il. 8.14b. Pozostaną tylko obiekty zaznaczone na tym rysunku kolorem czerwonym. Gdy dodamy je do obrazu początkowego, otrzymamy obraz z zalanymi otworami (il. 8.14c). Warto zwrócić uwagę, że jeśli otwór w obiekcie



Il. 8.14. Wypełnianie otworów (*hole fill*). Opis w tekście

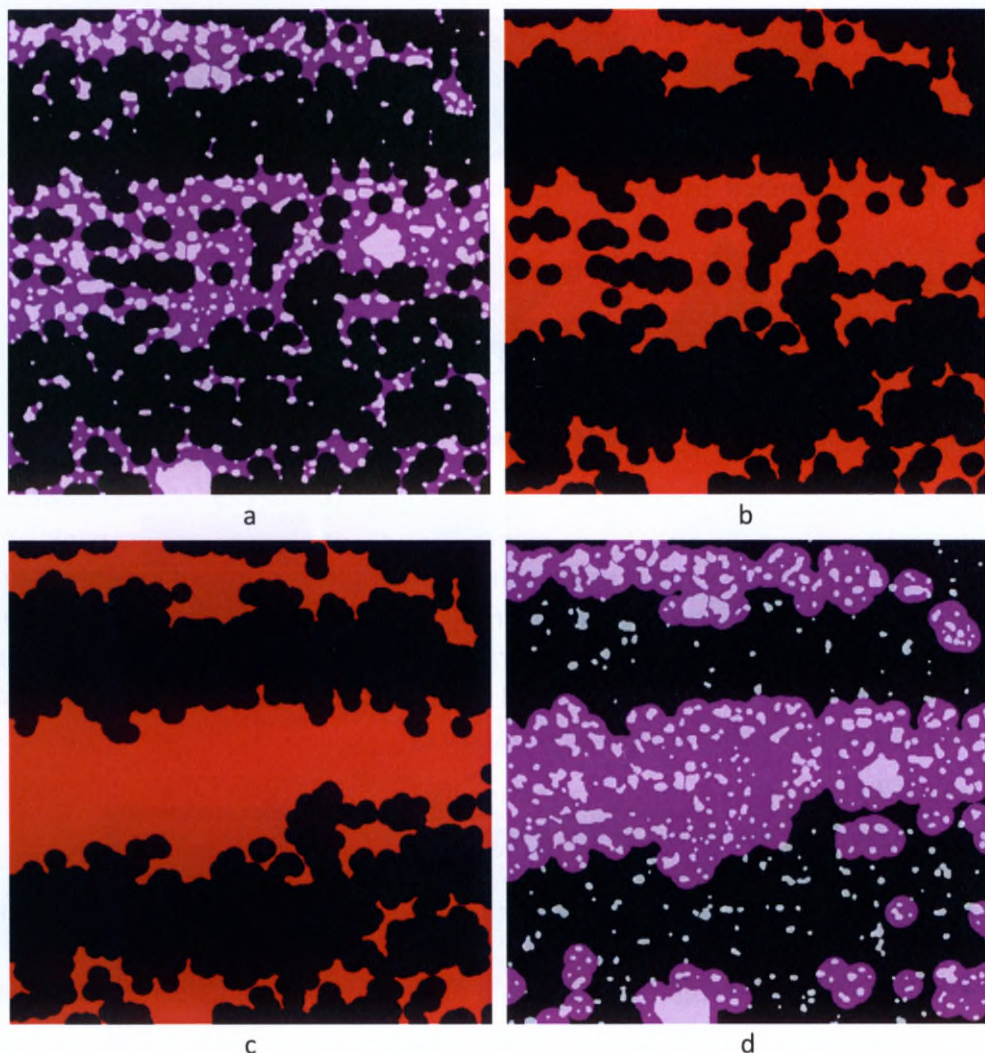
przecina lub dotyka brzegu figury, to nie zostanie zalany (pozostały takie otwory w obiektach z lewej strony il. 8.14c).

Wykorzystując opisane powyżej nowe operacje morfologiczne, możemy zmodyfikować obraz skupisk obiektów z il. 8.6a. Uzyskany za pomocą operacji zamknięcia obraz skupisk pokazuje ponownie il. 8.15a. Dzięki operacji rekonstrukcji udaje się usunąć drobne elementy, których nie uznajemy za skupiska. Pozostają jedynie duże obiekty, jednak ich kształt jest bardzo nieregularny (il. 8.15b). Można ten kształt łatwo skorygować przez zalanie otworów (il. 8.15c). Efekt wydaje się być dość dobry, ale nadal zewnętrzne brzegi obszarów skupisk pozostają bardzo nieregularne. Można je wygładzić przez operację otwarcia. Kończącym krokiem powinna być dylatacja, która powiększa obszar skupiska i sprawia, że należące do niego wydzielenia znajdują się zawsze wewnątrz skupiska, a nie leżą na jego brzegu, jak część wydzieleń na il. 8.15a. Końcowy wynik naszych działań pokazano na il. 8.15d. Obszar skupisk ma kolor purpurowy, a jasne plamy to miejsca wydzieleń.

Powyższe przykłady, choć dopiero zaczynamy poznawać przekształcenia morfologiczne, pokazują już, jak duże są ich możliwości. Takich operacji, jak tworzenie znaczników za pomocą erozji warunkowej, eliminacja części obiektów z zachowaniem w niezmienionej formie pozostałych, usuwanie obiektów przeciętych przez brzeg obrazu czy zalewanie otworów, nie jesteśmy w stanie wykonać za pomocą klasycznych filtrów.

Kolejnym względnie prostym przekształceniem jest ścienianie (ang. *thinning*). Polega ono na przyjmowaniu wartości 0 przez piksele, których otoczenie jest zgodne z elementem strukturalnym. Jest to podstawowa różnica w stosunku do erozji, w której piksel przyjmował wartość 0, jeżeli jego otoczenie nie było zgodne z elementem strukturalnym.

Konsekwencją tych różnic w definicji jest oczywiście różnica w wynikach operacji. Jeżeli przeprowadzimy pojedynczy krok erozji elementem pokazanym na il. 8.4b,



II. 8.15. Korekta obrazu skupisk. Opis w tekście

to usuniemy ze wszystkich obiektów pojedynczą warstwę pikseli. Jeżeli natomiast przeprowadzimy ścienianie tym samym elementem, to usuniemy całe wnętrza obiektów, czyli pozostawimy dokładnie te piksele, które usuwa erozja. Inaczej mówiąc, wygenerowalibyśmy obraz brzegu poszczególnych obiektów.

Przeciwieństwem ścieniania jest pogrubianie, czyli przypisywanie wartości 1 tym pikselom, których otoczenie spełnia warunki określone przez element strukturalny. Mamy tutaj pewną analogię do dyatacji, ale wyniki dla tych samych elementów strukturalnych będą zdecydowanie różne.



Il. 8.16. Szkielety wybranych figur

Poznamy teraz pewne informacje o szkieletach figur. Szkielety są zbiorami środków wszystkich kół wpisanych w figurę, opierających się co najmniej dwoma punktami o brzeg figury. Druga, równorzędna definicja mówi, że jest to zbiór punktów figury równoodległych od dwóch punktów na jej brzegu. W ten sposób (il. 8.16) szkieletem koła jest punkt, szkieletem kwadratu – jego przekątne, a szkieletem trójkąta – dwusieczne jego kątów na odcinku od wierzchołka trójkąta do punktu przecięcia dwusiecznych. Szkieletem pierścienia jest okrąg.

Własności szkieletów można opisać następująco:

- Szkielet figury leży zawsze w jej wnętrzu.
- Szkielet składa się z linii i nie ma pola powierzchni.
- Jeżeli w figurze występują otwory, to szkielet utworzy tyle zamkniętych pętli, ile jest otworów.
- Im bardziej postrzępiony jest brzeg figury, tym więcej gałęzi znajduje się w szkielecie.

Operację pozwalającą na wyznaczenie szkieletu figury nazywamy szkieletyzacją (ang. *skeletonization*). Można ją przeprowadzić na przykład za pomocą elementu strukturalnego, pokazanego na il. 8.17. Element ten został nazwany w tzw. alfabecie Go-
laya literą L i dlatego otrzymany z jego pomocą szkielet bywa nazywany szkieletem L. Prześledzimy teraz tworzenie szkieletu przy użyciu wirującego, czterospójnego elementu L. Kolejne konfiguracje tego elementu zostały pokazane na il. 8.9, ale dla lepszej czytelności powtarzamy je na il. 8.17.

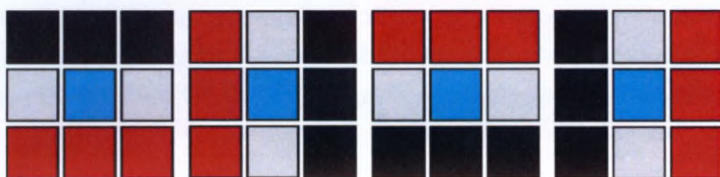
Algorytm generowania szkieletu jest następujący:

- Przykładamy element strukturalny do wszystkich pikseli obrazu. Jeżeli lokalna konfiguracja pikseli jest zgodna z elementem strukturalnym, to piksel

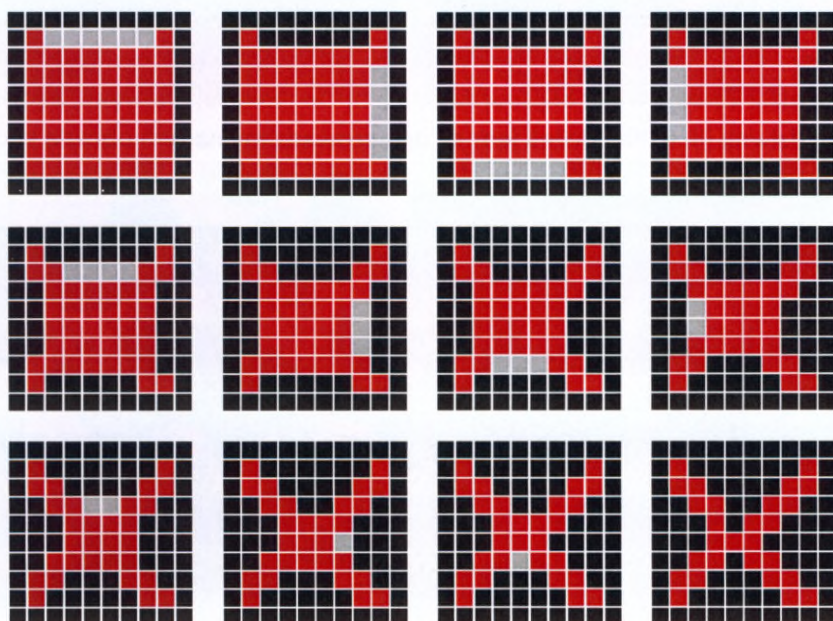
odpowiadający lokalizacji centralnego punktu z elementu strukturalnego przyjmuje wartość 0. Wykonujemy zatem ścienianie.

- Powtarzamy operację dla kolejnej konfiguracji wirującego elementu strukturalnego.
- Jeżeli kolejna iteracja nie wprowadza zmian do obrazu, budowa szkieletu jest zakończona.

Na il. 8.18 pokazano (od lewej do prawej w kolejnych wierszach) kolejne kroki budowy szkieletu kwadratu. Przy każdym kroku zaznaczono szarym kolorem te piksele, które powinny przyjąć wartość 0 ze względu na zgodność otoczenia z elementem strukturalnym. Prawy element dolnego rzędu składowych il. 8.18 nie ma zaznaczonego żadnego punktu szarym kolorem. Wynika to z faktu, że żaden punkt



Il. 8.17. Kolejne położenia wirującego elementu strukturalnego L przy czterospójnym sąsiedztwie



Il. 8.18. Powstawanie szkieletu kwadratu. Opis w tekście

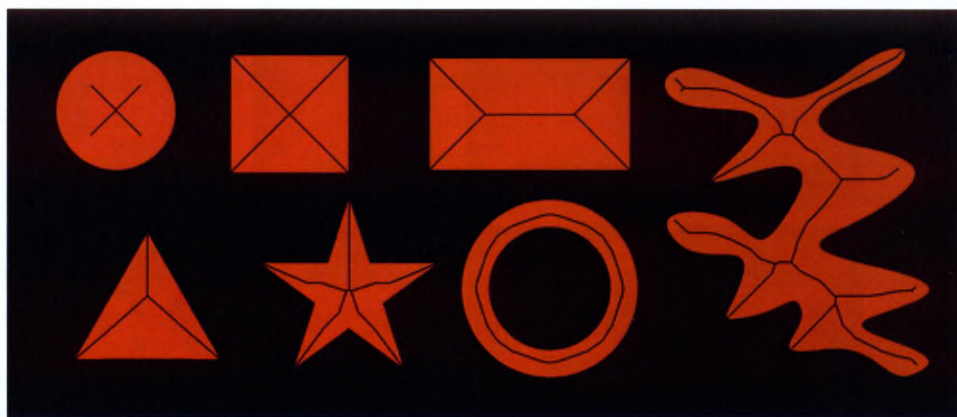
nie spełniał zgodności z elementem strukturalnym, a konkretnie jego konfiguracją pokazaną z prawej strony il. 8.17. Kolejny krok (obróć elementu) nie wprowadzi zmian do obrazu, zatem generowanie szkieletu zostało ukończone.

Można zatem powiedzieć, że szkieletyzacja to ścienianie obrazu odpowiednio dobranym wirującym elementem strukturalnym, prowadzone tak długo, jak długo kolejny krok wprowadza zmiany do obrazu.

Il. 8.19 przedstawia kilka figur wraz z nałożonymi na nie szkieletami. Jak widać, szkielety te są dobrym przybliżeniem szkieletów matematycznych, widocznych na il. 8.16. Jedynym wyjątkiem jest koło, w którym szkielet nie jest pojedynczym punktem, tylko przybiera kształt litery X.

Element strukturalny L nie jest jedynym elementem, który można wykorzystać do budowy szkieletu. Ma jednak istotną cechę, polegającą na tym, że tworzy szkielety o grubości jednego piksela, dobrze modelujące szkielety matematyczne. Inne elementy strukturalne prowadzą zazwyczaj do szkieletów bardziej odbiegających od teoretycznego modelu.

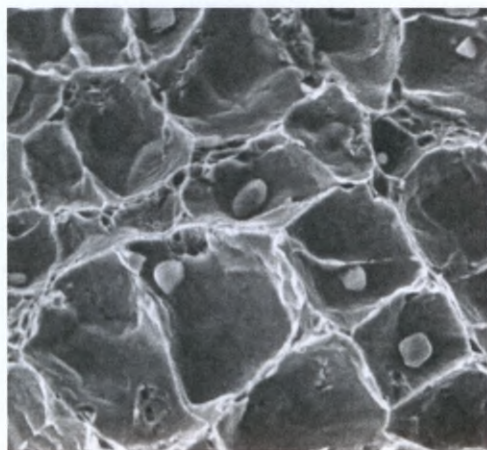
Szkieletyzację można wykorzystać wszędzie tam, gdzie potrzeba stworzyć cienkie elementy liniowe. Przykładem może być próba prostej detekcji granic pustek na powierzchni przełomu, obserwowanej za pomocą mikroskopu skaningowego. Granice pustek są widoczne na il. 8.20a jako najjaśniejsze elementy. Najprostsza próba detekcji drogą binaryzacji (il. 8.20b) nie daje satysfakcjonującego wyniku. Możemy wprowadzić usunąć wiele błędów za pomocą zamknięcia, ale wynik będzie co najmniej dziwny (il. 8.20c). Jeżeli jednak przeprowadzimy na tym obrazie szkieletyzację, to wynik tej operacji, nałożony na obraz początkowy, wykaże poprawną detekcję



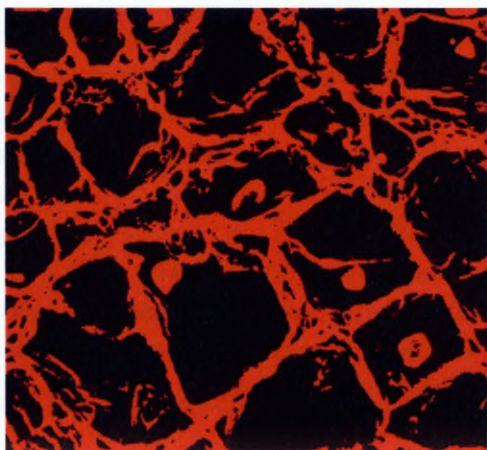
Il. 8.19. Przykłady szkieletów nałożonych na różne figury

dużej części granic pomiędzy pustkami (il. 8.20d). Przedstawiony na il. 8.20 wynik detekcji pokazuje potencjalne możliwości szkieletyzacji i nie należy go oceniać pod kątem poprawności samej detekcji, którą można przeprowadzić znacznie lepiej, ale wymaga to bardziej złożonego algorytmu.

Ponieważ szkielek może mieć grubość pojedynczego piksela, to można poszukiwać w nim punktów o pewnych cechach, np. punktów odizolowanych, leżących na końcach gałęzi szkieletu lub w jego węzłach. Jeżeli zamiast szkieletu będziemy analizowali figury, możemy szukać punktów leżących wewnątrz figury (będzie to odpowiadało pojedynczemu krokowi erozji) lub elementów leżących na brzegu (co z kolei będzie operacją analogiczną do ścieniania elementem strukturalnym z il. 8.4b).



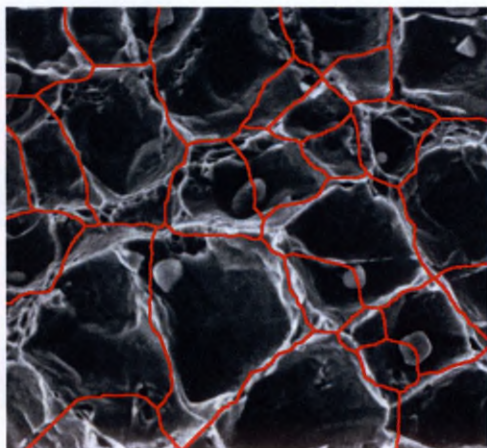
a



b

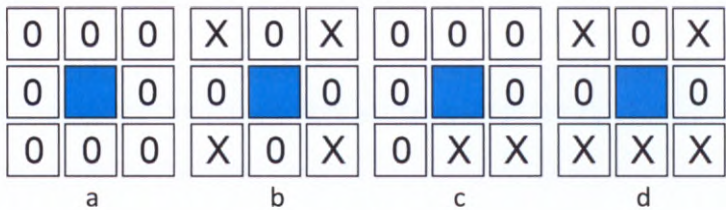


c



d

Il. 8.20. Detekcja granic pustek na przełomie. Opis w tekście

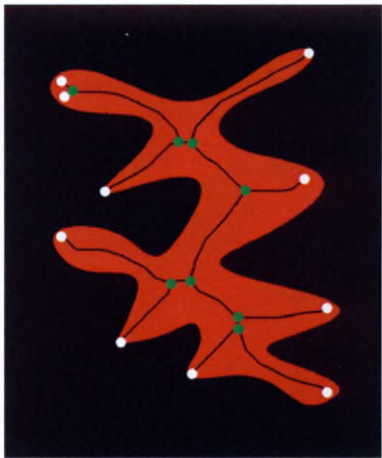


Il. 8.21. Elementy strukturalne do wykrywania odizolowanych punktów:
a) dla sąsiedztwa ósmiospójnego i b) dla sąsiedztwa czterospójnego, oraz punktów
końcowych: c) dla sąsiedztwa ósmiospójnego i d) dla sąsiedztwa czterospójnego.
Elementy strukturalne (c) oraz (d) są elementami wirującymi

Przekształcenie realizujące opisywane procesy możemy określić terminem trafi-nie-trafi (ang. *hit or miss*). Czasami spotyka się akronim HMT (*Hit or Miss Transformation*).

W przypadku sąsiedztwa ósmiospójnego wykrycie odizolowanych punktów nie stanowi żadnego problemu, jeśli użyjemy elementu strukturalnego, widocznego na il. 8.21a. Jego konstrukcja jest bardzo intuicyjna. Jeżeli przejdziemy do sąsiedztwa czterospójnego, to element ulegnie zmianie i przybierze postać z il. 8.21b. Nietrudno sobie wyobrazić, że w przypadku szachownicy złożonej z pojedynczych pikseli wszystkie piksele niestykające się z brzegiem zostaną w przypadku sąsiedztwa czterospójnego rozpoznane jako punkty odizolowane. Podobne problemy mogą towarzyszyć wykrywaniu punktów końcowych. W większości przypadków przedstawione elementy strukturalne dają jednak poprawne wyniki.

Wykrywanie punktów węzłowych, nazywanych też punktami potrójnymi (ang. *triple points*), wymaga innych elementów strukturalnych, w skład których wchodzi różne



Il. 8.22. Szkielet naniesiony na figurę, z której powstał, z zaznaczonymi na biało punktami końcowymi i na zielono punktami węzłowymi

kombinacje trzech jedynek oraz symboli X i 0. Liczby punktów węzłowych i końcowych szkieletu figury (il. 8.22) bywają używane do oceny kształtu figur.

Możliwość detekcji punktów końcowych prowadzi do kolejnego przekształcenia, jakim jest obcinanie lub skracanie gałęzi (ang. *pruning*). W przypadku szkieletów nie można obcinać gałęzi w obrębie zamkniętych pętli, bo nie występują w nich punkty końcowe. W ten sposób można wygenerować markery obiektów z otworami i wykorzystać je do rekonstrukcji tych obiektów. Dzięki temu możemy również spośród wszystkich obiektów wybrać te, które mają otwory. Oczywiście można to zrobić za pomocą innych przekształceń morfologicznych. Przykładowy algorytm mógłby być następujący:

- wykonanie negatywu obrazu,
- *border kill* (pozostaną tylko otwory),
- dylatacja o wielkości 1,
- logiczne AND obrazu po dylatacji oraz obrazu początkowego (dostaniemy krawędzie otworów, składające się z pikseli leżących wewnątrz obiektów),
- wykorzystanie tego obrazu jako markera do rekonstrukcji.

Powyższa sekwencja została pokazana, aby zilustrować różnorodność możliwości dawanych przez operacje morfologiczne.

Analizując rozmieszczenie obiektów lub punktów na obrazie, można każdemu z nich przypisać pewną strefę wpływów, która przypomina nieco ustalanie wód terytorialnych poszczególnych państw. Zwykle przyjmuje się, że strefa wpływu danego punktu lub obiektu zawiera wszystkie punkty, których odległość jest najmniejsza właśnie do tego punktu lub obiektu. Wynika stąd, że granica pomiędzy dwoma punktami powinna być symetryczną łączącego je odcinka. Problem budowy odpowiedniego zestawu granic jest znany jako triangulacja Woronoja (w literaturze angielskojęzycznej Voronoi). Często symuluje się ten proces za pomocą rozrostu obiektów i nazywa go aparatem komórkowym.

Odpowiednie przekształcenie można wykonać również za pomocą operacji morfologicznych. Nosi ono nazwę SKIZ, co jest akronimem angielskojęzycznego określenia *Skeleton by Influence Zone*, co można przetłumaczyć jako „szkielet określony przez strefę wpływu”. Tłumaczenie zresztą nie jest dobre, bo *de facto* podczas tego przekształcenia wykonuje się szkieletyzację obszaru stanowiącego sumę wszystkich stref wpływów, a ostatecznym wynikiem jest negatyw obrazu tego szkieletu, czyli po prostu zbiór poszczególnych stref wpływów. Nie wprowadzono żadnego odpowiednika dla akronimu SKIZ w języku polskim.

Jak wykonuje się SKIZ? Zwykle wystarczy wybrać w oprogramowaniu odpowiednie predefiniowane przekształcenie, ale można wykonać też tę operację krok po kroku



Il. 8.23. Negatyw obrazu obiektów oraz ich szkieletów (a) oraz negatyw szkieletu z il. (a) po obcięciu gałęzi, czyli SKIZ wyjściowego obrazu

za pomocą poznanych już wcześniej przekształceń. Najpierw wykonuje się negatyw wyjściowego obrazu, a następnie przeprowadza jego szkieletyzację. Połączone obrazy szkieletów oraz wyjściowych obiektów pokazano na il. 8.23a. Następnie obcina się gałęzie tak długo, aż kolejny krok nie wprowadzi zmian do obrazu. Ponieważ szkielet koła nie posiada zamkniętych pętli, to szkielety wszystkich otworów występujących w obrazie początkowym zostaną zlikwidowane. Ostatnim krokiem jest wykonanie negatywu oczyszczonego z gałęzi szkieletu, dzięki czemu otrzymujemy gotowy wynik SKIZ (il. 8.23b).



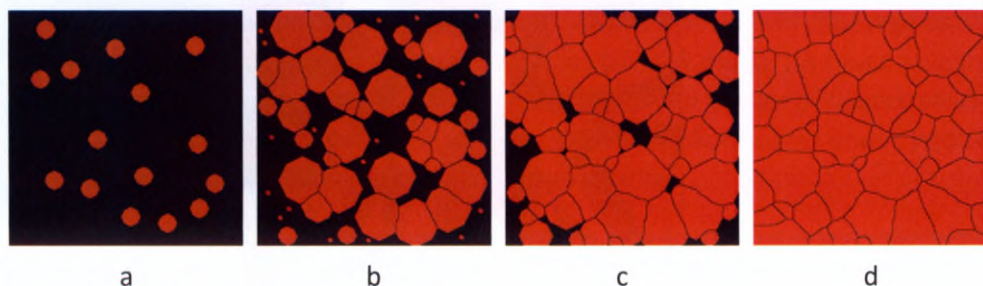
Il. 8.24. *Thickening* obrazu z il. 8.14a za pomocą elementu strukturalnego szkieletyzacji.
Opis w tekście

SKIZ jest wykorzystywany w różnych procesach symulacji wzrostu komórek, obszarów zurbanizowanych, zmian szaty roślinnej itp. Symulacje można komplikować przez wykonywanie tylko kilku kroków wzrostu obiektów, przerwanie procesu, dodanie nowych punktów, kolejny częściowy wzrost, znowu dodanie punktów itd. Pomocne w tym procesie jest pogrubianie (*thickening*) z użyciem elementu strukturalnego wykorzystywanego podczas szkieletyzacji. Poszczególne elementy mogą się powiększać, ale nie zlewają się ze sobą tak, jak to się dzieje podczas dylatacji (il. 8.24). Proces ten można byłoby nazwać również dylatacją bez dotykania. Kilka etapów opisanej powyżej symulacji ukazuje il. 8.25.

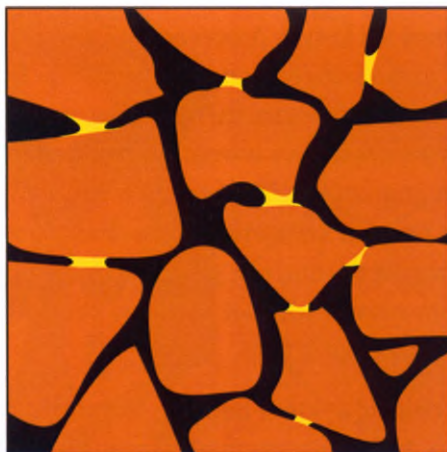
Na il. 8.25 mamy kilka obiektów, które będą się rozrastały. Il. 8.25b pokazuje rozrósnięte obiekty z il. 8.25a, kilka mniejszych obiektów, które pojawiły się podczas poprzednich kroków, oraz dodane losowo zarodki kolejnych komórek. Na il. 8.25c widać zapełnianie powierzchni z powodu powiększających się komórek. Widać, że komórki ograniczają możliwości rozrostu sąsiadów. Il. 8.25d ilustruje końcowy wynik symulacji.

Przegląd operacji morfologicznych, dostępnych dla obrazów binarnych, zakończymy omówieniem detekcji działów wodnych (ang. *watershed segmentation*). W tej procedurze pojawia się co prawda obraz pośredni w odcieniach szarości, ale zarówno obraz wyjściowy, jak i końcowy wynik powinny być obrazami binarnymi. Dodatkowym argumentem za omówieniem detekcji działów wodnych w obrazach binarnych jest fakt, że systemy analizy obrazu oferują procedury przeznaczone wyłącznie dla obrazów binarnych, które w tle wykorzystują detekcję działów wodnych. Przykładem może być procedura SplitConvex w systemie Aphelion.

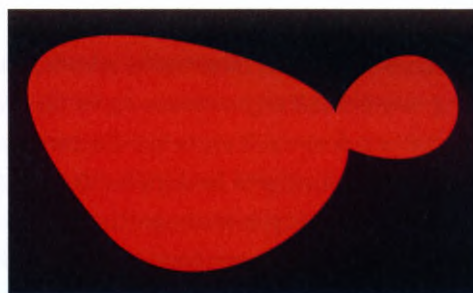
W przypadku obrazów binarnych detekcję działów wodnych wykorzystuje się do podziału sklejonych obiektów. Takie zadanie realizuje wymieniona już procedura SplitConvex, przy czym wynikiem jej działania nie zawsze jest zbiór wyłącznie wypukłych obiektów. Wynikowe obiekty mogą być wklęsłe, o ile nie zawierają lokalnych



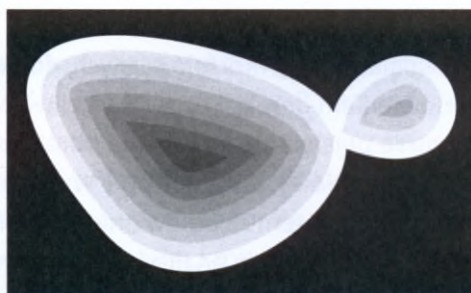
Il. 8.25. Kilka etapów symulacji wzrostu kolonii komórek, wykorzystującej operacje morfologiczne



Il. 8.26. Zestaw obiektów, które będą modyfikowane za pomocą detekcji działów wodnych.
Opis w tekście

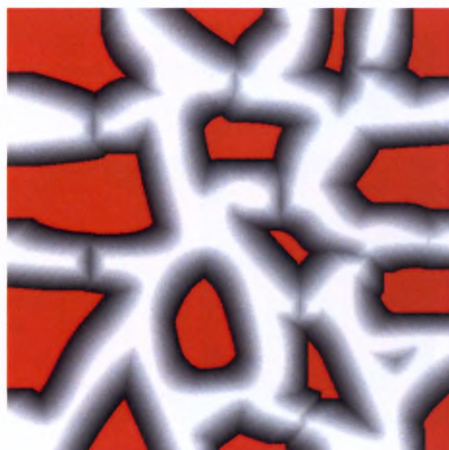


a



b

Il. 8.27. Przygotowanie obrazu do detekcji działów wodnych. Opis w tekście



Il. 8.28. Generowanie działów wodnych dla binarnej wersji obrazu z il. 8.26. Opis w tekście

przewężeń. W praktyce rzeczywiście najczęściej mamy do czynienia ze zbiorami sklejonych wypukłych obiektów, które trzeba rozdzielić. To tłumaczy pochodzenie i usprawiedliwia nazwę procedury, choć jej wynik nie zawsze odpowiada sugerowanemu przez jej nazwę zbiorowi ściśle wypukłych obiektów.

Il. 8.26 zawiera pomarańczowe obiekty, częściowo połączone za pomocą żółtych mostków. Binarna wersja tego obrazu (il. 8.29a) posłuży do demonstracji procesu rozdzielania sklejonych obiektów przy użyciu detekcji działów wodnych.

Drugim obszarem zastosowań detekcji działów wodnych w obrazach binarnych jest uzupełnianie brakujących odcinków siatek, stanowiących granice zbiorów różnego rodzaju obiektów. Warto jednak zwrócić uwagę, że jest to jedynie inna definicja tego samego problemu. Jeżeli binarny obraz niepełnej siatki granic pomiędzy obiektami zamienimy na jego negatyw, to otrzymamy obraz częściowo sklejonych obiektów.

Ideę przebiegu detekcji działów wodnych można wyjaśnić na podstawie il. 8.27. Widać na nim dwa sklezione obiekty (il. 8.27a). Jeżeli potraktujemy je jako obraz linii brzegowej dwóch połączonych jezior o głębokości proporcjonalnej do odległości od brzegu, to możemy wygenerować obraz izobat (linii punktów o tej samej głębokości). Przyjmując, że teren poza obrębem jezior jest poziomy i ma najwyższą wysokość, to można przypisać mu wartość 255, a poszczególne izobaty będą miały coraz to niższe wartości. Obraz takich izobat pokazano na il. 8.27b. Dla poprawy czytelności obszar na zewnątrz jezior ma kolor czarny, a nie biały.

Detekcja działów wodnych polega na wypełnianiu niecek jezior przy zachowaniu stałego poziomu wody. Oznacza to, że generujemy obraz lustra wody (czerwone obiekty na il. 8.28) jako wynik binaryzacji przy systematycznie podnoszonym progu binaryzacji. Jeżeli obiekty nie były sklezione ze sobą, to binaryzacja z progiem 254 dałaby obraz tych obiektów. Na wynik binaryzacji nakładamy jednak dodatkowe ograniczenie, które nie pozwala na sklejenie się lusterek wody z oddzielnych jezior. Można to uzyskać przez odpowiedni dobór operacji pogrubiania (zob. il. 8.24 oraz jego opis). Jeżeli opisywane postępowanie będziemy prowadzili tak długo, jak długo kolejny krok wprowadza zmiany do obrazu, to otrzymamy na końcu granice pomiędzy obiektami, które modelują dział wodny między jeziorami. Uzyskany obraz granic będzie miał oczywiście grubość zaledwie pojedynczego piksela.

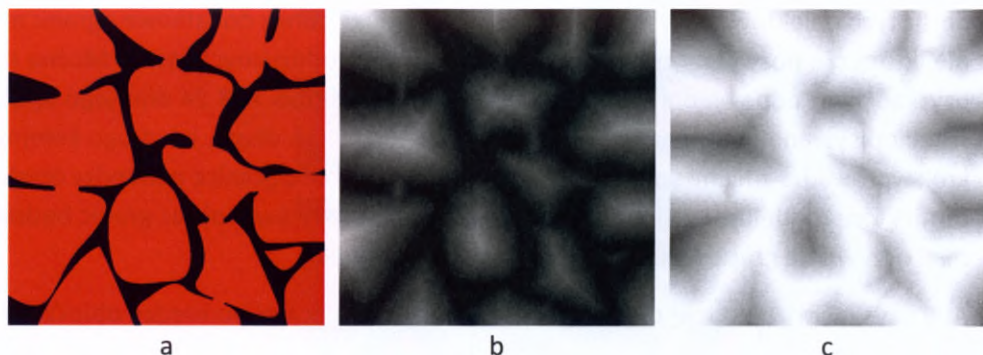
Aby uzyskać wizualizację obrazu początkowego z granicami pomiędzy obiektami, należy dokonać odpowiedniego połączenia obrazów. Najprostsze jest przeprowadzenie logicznego odejmowania wyniku detekcji działów wodnych od początkowego obrazu binarnego.

W celu przygotowania obrazu binarnego (il. 8.29a) do detekcji działów wodnych wyznacza się funkcję odległości (ang. *distance function*), której wartości podają odległość od brzegu zestawu obiektów (il. 8.29b). Wartości tej funkcji odpowiadają liczbie kroków erozji, potrzebnej do osiągnięcia miejsca odpowiednio oddalonego od brzegu obiektów. Obraz umownego dna jezior uzyskuje się po prostu wykonując negatyw funkcji odległości (il. 8.29c). W celu poprawy kontrastu wartości pikseli w obrazach na il. 8.29b i 8.29c rozciągnięto do zakresu 0-255.

Można teraz wykonać detekcję działów wodnych. Wynik jest poprawny z wyjątkiem miejsc zaznaczonych strzałkami (il. 8.30a). Te dodatkowe linie podziału wynikają z cyfrowego charakteru wszystkich przekształceń, gdzie lokalne skokowe zmiany wartości poszczególnych pikseli mogą wpływać na końcowy wynik detekcji.

Możliwe jest jednak pewne zmodyfikowanie omawianej procedury przez wprowadzenie znaczników (il. 8.30b). Znaczniki opisują początkowy moment całej procedury. Ich największa siła polega na tym, że w końcowym efekcie detekcja działów wodnych wygeneruje tyle obiektów, ile było znaczników. W ten sposób zabezpiecza się przed niepożądanymi, nadmiarowymi podziałami.

Osobnym problemem jest przygotowanie znaczników. Jeżeli wykorzysta się do tego celu erozję warunkową, to otrzymamy dużą liczbę nadmiarowych podziałów. Dlatego w tym przypadku zastosowano kilka kroków erozji. Ze względu na konieczność rozdzielania sklejonych obiektów (powstające znaczniki nie mogą być sklejo- ne) erozja musiała być na tyle obszerna, że usunęła wszystkie piksele niewielkiego obiektu, widocznego blisko prawego dolnego rogu obrazu z il. 8.29a. Odpowiedni znacznik (dla zapewnienia lepszej czytelności pokazany na il. 8.30b kolorem białym) dodano na podstawie wyników erozji warunkowej. Końcowy wynik detekcji

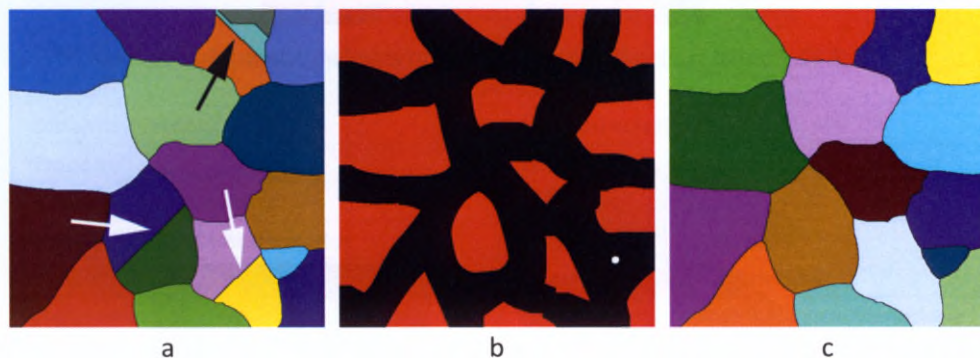


Il. 8.29. Przygotowanie obrazu binarnego do detekcji działów wodnych: a) wyjściowy obraz binarny, b) funkcja odległości, c) negatyw funkcji odległości

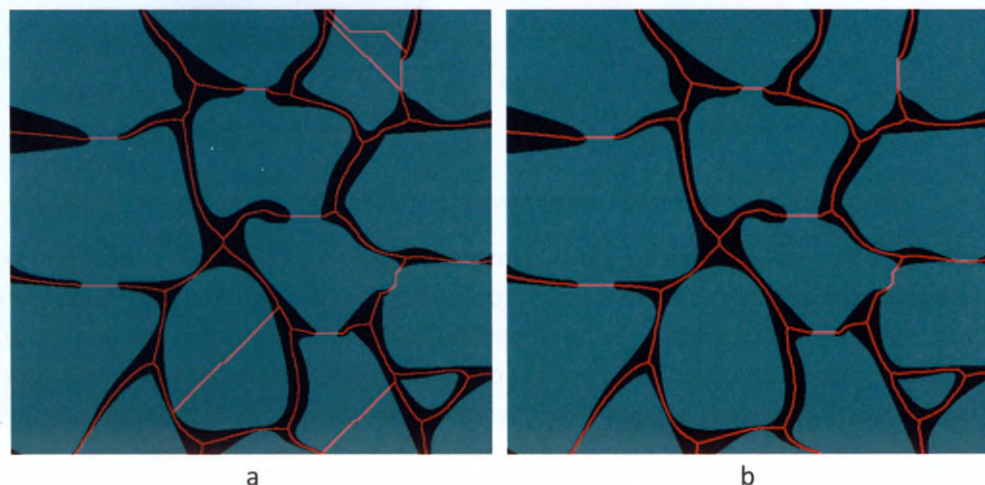
działów wodnych z wykorzystaniem znaczników (ang. *seeded watershed segmentation*) przedstawiono na il. 8.30c.

Porównanie wyników obu wariantów detekcji działów wodnych pokazuje il. 8.31. Wykorzystanie znaczników (markerów) zlikwidowało nadmiarowe podziały i pozwoliło na prawidłowe rozdzielenie obiektów.

Il. 8.32 przedstawia z kolei wynik podziału sklejonych obiektów, uzyskany za pomocą operacji SplitConvex, zaimplementowanej w systemie Aphelion. Operacja ta wymaga podania parametru opisującego umowną siłę (*Filter Strength*) podziału na obiekty wypukłe. Pokazany na il. 8.32 wynik detekcji z parametrem siły wynoszącym 6, jest



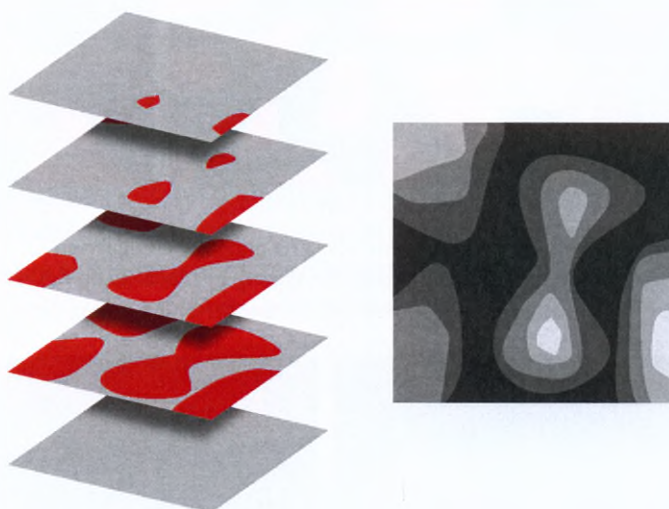
Il. 8.30. Detekcja działów wodnych: a) detekcja na podstawie negatywu funkcji odległości z nadmiarowymi podziałami zaznaczonymi strzałkami, b) znaczniki do detekcji, c) detekcja działów wodnych ze znacznikami



Il. 8.31. Wynik detekcji działów wodnych nałożony na obraz początkowy: a) detekcja działów wodnych bez znaczników, b) detekcja działów wodnych ze znacznikami



Il. 8.32. Podział sklejonych obiektów z wykorzystaniem operacji SplitConvex w systemie Aphelion



Il. 8.33. Interpretacja obrazu szarego jako sumy obrazów binarnych

identyczny z wynikiem detekcji z udziałem znaczników (il. 8.30c i 8.31b). Jeżeli przyjąć wartość tego parametru na poziomie 1, to uzyskamy wynik identyczny z detekcją bez znaczników (il. 8.30a i 8.31a). Wyższe wartości parametru siły filtru doprowadzą do redukcji liczby podziałów. Sposób, w jaki omawiany parametr wpływa na wynik filtracji, nie został opisany w dokumentacji systemu, ale omówione wcześniej procedury wskazują, że ma on związek z generowaniem znaczników do detekcji działów wodnych.

Przedstawiony przegląd operacji morfologii matematycznej dla obrazów binarnych pokazuje, że jest to zestaw bardzo elastycznych, różnorodnych filtrów, które

pozwalają na daleko idące modyfikacje obrazów. W przypadku klasycznych filtrów modyfikacje obrazów binarnych ograniczają się w praktyce do filtracji uśredniającej oraz filtrów rankingowych, w tym maksymalnego, minimalnego oraz medianowego. Ponieważ w systemach analizy obrazu nie tłumaczy się angielskojęzycznych terminów związanych z operacjami morfologii matematycznej, a w tym rozdziale pojawiło się dużo nowych terminów, zabrano je w postaci tabelarycznej. Tabela 8.1 zawiera słownik wybranych terminów.

Tabela 8.1. Słownik wybranych terminów związanych z morfologią matematyczną

Termin angielski	Termin polski
Border kill	Usunięcie obiektów przeciętych przez brzeg obrazu (dosł. zabicie granicy)
Closing	Zamknięcie
Dilation	Dylatacja
Distance (function)	(funkcja) odległości
Erosion	Erozja
Geodesic (operations)	(operacje) geodezyjne
Hit or miss (HMT) transformation	Przekształcenie „trafi, nie trafi”
Hole fill	Wypełnienie (zalanie) otworów
Linear erosion/dilation/opening/closing	Liniowa erozja/dylatacja/otwarcie/zamknięcie
Mask	Maska
Multiple points	Punkty wielokrotne (mające więcej niż dwóch sąsiadów)
Opening	Otwarcie
Pruning	Obcinanie gałęzi
Reconstruction	Rekonstrukcja
Rotating structuring element	Wirujący element strukturalny
Seeded watershed segmentation	Detekcja działów wodnych ze znacznikami
Single points	Punkty pojedyncze (odizolowane)
Skeletonization	Szkieletyzacja
SKIZ (Skeleton by Influence Zone)	SKIZ (szkieletyzacja przez strefę wpływów)
Split convex	Podziel na wypukłe
Structuring (structurizing) element	Element strukturalny
Thickening	Pogrubianie
Thinning	Scienianie
Triple points	Punkty węzłowe (potrójne)
Ultimate eroded set	Wynik erozji warunkowej
Ultimate erosion	Erozja warunkowa
Watershed segmentation	Detekcja działów wodnych

W dalszej części rozdziału poświęconego operacjom morfologicznym zajmiemy się działaniem poznanych już przekształceń na obrazach szarych. Omawianie przekształceń morfologii matematycznej dla obrazów szarych rozpoczniemy od przedstawienia prostego związku łączącego obrazy szare i binarne (il. 8.33).

Obraz szary (prawa strona il. 8.33) możemy interpretować jako sumę obrazów binarnych, pokazanych z lewej strony il. 8.33. Na samym dole będziemy mieli obraz binarny, odpowiedzialny za wartość 0. Będzie on zawierał wszystkie piksele obrazu oraz informację o jego rozmiarach. Następna będzie warstwa odpowiedzialna za wszystkie piksele o wartości co najmniej 1. Kolejna będzie odpowiadała za wszystkie piksele o wartości co najmniej 2, następna za wszystkie piksele o wartości co najmniej 3 itd. Poziom szarości pojedynczego piksela będzie równy liczbie warstw (poza zerową), w których dany piksel ma wartość 1. Oczywiście jest, że każda kolejna warstwa nie może mieć większej liczby pikseli o wartości 1 niż warstwy leżące pod nią. Jeżeli jakikolwiek piksel leżący na warstwie o numerze x ma wartość 1, to piksele o tych samych współrzędnych, leżące na wszystkich warstwach od 1 do x , również mają wartość 1.

Na podstawie tej interpretacji obrazu szarego można łatwo zdefiniować operacje morfologiczne na obrazach szarych. Po prostu wykonujemy daną operację na wszystkich warstwach binarnych (już wiemy, jak to zrobić), a suma tych warstw po operacji daje wynik przekształcenia.

I na tym można byłoby zakończyć rozważania w sposób szybki, ale nie dający zbyt wielu informacji przydatnych w praktyce. Niemniej przedstawiona interpretacja obrazu szarego jako sumy obrazów binarnych pokazuje, że „nie taki diabeł straszny”.

Oczywiście, gdybyśmy chcieli dzielić obraz szary na obrazy binarne, wykonywać na nich operacje i następnie składać z powrotem w obraz szary, to tracilibyśmy bardzo dużo czasu na obliczenia. Można zrobić to inaczej dzięki nowej interpretacji elementu strukturalnego. Definiujemy go tak samo, jak było to w przypadku obrazów binarnych, za pomocą trzech symboli: 0, 1 oraz X. Ich znaczenie jest teraz następujące:

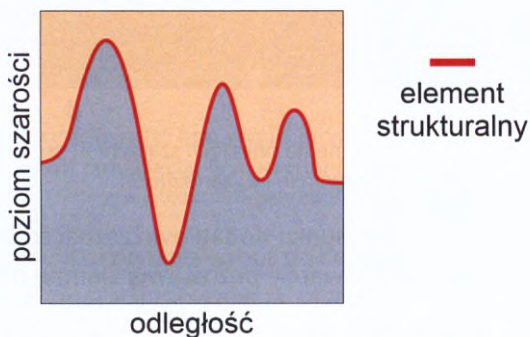
- 0 oznacza, że piksel znajdujący się w tym miejscu powinien mieć wartość mniejszą od punktu centralnego,
- 1 oznacza, że piksel znajdujący się w tym miejscu powinien mieć wartość równą lub większą od punktu centralnego,
- X oznacza, podobnie jak w przypadku obrazów binarnych, że wartość tego piksela nie jest brana pod uwagę.

Zmienia się również traktowanie piksela odpowiadającego punktowi centralnemu. Jeżeli w przekształceniach dla obrazów binarnych miał przyjąć wartość 0, to

w przypadku obrazów szarych przyjmuje najniższą wartość spośród wszystkich pikseli branych pod uwagę przez element strukturalny. Analogicznie, jeżeli w przekształceniach dla obrazów binarnych miał przyjąć wartość 1, to w przypadku obrazów szarych przyjmuje najwyższą wartość spośród wszystkich pikseli branych pod uwagę przez element strukturalny.

Tylko tyle i aż tyle zmian. Oczywiście obrazy szare pozwalają zapisać o wiele więcej informacji niż binarne. Stąd też wyników operacji morfologicznych będzie więcej, a w dodatku możliwe są nowe operacje, których nie da się przeprowadzić na obrazach binarnych. Warto przy okazji przypomnieć, że pewne operacje typowe dla obrazów binarnych, np. szkieletyzacja, dają w przypadku obrazów szarych dość zaskakujące wyniki.

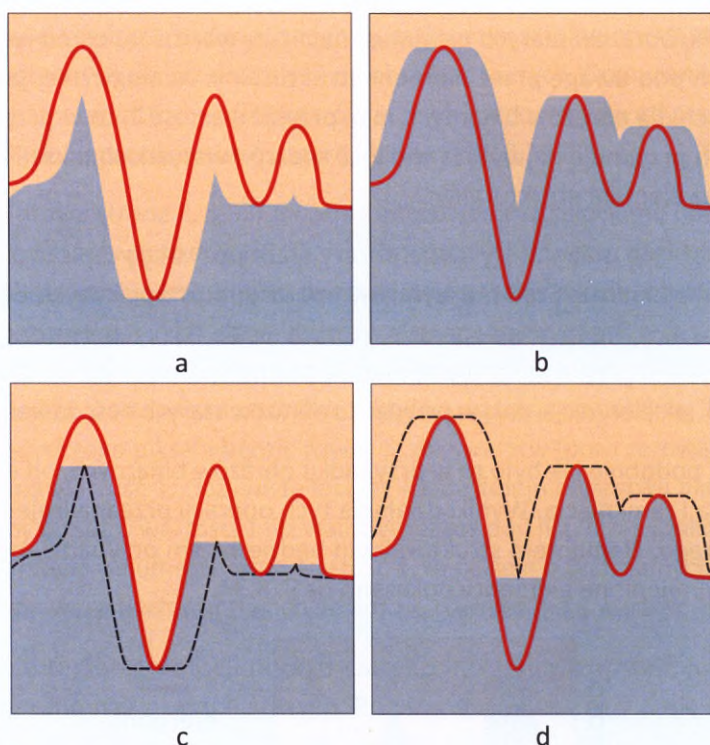
Zacznijmy, podobnie jak było to w przypadku obrazów binarnych, od erozji, dylatacji, otwarcia i zamknięcia. Wyniki działania tych operacji przeanalizujemy na profilu obrazu szarego. Elementem strukturalnym będzie w tym przypadku poziomy odcinek. Oba wymienione elementy pokazano na il. 8.34.



Il. 8.34. Profil obrazu szarego oraz element strukturalny

Na il. 8.35 przedstawiono wyniki czterech podstawowych operacji:

- Erozja (il. 8.35a). Początkowy profil pokazano w postaci czerwonej linii, natomiast profil po erozji reprezentuje górna część szarego obszaru. Erozja powoduje przyciemnienie obrazu, ograniczenie wielkości jego jasnych partii oraz powiększenie obszarów najciemniejszych. Potwierdza to porównanie il. 8.36a oraz 8.36b. Podobnie jak w przypadku obrazów binarnych, erozja jest addytywna.
- Dylatacja (il. 8.35b). Znaczenie poszczególnych elementów jest takie samo jak na il. 8.35a. Wyraźnie widać, że dylatacja jest przeciwieństwem erozji. Powoduje rozjaśnienie obrazu i ograniczenie wielkości ciemnych elementów. Potwierdza to porównanie il. 8.36a oraz 8.36c.



Il. 8.35. Operacje morfologiczne na obrazach szarych: a) erozja, b) dylatacja, c) otwarcie i d) zamknięcie

- Otwarcie (il. 8.35c). Na tym rysunku dodatkowo zaznaczono czarną przerywaną linią profil po erozji obrazu. Otwarcie pozostawia ciemne obszary bez większych zmian, natomiast odcina najjaśniejsze elementy obrazu.
- Zamknięcie (il. 8.35d). Czarna przerywana linia oznacza profil po dylatacji. Zamknięcie pozostawia bez zmiany obszary najjaśniejsze i zalewa wyższymi wartościami poziomów szarości miejsca najciemniejsze (lokalne minima). Dzięki temu obraz staje się jaśniejszy, choć nie jest to tak wyraźne jak w przypadku dylatacji (il. 8.35b). Jeżeli wykorzysta się dostatecznie duży element strukturalny, to dylatacja całkowicie usunie ciemne elementy i zamknięcie będzie stanowiło obwiednię najjaśniejszych pikseli obrazu. Dlatego zamknięcie jest powszechnie używane w procedurach korekcji cienia.

Operacje erozji, dylatacji, zamknięcia oraz otwarcia wykorzystuje się w celu modyfikacji obrazu stanowiącej element całego ciągu operacji, których końcowym wynikiem ma być najczęściej detekcja obiektów. Przykładowe wyniki działania erozji, dylatacji i zamknięcia pokazuje il. 8.36. Niektóre operacje mogą jednak przynieść wyniki, które oprócz wartości merytorycznej dla analizy obrazu, są interesujące



a



b



c



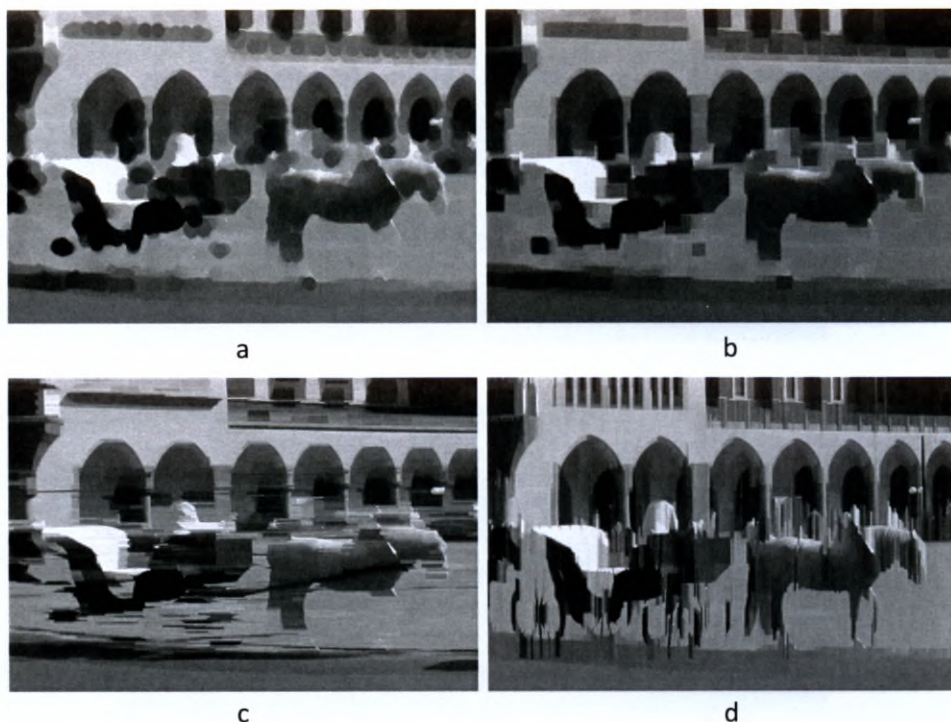
d

Il. 8.36. Obraz wyjściowy (a) oraz wyniki prostych operacji morfologicznych: b) erozji, c) dylatacji, d) zamknięcia. Wykorzystano kwadratowy element strukturalny o wielkości 1% szerokości obrazu

wizualnie i mogą być wykorzystane wręcz do celów artystycznych. Pewne przykłady, zawierające wyniki operacji zamknięcia, przedstawia il. 8.37.

Il. 8.37a i 8.37b przedstawiają obraz wyjściowy (il. 8.36a) po zamknięciu dużym elementem, odpowiednio kołowym i kwadratowym. Mamy bardzo daleko idące uogólnienie obrazu, ale wciąż możliwe jest rozpoznanie konia, dorożki i budynku Sukiennic w Krakowie jako tła. W latach osiemdziesiątych ubiegłego wieku podobnej generalizacji poddano kolorowe reprodukcje obrazów Paula Gauguina. Wybór artysty nie dziwi, skoro był Francuzem, a twórcami morfologii matematycznej są też Francuzi. Okazało się, że na uogólnionych obrazach ujawniły się pewne powtarzalne cechy, będące wynikiem techniki malarskiej Gauguina. Najciekawsze w tych analizach było to, że oryginały poddane takiej analizie wyraźnie różniły się od wysokiej jakości kopii, które były trudne do identyfikacji nawet dla fachowców.

Liniowe zamknięcia dają efekty zbliżone do znanych z programów do edycji grafiki efektów specjalnych, dających wrażenie ruchu (ang. *motion blur*) lub ściekającej po ścianie świeżej farby (ang. *wet paint*).



Il. 8.37. Różne rodzaje zamknięcia obrazu z il. 8.36a: a) kołowy element strukturalny, b) kwadratowy element strukturalny, c) poziomy element liniowy, d) pionowy element liniowy. Wielkość elementów wynosiła ok. 2% szerokości obrazu w przypadku elementu kołowego i kwadratowego oraz ok. 3% szerokości obrazu w przypadku liniowych elementów strukturalnych

Przekształcenia morfologiczne obrazów szarych pozwoliły na wprowadzenie dwóch operacji, znanych pod nazwą *white top hat* oraz *black top hat*. Nie ma sensu dosłownie tłumaczyć tych nazw. *White top hat* to po prostu detekcja lokalnych maksimów, a *black top hat* to detekcja lokalnych minimów na obrazie.

Pomocne w zrozumieniu tych przekształceń będą odpowiednio il. 8.35c i 8.35d. *White top hat* to różnica pomiędzy obrazem początkowym i obrazem po otwarciu. Otrzymamy wartości odpowiadające jasnobrązowemu obszarowi pomiędzy profilem obrazu i obszarem po otwarciu na il. 8.35c. Będą to lokalne maksima obrazu. Z kolei *black top hat* to negatyw różnicy obrazu po zamknięciu i obrazu początkowego, co daje detekcję lokalnych minimów na obrazie.

Nazwa *top hat*, czyli czubek kapelusza, jest najprawdopodobniej odbiciem poczucia humoru twórcy lub twórców tego przekształcenia. Jeżeli weźmiemy pod uwagę fragment czerwonej linii profilu obrazu, to może się on kojarzyć z kapeluszem (il. 8.38).



Il. 8.38. Żartobliwy rysunek wyjaśniający nazwę operacji *top hat*. Opis w tekście

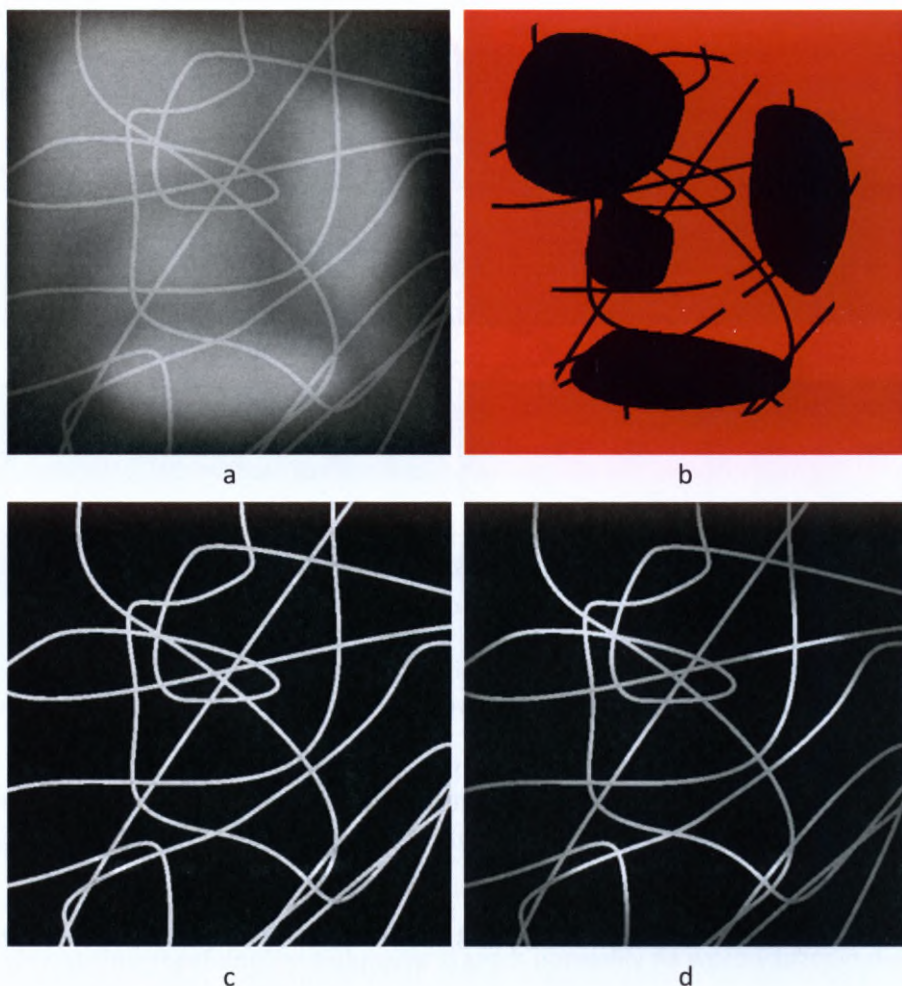
Jeżeli odetniemy od niego obszar po otwarciu (szara powierzchnia na il. 8.38), to pozostanie nam faktycznie czubek kapelusza, widoczny na rysunku jako biały obszar pod czerwoną linią symbolizującą profil obrazu.

Lokalne minima są czasami, oprócz erozji warunkowej, narzędziem do przygotowywania znaczników do detekcji działów wodnych. Uzyskuje się te znaczniki dzięki binaryzacji obrazu lokalnych minimów.

Przekształcenie *top hat* znajduje dość często zastosowanie podczas detekcji krawędzi, jeżeli stanowią one lokalne minima lub maksima. Dzieje się tak np. w przypadku kostki brukowej lub płytek w łazience.

Jeżeli mamy bardzo zróżnicowane tło (il. 8.39a), to detekcja obiektów za pomocą binaryzacji może okazać się niemożliwa (il. 8.39b). W takim przypadku warto spróbować wykorzystać operację *white top hat*. Wynik detekcji lokalnych maksimów (*white top hat*) pokazano na il. 8.39c i porównano z wynikiem korekcji cienia. Widać wyraźnie, że obraz po korekcji cienia (il. 8.39d) ma silniej zróżnicowane wartości pikseli w obrębie jasnych linii. W dodatku korekcja cienia zajmuje więcej czasu, gdyż zamiast odejmowania trzeba wykonać dzielenie i mnożenie. Obraz po operacji *top hat* można teraz z łatwością poddać binaryzacji. Warto zwrócić uwagę, że wykonując binaryzację po operacji *top hat*, wykonujemy w rzeczywistości binaryzację z lokalnie zmiennym progiem czułości. Innymi słowy, przeprowadzamy binaryzację adaptacyjną lub dynamiczną.

Wyniki operacji *top hat* mogą być zaskakujące w przypadku obrazów zawierających wiele drobnych elementów o szybko zmieniającym się poziomie szarości, jak np. obraz konia z dorożką z il. 8.36a. Jeżeli wykonamy operację *white top hat*, to wzmocnimy również stosunkowo ciemne obszary, jeżeli tylko są lokalnie jaśniejsze



Il. 8.39. Wykorzystanie operacji *white top hat*: a) testowy obraz do demonstracji zastosowania operacji *white top hat*, b) błędny wynik binaryzacji, c) wynik operacji *top hat*, d) wynik korekcji cienia. Początkowe poziomy szarości zostały w celu zwiększenia czytelności obrazów rozciągnięte do zakresu 0-255

od otoczenia. Z tego powodu po wyrównaniu histogramu dostajemy zaskakująco interesującą grafikę, przypominającą rysunek ołówkiem (il. 8.40).

Lokalnie najciemniejsze punkty (il. 8.41a) można wykryć za pomocą operacji *black top hat* (il. 8.41b). Jeżeli wykonamy negatyw tego obrazu (il. 8.41c), to z łatwością możemy już przeprowadzić jego binaryzację. W tym konkretnym przypadku równomiernie zmieniające się tło bardzo ułatwia korekcję cienia, która daje znakomity wynik (il. 8.41d). Obraz widoczny na il. 8.41c jest jednocześnie obrazem z korekcją cienia, otrzymaną przez dodanie do obrazu wyjściowego negatywu tła. Rozwiązanie



Il. 8.40. Obraz z il. 8.36a poddany operacji *white top hat* z późniejszym wyrównaniem histogramu

to było dyskutowane w rozdziale dotyczącym korekcji cienia i ocenione jako takie, którego należy unikać.

Omówione powyżej przykłady, zwłaszcza przedstawione na il. 8.39, 8.40 i 8.41, nie są wcale reprezentatywne dla wszystkich problemów związanych z operacją *top hat*. W szczególności porównanie z prawidłowo wykonaną korekcją cienia pokazuje, że każde z tych przekształceń może dać lepsze wyniki niż drugie. Wszystko zależy od konkretnych obrazów oraz zagadnień, które mamy rozwiązać. Jednocześnie przykłady te pokazują, że wszystkie przekształcenia analizy obrazu stanowią pewną całość i często podobne wyniki można otrzymać, stosując wyraźnie odmienne algorytmy.

Mamy również operacje morfologiczne, które mogą być używane w przypadku obrazów szarych, ale w odróżnieniu od obrazów binarnych dają często nieprzewidywalne wyniki. Przykładem takiej operacji jest szkieletyzacja. Jeżeli zastosujemy ją do szarego obrazu, przedstawiającego trzy w przybliżeniu koliste obiekty o rozmytych kształtach, to otrzymamy bardzo zaskakujący wynik (il. 8.42). W pewnych przypadkach szkieletyzacja może spowodować np. wzrost kontrastu i ułatwić detekcję, ale są to dość rzadkie przypadki.

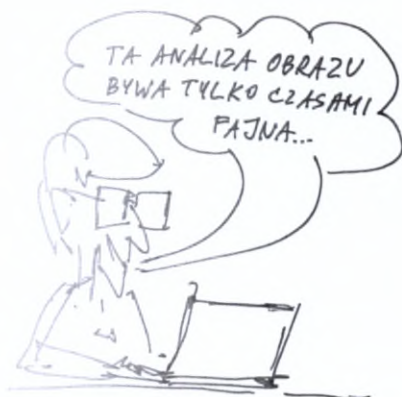
W odróżnieniu od szkieletyzacji detekcja działów wodnych bywa z powodzeniem wykorzystywana w przetwarzaniu obrazów szarych, bo nawet dla obrazów binarnych detekcję działów wodnych wykonuje się na obrazach pomocniczych (funkcja odległości), które są obrazami szarymi. Detekcja działów wodnych jest cennym narzędziem do detekcji obiektów o rozmytych brzegach (il. 8.42a). Rozmyte w mniejszym lub



a



b



c



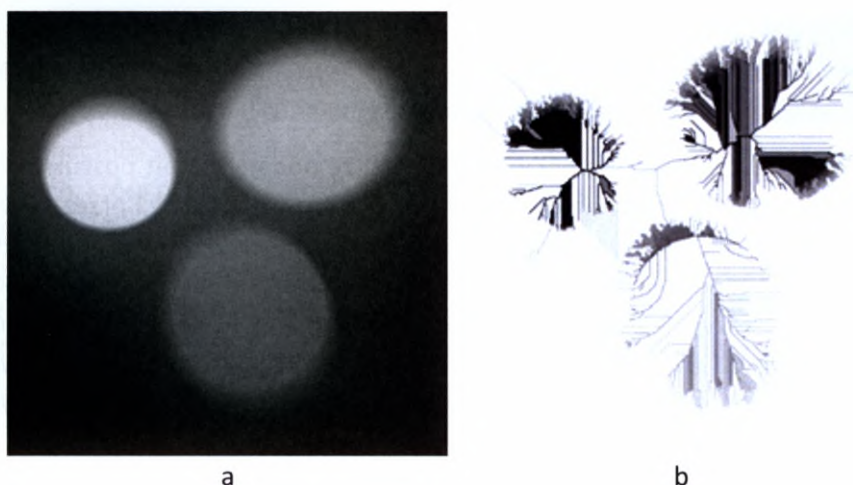
d

Il. 8.41. Obraz z silnym efektem cienia (a), wynik operacji *black top hat* (b), jego negatyw (c) oraz obraz po korekcji cienia (d)

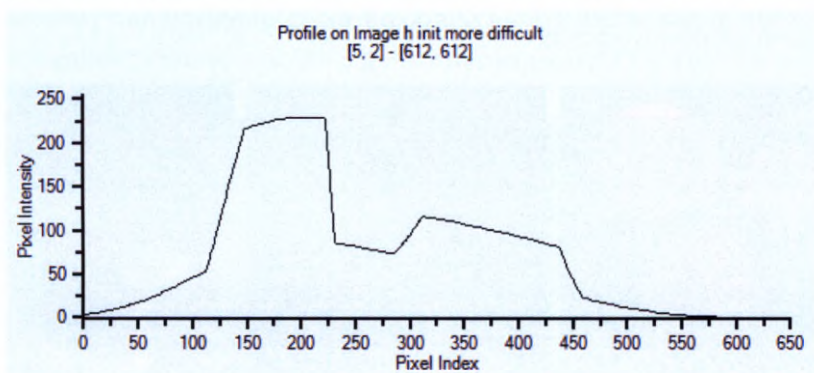
większym stopniu krawędzie obiektów są typowym zjawiskiem, ale radzimy sobie z tym, dobierając po prostu odpowiedni próg binaryzacji. W tym przypadku mamy jednak znacznie bardziej złożony problem:

Obiekty wykazują zróżnicowane poziomy szarości, zatem zastosowanie wspólnego poziomu szarości jako kryterium detekcji spowoduje, że zawsze któryś obiektów będzie binaryzowany z niewłaściwym progiem detekcji.

Dodatkowo obraz ma nie tylko nierówne tło. Poziomy szarości obiektów są również zmienne, co dodatkowo bardzo utrudnia detekcję (il. 8.43).



Il. 8.42. Trzy obiekty o rozmytych brzegach (a) oraz zaskakujący wynik ich szkieletyzacji (b). W celu poprawy czytelności obraz (b) wyświetlany jest jako negatyw wyniku szkieletyzacji



Il. 8.43. Profil poziomów szarości obrazu z il. 8.42a, wyznaczony wzdłuż linii biegnącej po przekątnej od lewego górnego do prawego dolnego rogu

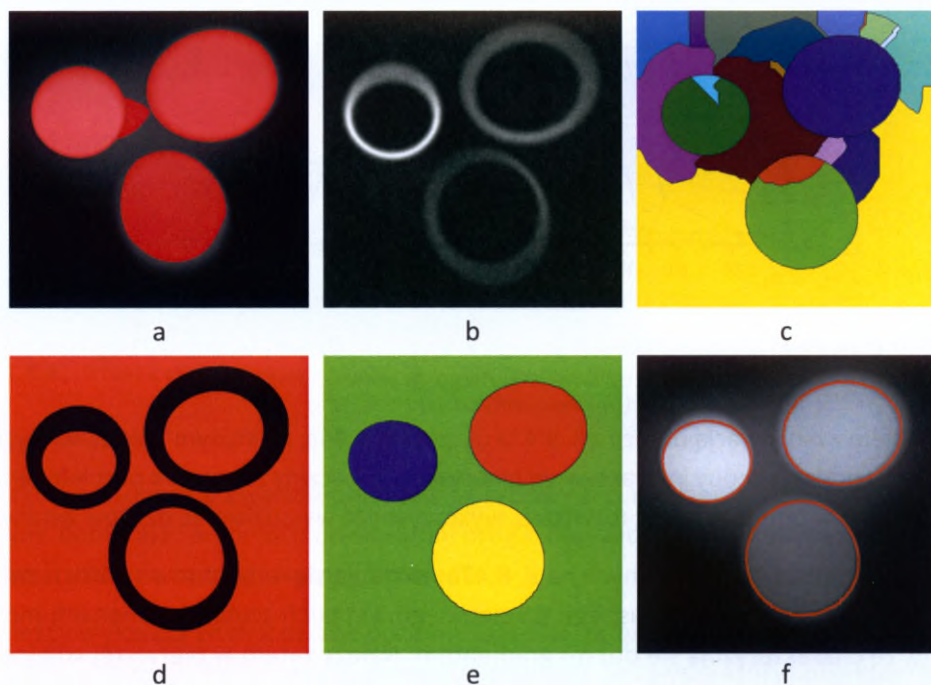
Problem z binaryzacją obrazu z il. 8.42a ukazuje il. 8.44a, na którym pokazano wygląd obrazu podczas próby interaktywnej binaryzacji. Zobaczmy zatem, jak problem detekcji można w tym przypadku rozwiązać, wykorzystując wyszukiwanie działów wodnych.

Granice obiektów, widocznych na il. 8.42a, znajdują się w obszarze szybkich zmian poziomów szarości – od jasnego do ciemnego. O takich szybkich zmianach mówimy, że charakteryzują się dużym gradientem. Należy zatem w pierwszej kolejności stworzyć obraz gradientowy. Można zrobić to bardzo łatwo, odejmując od obrazu po dylatacji obraz po erozji. Niestety, pracujemy w przestrzeni dyskretniej, a nie ciągłej, co powoduje, że obraz gradientowy będzie miał w niektórych miejscach gwałtowne zmiany poziomów szarości, co może spowodować poszarpany brzeg obiektów

wykrywanych za pomocą działów wodnych. Dlatego warto ten surowy obraz gradientowy poddać rozmyciu, np. filtrem Gaussa. Tak przygotowany obraz gradientowy może stanowić podstawę do detekcji działów wodnych (il. 8.44b).

Automatyczna detekcja działów wodnych dobrze oddaje kształt poszukiwanych obiektów, ale obraz wykazuje dodatkowe, nadmiarowe podziały i przez to wynik jest w praktyce bezużyteczny (il. 8.44c). Musimy zatem przygotować odpowiednie markery, które zapewnią ochronę przed nadmiarowymi podziałami. Wobec trudności z binaryzacją wyjściowego obrazu, można je w tym przypadku uzyskać na drodze binaryzacji obrazu gradientowego (il. 8.44d). Warto w tym miejscu zwrócić uwagę, że sam obraz gradientowy nie nadaje się do detekcji obiektów, choć dobry wynik dałby SKIZ obrazu z il. 8.44d.

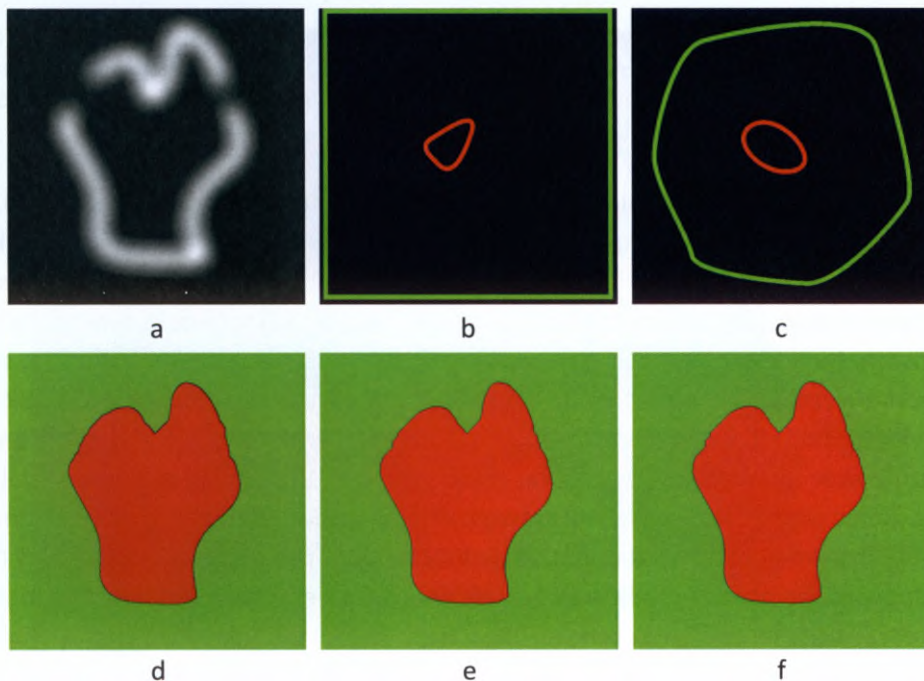
Możemy teraz bez przeszkód dokonać detekcji działów wodnych z ograniczeniem przez znaczniki. Wynik pokazuje il. 8.44e. Nałożenie granic wykrytych obiektów na obraz początkowy (il. 8.44f) dowodzi skuteczności i poprawności detekcji z wykorzystaniem działów wodnych.



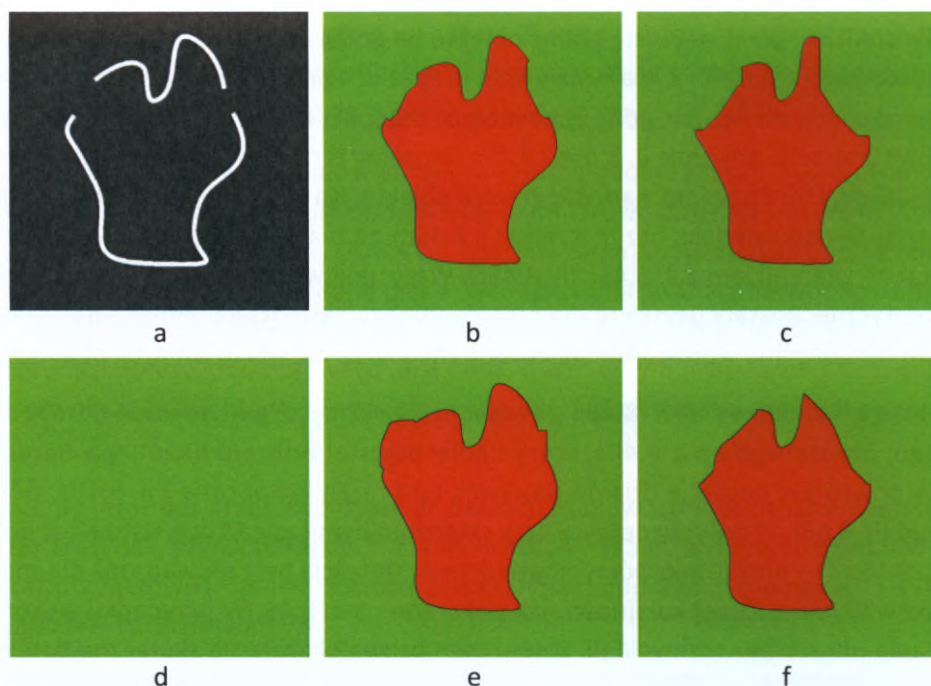
Il. 8.44. Wykorzystanie działów wodnych do detekcji obiektów: a) ilustracja problemów z binaryzacją, b) obraz gradientowy, c) wynik detekcji działów wodnych, d) markery (znaczniki) do detekcji, uzyskane na podstawie binaryzacji obrazu gradientowego, e) wynik detekcji działów wodnych ze znacznikami oraz f) wynik detekcji nałożony na obraz początkowy

Prześledzimy teraz detekcję granic obiektu na podstawie niepełnego obrazu gradientowego granic (il. 8.45a). Zazwyczaj przy takich próbach detekcji używa się systemów interaktywnych, gdzie odpowiednie znaczniki wprowadza operator. Przykładowe znaczniki pokazuje il. 8.45b i 8.45c. Możemy teraz wykonać detekcję działów wodnych, która powinna nam dać granice obiektu. Na il. 8.45 pokazano trzy wyniki, uzyskane odpowiednio: bez znacznika (il. 8.45d), ze znacznikiem z il. 8.45b (il. 8.45e) oraz ze znacznikiem z il. 8.45c (il. 8.45f). Wszystkie wyznaczone zarysy są niemal identyczne. Zostały uzyskane dla czterospójnego sąsiedztwa. Zmiana na sąsiedztwo ośmiospójne wprowadza tak drobne zmiany, że nie warto ich tutaj pokazywać.

Obraz gradientowy na il. 8.45a powstał przez rozmycie gaussowskie obrazu szarego, pokazanego na il. 8.46a, który z kolei powstał z obrazu binarnego. Spróbujmy powtórzyć detekcję działów wodnych tym razem na obrazie z il. 8.46a. Górny rząd (il. 8.46b-c) dotyczy detekcji przy czterospójnym sąsiedztwie, natomiast dolny (il. 8.46d-f) – przy ośmiospójnym sąsiedztwie. Detekcja była prowadzona dla znaczników w takiej samej kolejności, jak pokazano na il. 8.45. W przypadku wykorzystania obrazu bez rozmycia (il. 8.46a) niemożliwa okazała się skuteczna detekcja



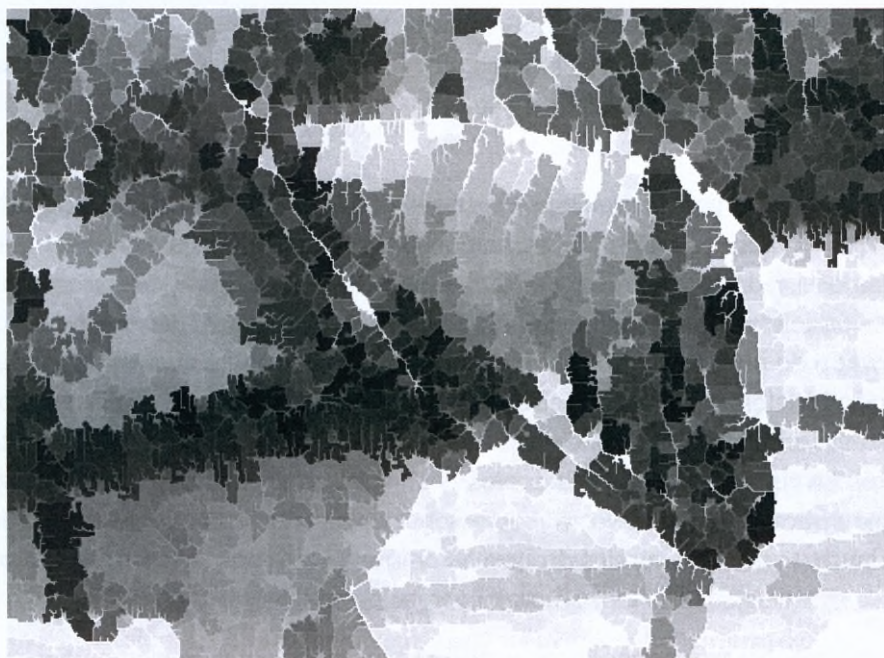
Il. 8.45. Detekcja obiektu na podstawie niepełnego obrazu gradientowego (a) jego brzegu: b-c) znaczniki, d) detekcja działów wodnych bez znacznika, e) detekcja ze znacznikiem z obrazu (b), f) detekcja ze znacznikiem z obrazu (c)



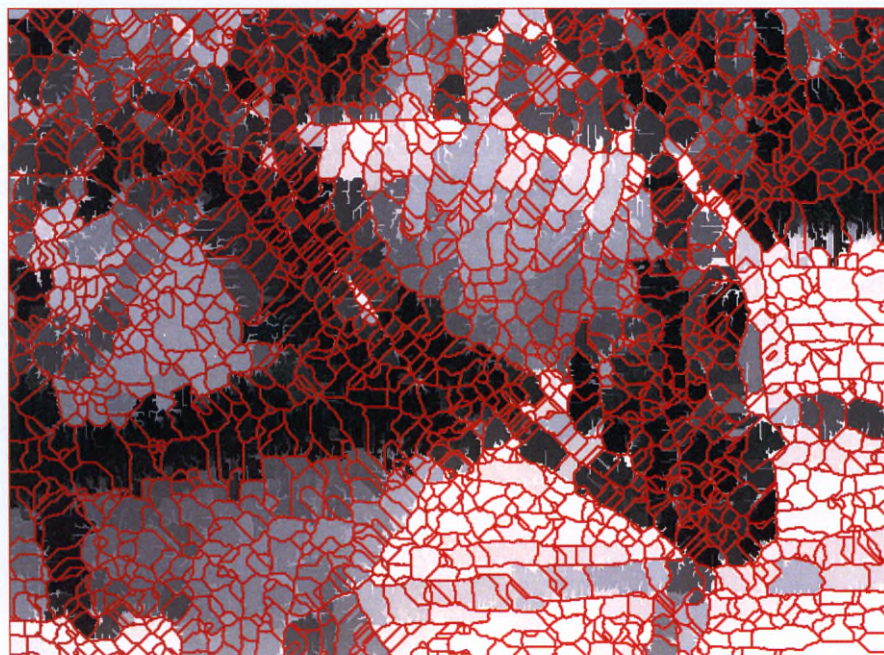
Il. 8.46. Wykorzystanie działów wodnych do detekcji granic obiektu, widocznych na obrazie (a): b) znacznik z il. 8.45b, czterospójne sąsiedztwo, c) znacznik z il. 8.45c, czterospójne sąsiedztwo, d) bez znacznika, ośmiospójne sąsiedztwo, e) znacznik z il. 8.45b, ośmiospójne sąsiedztwo, f) znacznik z il. 8.45c, ośmiospójne sąsiedztwo

bez znaczników, niezależnie od rodzaju sąsiedztwa (il. 8.46d). Wynika to stąd, że wszystkie ciemne punkty na il. 8.46a mają wartość 0. Na początku detekcji działów wodnych powstał jeden obiekt z otworami, które zostały zalane. Po rozmyciu gaussowskim powstały punkty o wartościach powyżej zera, które dały dwa rozłączne obszary punktów o wartości 0 w obrazie z il. 8.45a. Dzięki temu możliwe było skuteczne wyznaczenie obiektów (il. 8.45d-f). Natomiast w pozostałych przypadkach wykorzystania il. 8.46a jako obrazu gradientowego wynik detekcji zmieniał się wyraźnie wraz ze zmianą znacznika oraz zasad spójności. Płyne stąd wniosek, że detekcja granic obiektów z rozmytymi krawędziami za pomocą detekcji działów wodnych może być skutecznym narzędziem, ale jednocześnie w pewnych warunkach, przy nieodpowiednim obrazie gradientowym, nie sprawdza się. Trudno jest, niestety, podać precyzyjne zalecenia, w jakich warunkach sprawdzi się konkretny wariant analizy. Trzeba to każdorazowo wypróbować na starannie dobranych obrazach testowych.

Dotychczasowa analiza pokazuje, że operacje morfologiczne na obrazach szarych mogą dać wiele interesujących wyników, ale są mniej przewidywalne (przynajmniej



a



b

Il. 8.47. a) szkieletyzacja głowy konia (fragmentu il. 8.36a), b) wynik detekcji działów wodnych z tego obrazu, nałożony na wynik szkieletyzacji z il. (a)

niektóre) niż w przypadku obrazów binarnych. Systemy analizy obrazu przewidują jeszcze inne złożone operacje, których nie omawialiśmy, ale ogólnie rzecz biorąc, są one złożeniem kilku operacji omawianych w ramach tego rozdziału, zatem ich zrozumienie nie powinno stanowić problemu.

Interesujące jest prześledzenie pewnych związków pomiędzy szkieletyzacją i detekcją działów wodnych. Wydaje się, że oba przekształcenia mają zupełnie inne cele i własności, jednak w przypadku obrazów szarych zdarzają się zaskakujące podobieństwa. Szkieletyzacja fragmentu obrazu z il. 8.36a, przedstawiającego głowę konia, prowadzi do powstania mozaiki niewielkich obszarów o tym samym poziomie szarości, oddzielonych nieco jaśniejszą siatką o grubości pojedynczego piksela. Poziom szarości poszczególnych elementów mozaiki zmienia się podobnie do zmian poziomu szarości wyjściowego obrazu, ale całość jest nieco jaśniejsza (il. 8.47a). Jeżeli na tym samym obrazie wyjściowym wykonamy bez udziału znaczników detekcję działów wodnych, to otrzymamy sieć drobnych komórek, która dość wiernie odwzorowuje siatkę rozdzielającą elementy mozaiki obrazu po szkieletyzacji (il. 8.47b).

Ta pozornie zaskakująca zgodność wynika z używania podobnych elementów strukturalnych i zbliżonych operacji na tych elementach. Jest to sytuacja podobna jak w przypadku porównania wyników szkieletyzacji oraz SKIZ. SKIZ można uzyskać za pomocą sekwencji przekształceń: negatyw – szkieletyzacja – obcięcie gałęzi – negatyw. Są to, jak gdyby, dwa warianty szkieletyzacji. Analogiczne podobieństwa mamy w przypadku erozji i dylatacji lub otwarcia i zamknięcia. Z drugiej strony, wyróżnienie np. erozji i dylatacji jako dwóch odrębnych przekształceń jest bardzo wygodne z praktycznego punktu widzenia, gdyż skraca i upraszcza algorytmy przetwarzania obrazu.

Podsumowując, można stwierdzić, że operacje morfologiczne stanowią rozbudowaną grupę bardzo uniwersalnych filtrów, odgrywających ważną rolę w procesach analizy obrazu. Podobnie jednak jak pozostałe przekształcenia, potwierdzają w pełni swoją użyteczność dopiero we współpracy z innymi przekształceniami.

9. Analiza obrazów kolorowych

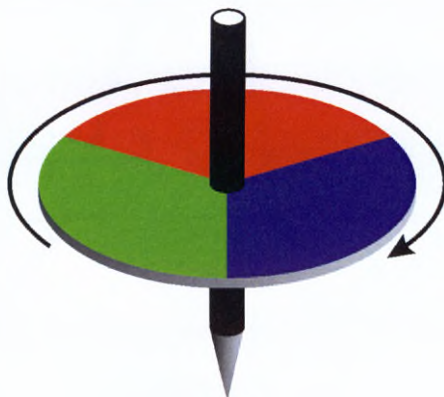
Analizę obrazów kolorowych zaczniemy od omówienia podstawowego sposobu ich zapisu. Powszechnie stosuje się w tym celu model RGB, którego nazwa jest akronimem składowych barw: Red, Green, Blue (z ang. 'czerwony', 'zielony', 'niebieski'). Poszczególne kolory powstają przez dodawanie składowych, dlatego też model RGB bywa nazywany modelem addytywnym (il. 9.1a.).

Punktem wyjścia w modelu RGB jest brak koloru, czyli czerń. Poszczególne kolory uzyskuje się przez dodanie w odpowiednich proporcjach jednej, dwóch lub wszystkich trzech składowych RGB. Jeżeli w pełni nasycone podstawowe barwy skojarzymy w pary, to otrzymamy trzy kolejne kolory. Suma czerwonego z zielonym daje żółty (ang. *yellow*), czerwonego z niebieskim – purpurę (ang. *magenta*), a suma zielonego i niebieskiego prowadzi do koloru niebiesko-zielonego, który na ogół jest odbierany jako jasno-niebieski. Nie ma w języku polskim dobrej nazwy dla tego koloru i dlatego często używa się angielskiej nazwy *cyan*, niekiedy w spolszczonej wersji jako cyjan. Suma wszystkich składowych RGB prowadzi do koloru białego, co schematycznie przedstawia il. 9.1a.

Jeśli trudno jest nam uwierzyć, że suma trzech podstawowych barw modelu RGB daje biały kolor, to można wykonać proste doświadczenie. Z kawałka kartonu i ołówka można zrobić prosty bączek, pokazany na il. 9.1b, którego krążek jest oklejony kolorowym



a



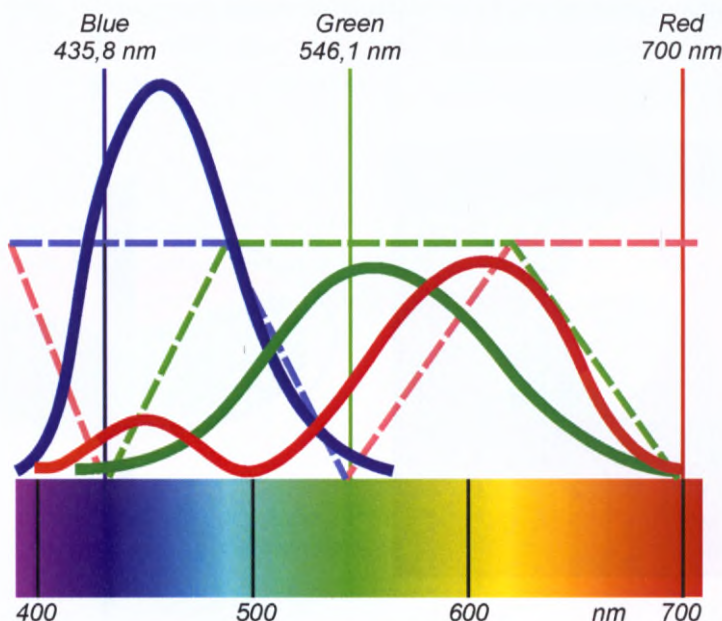
b

Il. 9.1. Model RGB. Opis w tekście

papierem, pomalowany lub pokolorowany kredkami tak, aby uzyskać trzy składowe RGB. Po nadaniu bączkowi szybkiego ruchu obrotowego powierzchnia krążka będzie sprawiała wrażenie jasnej plamy. Ze względu na niedoskonałości barwników zapewne nie uzyskamy w tym prostym doświadczeniu idealnej bieli.

Model RGB wykorzystuje cechy naszego zmysłu wzroku. Ludzkie oko posiada na powierzchni wewnętrznej części gałki ocznej dwie rodziny uczulonych na światło komórek, nazywanych pręcikami i czopkami. Czopki są wrażliwe na zmiany długości fali światła, czyli zmiany koloru. Z kolei pręciki reagują jedynie na zmiany natężenia światła, ale za to są znacznie bardziej czułe niż czopki. Dzięki temu przy słabym oświetleniu wciąż dostrzegamy kształty różnych obiektów, choć przestajemy widzieć ich kolory. Obserwacja tego zjawiska znalazła swoje odbicie w znanym przysłowiu, które mówi, że „w nocy wszystkie koty są szare”.

Badania ludzkiego zmysłu wzroku wykazały, że posiadamy trzy rodzaje czopków, które są odmiennie uczulone na światło o różnych długościach fali, co w sposób przybliżony pokazuje il. 9.2. Najbardziej czułe są czopki B, które osiągają maksimum czułości dla niebieskiego światła o długości fali ok. 450 nm. Czopki G maksimum czułości osiągają dla fali o długości ok. 550 nm, co odpowiada kolorowi zielonemu. Czułość czopków R jest najbardziej zróżnicowana. Swoje maksimum osiąga dla fali o długości nieco ponad 600 nm, co odpowiada bardziej barwie pomarańczowej niż czerwonej.



Il. 9.2. Porównanie czułości czopków oraz modelu RGB. Opis w tekście

Maksimum sumy sygnałów od czopków G i R przypada na barwę żółtą, co tłumaczy, dlaczego żółte światło lamp sodowych sprawia wrażenie jaśniejszego oświetlenia i dlaczego szkła okularów pomagających podczas jazdy we mgle są żółte. Nietrudno też zauważyć, że zaburzenia w działaniu jednej odmiany czopków spowodują upośledzenie rozpoznawania kolorów, znane pod nazwą daltonizmu.

Model RGB stanowi jedynie przybliżenie zmian czułości oka ludzkiego na różne długości fali światła. Sposób modelowania poszczególnych kolorów pokazują linie przerywane na il. 9.2. W górnej części rysunku zaznaczono długości fali światła składowych modelu RGB, natomiast w dolnej części pokazano tęczę, uzyskaną za jego pomocą.

Dzięki podobieństwu do czułości ludzkiego oka model RGB umożliwia rejestrację oraz wyświetlanie kolorowych obrazów. Wykorzystują go ekrany telewizorów oraz smartfonów, monitory komputerów, kamery, skanery i projektory. Jest zatem wykorzystywany również w grafice komputerowej i analizie obrazu.

Najczęściej kolory w modelu RGB zapisywane są za pomocą 24 bitów (po 8 bitów dla każdej składowej). Daje to łącznie potężną liczbę 16 777 216 kolorów, ale ze względu na uproszczenia modelu nie pozwala na oddanie wszystkich kolorów dostrzeganych przez ludzkie oko. Najłatwiejsze do zauważenia są niedostatki modelu RGB w odwzorowaniu różnych odcieni czerwieni. Jeżeli na ekranie widzimy czerwony obiekt, to jest on mało zróżnicowany i sprawia wrażenie jednolitej plamy. Jest to dobrze widoczne na il. 9.3.

Niezależnie od braku możliwości odwzorowania wszystkich barw dostrzeganych przez ludzkie oko, który w praktyce nie jest zbyt dokuczliwy, obrazy zapisane zgodnie z modelem RGB mają inną bardzo poważną wadę: są trudne do korekcji. Problem ten ukazuje il. 9.4. Dla naszego zmysłu wzroku wcale nie jest oczywiste, że w obrazie z il. 9.4a

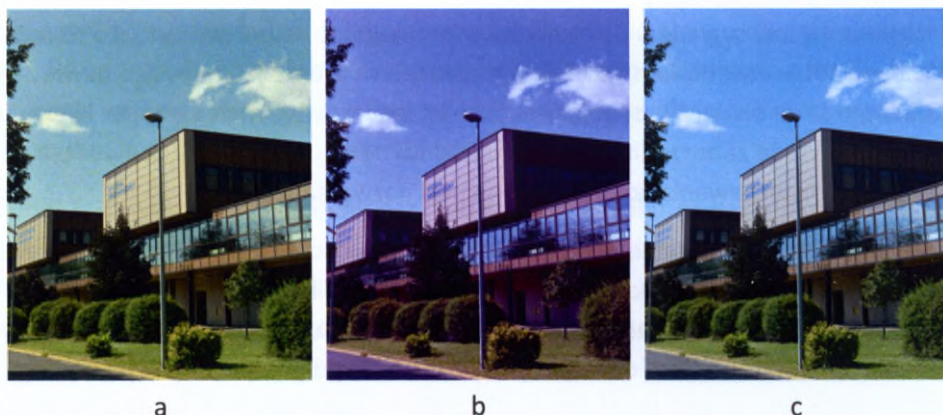


a



b

Il. 9.3. Niedoskonałości modelu RGB. Zdjęcie czerwonego nadwozia legendarnego Jaguara E-type z 1969 roku (a) wykazuje wyraźnie mniejsze zróżnicowanie niż kremowego nadwozia Lancii Aurelii z 1955 roku (b). Zdjęcia wykonano w Muzeum Motoryzacji w Turynie

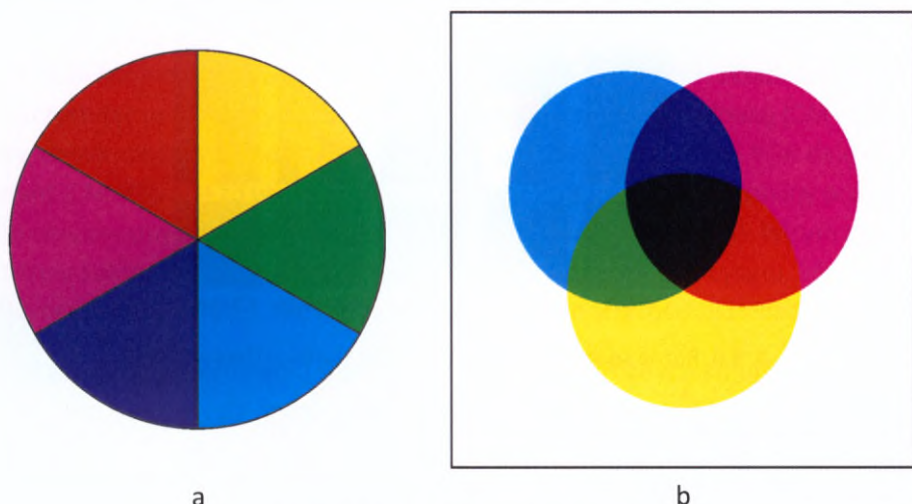


Il. 9.4. Korekcja obrazów RGB: a) obraz z niedostatkim składowej niebieskiej, b) obraz z niedostatkim składowej zielonej, c) obraz z prawidłowymi barwami

trzeba wzmocnić składową niebieską, a w obrazie z il. 9.4b trzeba wzmocnić składową zieloną, aby otrzymać prawidłowe barwy widoczne na il. 9.4c.

Dlaczego tak się dzieje, skoro model RGB został opracowany na podstawie badania cech ludzkiego zmysłu wzroku? Wyjaśnienie leży w analizie sposobu powstawania barw w naturze. Najbardziej naturalnym źródłem światła jest Słońce, które oświetla poszczególne obiekty światłem białym, które stanowi mieszaninę różnych długości fal z zakresu od ok. 400 nm do ok. 700 nm. Jeżeli obiekt oświetlony Słońcem odbija wszystkie składowe widma, to widzimy go jako biały. Jeżeli obserwowany obiekt jest np. zielony, to dzieje się tak dlatego, że jego powierzchnia odbija tylko fale o długości z zakresu ok. 520-570 nm, a pozostałe długości fal są pochłaniane. Tak więc kolory dostrzegamy dzięki temu, że z widma światła białego pewne składowe są odejmowane. Możemy to zjawisko w przybliżeniu opisać za pomocą modelu CMY, zwanego też modelem subtraktywnym. Na il. 9.5a pokazane jest koło barw, wynikające z modelu RGB. Trzy kolory uzupełniające składowe RGB mogą powstać dzięki dodawaniu odpowiednio skojarzonych barw podstawowych. Jest zatem logiczne, że można też definiować barwy na podstawie tych kolorów uzupełniających, czyli niebiesko-zielonego, purpury i żółtego. Pierwsze litery angielskich nazw tych kolorów (*Cyan*, *Magenta*, *Yellow*) tworzą nazwę modelu, który jest pokazany na il. 9.5b. Ten model kolorów jest dla nas znacznie łatwiejszy do interpretacji, gdyż pasuje do naszego doświadczenia ze światłem słonecznym. Na przykład bez problemu zgadniemy, że nadmiar żółtego koloru spowoduje zmianę czerwonego na pomarańczowy.

Dotychczasowe rozważania można podsumować następująco: Model RGB opisuje sposób generowania obrazów przez monitory, które są źródłem kolorowego światła.



II. 9.5. Koło barw (a) i model CMY (b). Opis w tekście

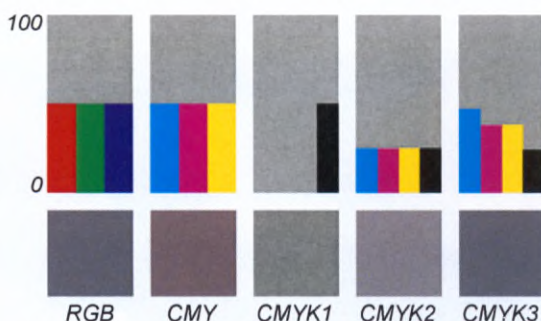
Takie zjawisko nie występuje w naturze i dlatego korekcja obrazów RGB jest dla nas trudna. Nie mamy po prostu doświadczenia z takim sposobem generowania kolorów. Natomiast model CMY dość dobrze opisuje powstawanie kolorów dzięki oświetleniu kolorowej powierzchni białym światłem, na przykład słonecznym.

Najczęściej składowe RGB opisuje się w skali 0-255, co odpowiada 8-bitowej głębi poszczególnych składowych. Natomiast model CMY częściej opisuje się w skali 0-100, co pozwala na definiowanie udziału tych składowych w procentach. Przejście z modelu RGB na CMY można wtedy zapisać następująco:

$$\begin{aligned}C &= (255-R)/2,55 \\M &= (255-G)/2,55 \\Y &= (255-B)/2,55\end{aligned}\tag{9.1}$$

gdzie C , M , Y , R , G oraz B oznaczają składowe kolory modelu.

Model CMY nie jest w praktyce używany w analizie obrazu, ale omówimy szerzej jego modyfikację, czyli model CMYK, gdyż jest on powszechnie wykorzystywany w druku. Ponieważ model CMY dość dobrze odwzorowuje nasze doświadczenie z kolorami, powstającymi dzięki częściowemu odbiciu światła białego, to wydaje się być dobrym narzędziem do konstrukcji urządzeń drukujących. W istocie zrobiono tak w przypadku pierwszych drukarek atramentowych. Niestety, wyniki były dalekie od doskonałości. Wynikowe kolory zależą bowiem od barwników, rodzaju podłoża, w tym jego nasiąkliwości dla atramentu lub farby, a także własności optycznych powierzchni.



Il. 9.6. Różne sposoby generowania 50% szarości. Opis w tekście

Na przykład pokrycie kolorowego wydruku na matowym papierze błyszczącą folią daje wrażenie zdecydowanie bardziej wyrazistych kolorów. Największym niedostatkim drukarek pracujących zgodnie z modelem CMY był brak możliwości uzyskania dobrej czerni. Druk zawierający trzy kolory był nieco rozmyty, a ponadto zamiast czystej czerni zwykle otrzymywało się brzydki, ciemnobrunatny kolor. Skutecznym rozwiązaniem okazało się dodanie czwartego koloru, którym jest czern. Do nazwy modelu dodano ostatnią literę angielskiego słowa *black*.

Przyjrzyjmy się teraz, jak zaczerńnienie na poziomie 50% może być przeprowadzone z wykorzystaniem poszczególnych modeli (il. 9.6). W górnej części rysunku mamy wykresy słupkowe, ilustrujące udział poszczególnych składowych z modeli RGB, CMY oraz CMYK. W dolnej części rysunku mamy przybliżony wygląd danego koloru. W przypadku modelu CMYK wykorzystano trzy różne warianty, oznaczone odpowiednio: CMYK1, CMYK2 i CMYK3. 50% szarości w modelu RGB uzyskujemy przez przypisanie wartości 128 (50% zakresu 0-255) wszystkim składowym. Przejście z modelu RGB na CMY opisuje równanie (9.1) i tutaj również wszystkie składowe są sobie równe i mają wartość 50%. Warto zwrócić uwagę, że wygenerowany w ten sposób kolor jest raczej brązowy, a nie szary. Jeżeli dysponujemy czarnym kolorem w modelu CMYK, to możemy szarość uzyskać za pomocą tylko jednego koloru, czyli czarnego (CMYK1). Takie rozwiązanie daje znaczne oszczędności w zużyciu tuszu lub farby drukarskiej. Przejście z modelu CMY na CMYK można w tym przypadku opisać równaniem:

$$\begin{aligned}
 C &= C1 - \min(C1, M1, Y1) \\
 M &= M1 - \min(C1, M1, Y1) \\
 Y &= Y1 - \min(C1, M1, Y1) \\
 K &= \min(C1, M1, Y1)
 \end{aligned}
 \tag{9.2}$$

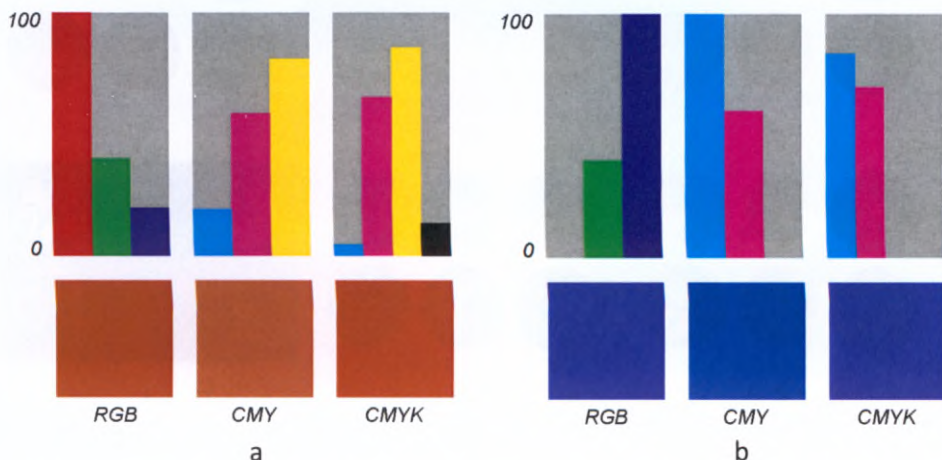
gdzie C , M , Y oraz K oznaczają składowe modelu CMYK, a $C1$, $M1$ oraz $Y1$ – składowe modelu CMY.

Opisana powyżej konwersja daje szarość zamiast brązu, ale jest dość odległa od szarości RGB. Można próbować generować szarość ze wszystkich składowych CMYK, wykorzystując w tym przypadku 25% nasycenia dla każdej składowej (CMYK2). Niestety, wynik nie jest wcale lepszy. Najlepszą zgodność z modelem RGB daje ostatni wariant (CMYK3). Jest to wynik konwersji modelu RGB na CMYK, dokonany przez algorytm wbudowany do profesjonalnego oprogramowania graficznego. W tym przypadku mamy kolor oznaczony jako C47M38Y38K24, co oznacza 47% cyjanu, 38% purpury i żółci oraz 24% czarnego. Nie ma tu już jednakowych wartości dla składowych CMY.

Udział czerni w generowaniu szarości zmienia się dla różnych kolorów. Odpowiedni algorytm opracowała firma Adobe i opatentowała go. Patentowanie algorytmów jest wprawdzie niemożliwe w Europie, ale w USA nie ma już takich ograniczeń. Dzięki temu firma Adobe czerpie zyski z udostępniania algorytmu producentom oprogramowania graficznego, np. firmie Corel. Tłumaczy to też, dlaczego nie ma darmowego oprogramowania, realizującego konwersję RGB na CMYK.

Niestety, nawet najbardziej zaawansowane algorytmy nie pozwalają na wierne oddanie części kolorów RGB za pomocą druku. Najbardziej dotkliwe jest to w przypadku różnych oieni niebieskiego, natomiast wiele innych kolorów jest odwzorowywanych bardzo dobrze (il. 9.7).

Brązowy kolor jest dość dobrze oddawany nawet przez model CMY, natomiast profesjonalny algorytm zamiany na paletę CMYK daje bardzo dobrą zgodność z modelem RGB (il. 9.7a). CMYK, uzyskany zgodnie z równaniem (9.2), niepokazany na il. 9.7, nie radzi sobie z zamianą z modelu RGB i daje zbyt jasny kolor. W przypadku koloru

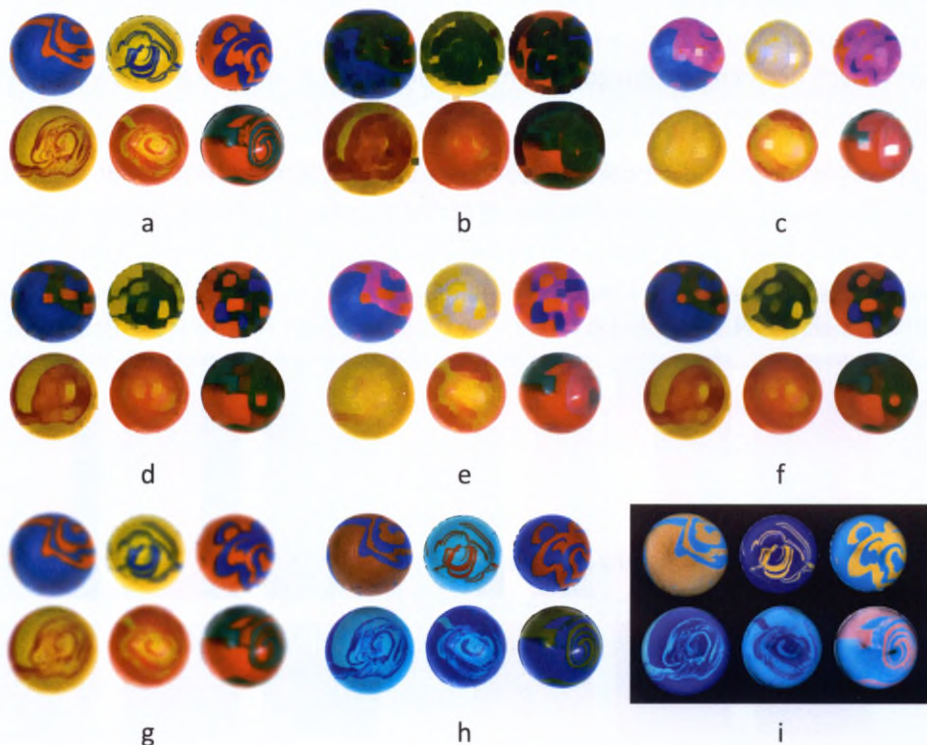


Il. 9.7. Konwersja koloru RGB na kolory zgodne z modelem CMY oraz CMYK. Opis w tekście

fioletowego model CMY oferuje kiepskie odwzorowanie modelu RGB. Znacznie lepiej radzi sobie komercyjny algorytm CMYK (il. 9.7b).

Analizując ilustracje w tym rozdziale, należy pamiętać, że wszystkie obrazy są drukowane zgodnie z modelem CMYK. Dotyczy to również obrazów RGB. Zatem różnice pomiędzy RGB, CMY i CMYK są tutaj jedynie pewnym przybliżeniem różnic występujących w rzeczywistości.

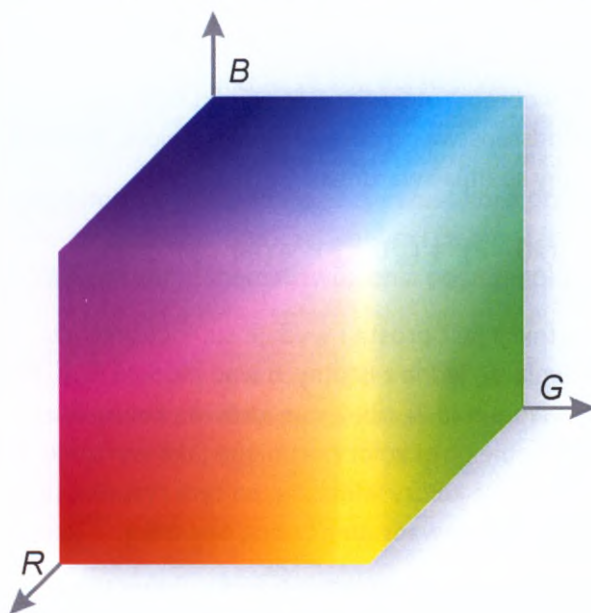
Ponieważ nie wszystkie kolory z przestrzeni RGB można wiernie oddać za pomocą urządzeń korzystających z modelu CMYK, to konwersja RGB-CMYK nie jest jednoznaczna. Inaczej mówiąc, część nieznacznie różniących się kolorów RGB jest przybliżana za pomocą jednego zestawu barw CMYK. Jeżeli zatem chcielibyśmy dokonać powrotnej zamiany CMYK na RGB, to otrzymamy kolory nieco inne od początkowych. W zaawansowanych systemach druku, nastawionych na możliwie jak najbardziej wierne oddawanie kolorów, używa się dodatkowych barw, aby pokonać ograniczenia modelu CMYK. Na tym kończymy analizę modelu CMYK, gdyż zajmujemy się analizą obrazu, a nie problemami druku.



Il. 9.8. Kolorowy obraz piłek (a) oraz jego proste przekształcenia: b) erozja, c) dylatacja, d) otwarcie, e) zamknięcie, f) mediana, g) rozmycie Gaussa, h) składowa R zamieniona miejscami ze składową B, i) negatyw

Podstawowe przekształcenia kolorowych obrazów, wbudowane w systemy analizy obrazu, polegają na identycznej obróbce poszczególnych składowych, traktowanych jako obrazy szare. Wyniki ich działania są zatem dla nas dość łatwe do przewidzenia. Przykłady przetwarzania obrazu kolorowego przedstawia il. 9.8. Możliwości przetwarzania obrazów kolorowych są jednak zdecydowanie większe, gdyż możemy każdą składową przetwarzać w inny sposób, a ponadto wykonywać na tych składowych różne operacje, co zostanie omówione w dalszej części tego rozdziału. Przykłady takich możliwości zawierają il. 9.8h oraz il. 9.8i. W pierwszym przypadku wystarczyła zamiana składowej R na składową B oraz składowej B na składową R. Uzyskany obraz jest zdecydowanie różny w odbiorze od wyjściowego obrazu, pokazanego na il. 9.8a. Również wykonanie negatywu obrazu kolorowego (il. 9.8i) daje zupełnie inne odczucia w porównaniu z obrazem początkowym. Tego typu operacje są często wykorzystywane w celu uczynienia obrazu bardziej czytelnym, co nie ma jednak bezpośredniego związku z analizą obrazu.

W przypadku analizy obrazów kolorowych najbardziej interesujące nie są jednak efekty ich przetwarzania, tylko możliwość detekcji obiektów na podstawie informacji o ich kolorach. Na obrazie szarym obiekty czerwone, zielone i niebieskie wyglądają identycznie i nie ma sposobu ich odseparowania, jeżeli nie różnią się innymi cechami, czyli np. wielkością, kształtem lub fakturą powierzchni. Włączenie koloru daje



Il. 9.9. Sześcian RGB. Opis w tekście

natychmiast dodatkowe możliwości analizy. Jest to ważne m.in. w badaniach histologicznych, gdzie pierwotnie bezbarwne preparaty są specjalnie barwione, aby móc analizować budowę tkanek.

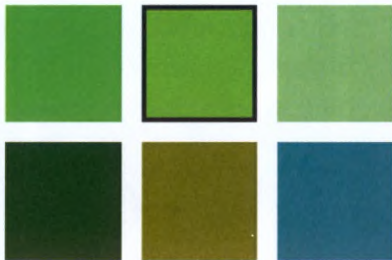
Model RGB jest bardzo popularnym narzędziem do opisu koloru nie tylko z powodu jego wykorzystania w wielu urządzeniach, ale też ze względu na swoją prostotę. Poszczególne kolory można łatwo interpretować dzięki wspomnianej już addytywności modelu. Wygodnym sposobem przedstawiania poszczególnych kolorów jest sześciian RGB (il. 9.9). Jest on umieszczony w przestrzennym układzie współrzędnych, którego osie odpowiadają składowym modelowi. Oczywiście płaska grafika nie pozwala na pokazanie wszystkich kolorów. To, co widzimy na il. 9.9, to symulacja tych kolorów, w których co najmniej jedna ze składowych RGB osiąga wartość maksymalną. Chcąc pokazać inne kolory, musielibyśmy analizować przekroje tego sześcianu.

Sześciian RGB pozwala na geometryczną interpretację barw. Każdy z 16 777 216 kolorów w 24-bitowym obrazie jest reprezentowany przez jeden punkt, którego współrzędne oznaczmy jako r_0 , g_0 i b_0 . Załóżmy, że chcemy uzyskać binarny obraz obiektu, który ma kolor opisany przez podane powyżej współrzędne. Wybór wszystkich punktów o tych współrzędnych najprawdopodobniej nie da nam dobrego wyniku detekcji, gdyż nawet w przypadku jednolitej barwy obiektu jego punkty będą się na obrazie nieco różniły. Możemy jednak łatwo zdefiniować odpowiednią procedurę detekcji. Wystarczy przyjąć, że poszukujemy obiektu, którego kolory są zbliżone do początkowego punktu o współrzędnych r_0 , g_0 i b_0 . Przez „zbliżone” będziemy rozumieli kolory leżące w obrębie otaczającej ten punkt kuli o promieniu r , który jest miarą dopuszczalnego odstępstwa od poszukiwanego koloru. Wykryte punkty o składowych r , g i b powinny spełniać warunek:

$$(r_0 - r)^2 + (g_0 - g)^2 + (b_0 - b)^2 \leq r^2 \quad (9.3)$$

Taka procedura detekcji jest prosta i wydaje się logicznie uzasadniona. Niestety, w tym miejscu ujawnia się jedna z istotnych wad modelu RGB. Podobieństwo kolorów, zdefiniowane wzorem (9.3), obejmuje zarówno kolory oceniane przez człowieka jako zbliżone do siebie, jak i kolory oceniane jako wyraźnie różne. Problem ten ukazuje il. 9.10. Wszystkie kolory widoczne na tym rysunku są jednakowo odległe w sześcianie RGB od koloru kwadratu z czarną obwódką.

Zatem detekcja obiektów na podstawie ich koloru okazuje się w praktyce znacznie trudniejsza, niż przypuszczaliśmy. W dalszej części rozdziału zajmiemy się różnymi sposobami, pozwalającymi na przynajmniej częściowe rozwiązanie tego problemu.



Il. 9.10. Zielony kwadrat w czarnej obwódce, otoczony kwadratami, których kolory są jednakowo odległe od zielonego w sześciu RGB

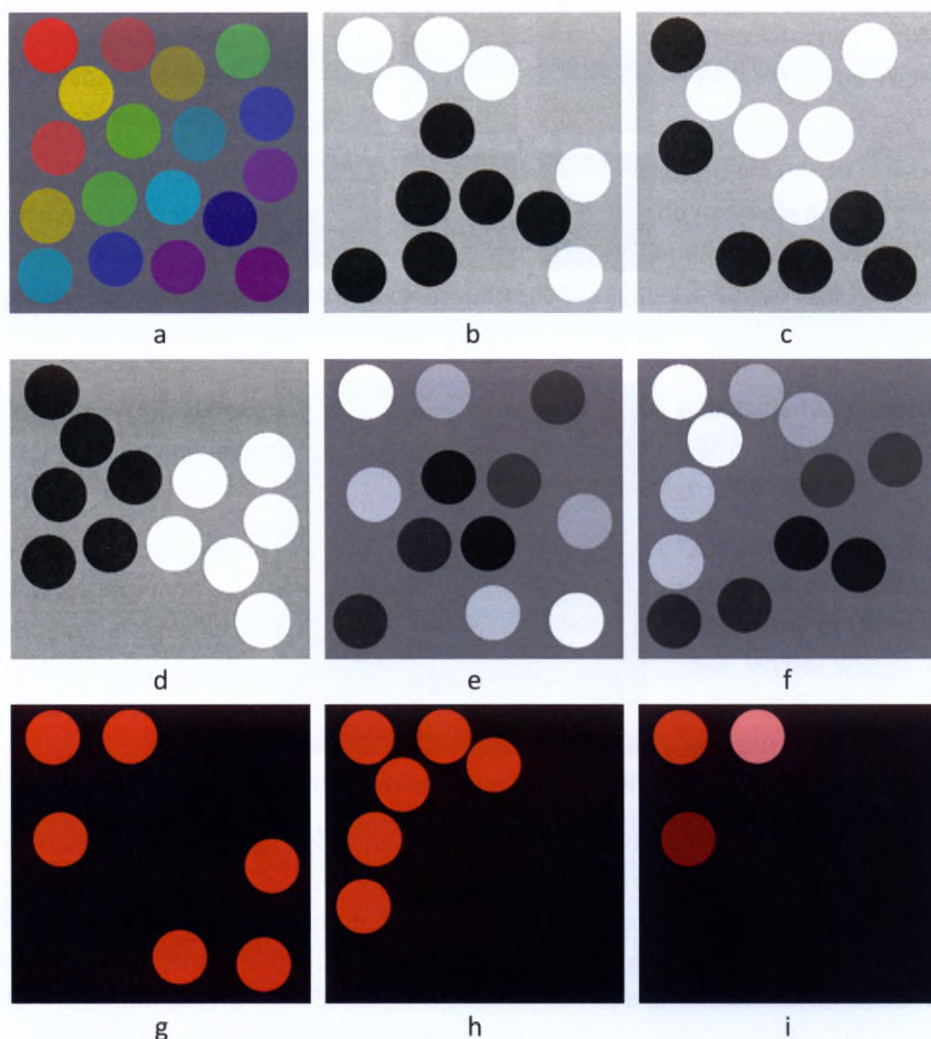


Il. 9.11. Kolorowe kółka. Opis w tekście

Przeanalizujemy teraz problem detekcji kolorowych kół na testowym obrazie, pokazanym na il. 9.11. Zawiera on łącznie 18 kół. Sześć z nich ma kolory znane z koła barw. Są one rozmieszczone w przybliżeniu wzdłuż przekątnej kwadratu. Do celów wizualizacji połączono je na il. 9.11 czarną linią. Powyżej przekątnej mamy sześć kół z rozjaśnionymi kolorami, a poniżej znajduje się sześć kół z przyciemnionymi kolorami. Koła w kolorze czerwonym, jasnoczerwonym i ciemnoczerwonym zaznaczono na il. 9.11 białymi obwódkami. Schemat tworzenia poszczególnych kolorów można prześledzić w tabeli 9.1 na przykładzie koloru zielonego oraz purpurowego. Pozostałe kolory są tworzone w analogiczny sposób.

Tabela 9.1. Składowe RGB koloru zielonego i purpurowego w obrazie z il. 9.11

Kolor kółka	Składowa R	Składowa G	Składowa B
Zielony	0	255	0
Ciemnozielony	0	152	0
Jasnozielony	152	255	152
Purpurowy	255	0	255
Ciemnopurpurowy	152	0	152
Jasnopurpurowy	255	152	255



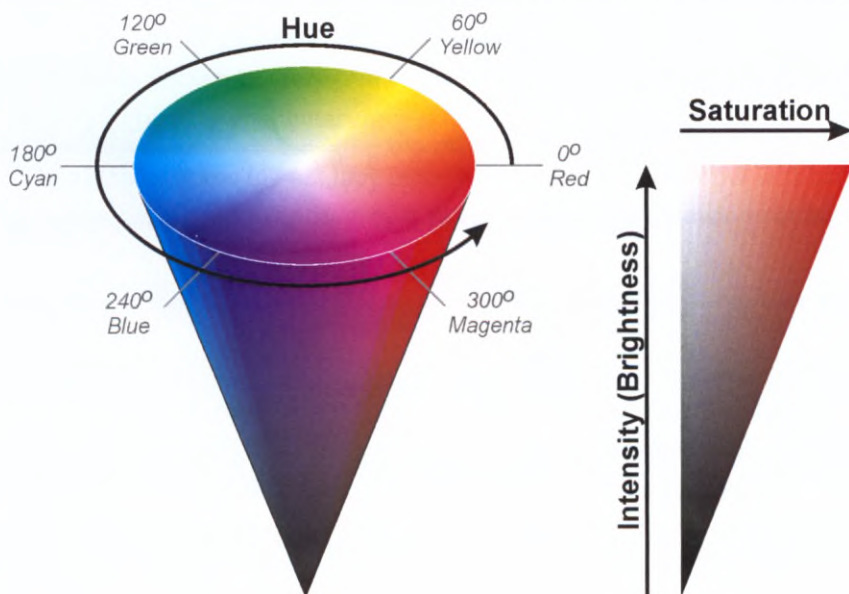
Il. 9.12. Detekcja czerwonych kółek: a) obraz kolorowych kółek; składowe obrazu: b) czerwona R, c) zielona G i d) niebieska B; różnice składowych: e) R-G, f) R-B; obrazy binarne różnic: g) R-G, h) R-B oraz i) końcowy wynik detekcji

Spróbujmy wykonać proste zadanie, polegające na wykryciu kół mających kolor czerwony, który może być rozjaśniony lub przyciemniony. Rozbicie obrazu wyjściowego (il. 9.12a) na składowe RGB (il. 9.12b-d) pokazuje, że elementy składowej czerwonej wchodzą w skład koloru żółtego oraz purpurowego. Jeżeli jednak odejmiemy składową zieloną i niebieską od składowej czerwonej (il. 9.12e oraz 9.12f), to wyeliminujemy część kół, w których występuje składowa czerwona. Ponieważ wynik odejmowania może być ujemny, to 50% szarości na il. 9.12e oraz 9.12f odpowiada wartości zerowej. Miejsca ciemniejsze niż 50% szarości odpowiadają wartościom ujemnym.

Binaryzacja obrazów różnic prowadzi do powstania obrazów 9.12g oraz 9.12h. Iloczyn logiczny tych obrazów (AND) da nam poszukiwany obszar występowania koloru czerwonego i pozwoli na skuteczną końcową detekcję, której wynik przedstawia il. 9.12i.

Detekcja kół koloru zielonego i niebieskiego będzie miała podobny przebieg, jak omówiona powyżej detekcja czerwonych kół. Zmienia się jedynie wykorzystywane do przekształceń obrazy. Bardziej złożona będzie jednak detekcja pozostałych kolorów: niebiesko-zielonego, purpurowego oraz żółtego.

Omówimy teraz detekcję żółtych kół. W tym przypadku wykryjemy najpierw obszary, w których występuje kolor czerwony lub zielony, gdyż właśnie ich suma daje w modelu RGB kolor żółty. Następnie odejmiemy od nich obszary występowania barwy czerwonej (il. 9.12i), zielonej oraz niebieskiej. Pamiętamy przy tym, że procedura ich wyznaczania również jest dość złożona. Omawiany sposób wykrywania żółtych obiektów jest skomplikowany, choć dotyczy bardzo prostego obrazu testowego. Nie można wprost wykorzystać przedstawionych powyżej procedur do wykrywania obszarów o dowolnej barwie w złożonych obrazach, zawierających wiele odcieni różnych kolorów. Prosta detekcja na podstawie zdefiniowania jednego koloru oraz odległości do zbliżonych barw, wchodzących w skład wykrywanego obiektu, może prowadzić w przypadku modelu RGB do poważnych błędów, co pokazano już na il. 9.10. Musimy zatem poszukiwać innych narzędzi do detekcji kolorowych obiektów.



Il. 9.13. Model HSI. Opis w tekście

Często pomocne (i co ważniejsze, skuteczne) może okazać się wykorzystanie innego modelu barw, nazywanego HSI (ang. *Hue, Saturation, Intensity*), który kolory opisuje za pomocą nowych trzech składowych: barwy, nasycenia oraz intensywności. Model ten bywa czasami nazywany również modelem HSB (*Hue, Saturation, Brightness*) lub HSV (*Hue, Saturation, Value*). Wszystkie te nazwy dotyczą jednak tego samego modelu, który został pokazany na il. 9.13.

Składowa barwy (*Hue*) zawiera kolory znane nam już z modelu RGB i pokazane w kole barw (il. 9.5a). Kolory zmieniają się w kierunku przeciwnym do wskazówek zegara i każdy kolor jest opisany za pomocą kąta. Kąt o wartości zerowej przypisano czerwieni. Poszczególne implementacje modelu HSI używają albo kątów wyrażonych w stopniach, albo w mierze łukowej, czyli w radianach.

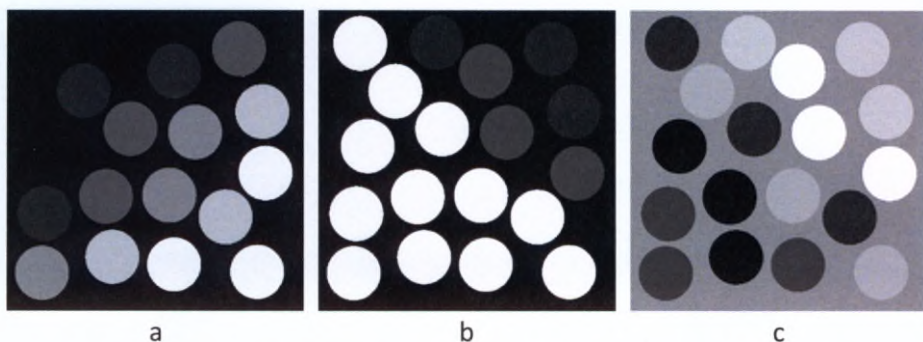
Kolejną składową jest nasycenie (*Saturation*). Przyjmuje się, że kolor biały ma zerowe nasycenie, a pełne 100% nasycenia osiągają kolory z koła barw oraz ich kombinacje. Ostatnią składową jest jasność (*Brightness*). Przyjmuje się, że kolor czarny ma zerową jasność, a biały wykazuje maksimum jasności. Składowa jasności to po prostu obraz szary, czyli średnia arytmetyczna składowych RGB. Przejście ze składowych RGB na składowe HSI opisuje równanie (9.4), które wymaga wcześniejszej normalizacji obrazu RGB, czyli przeliczenia poszczególnych składowych z zakresu 0-255 na zakres 0-1. Równanie (9.4) jest dość złożone dla składowej H . Znane są też inne formuły obliczania składowej H , które nie korzystają z funkcji trygonometrycznych i są dzięki temu mniej złożone obliczeniowo.

$$H = \begin{cases} \theta & \text{dla } B \leq G \\ 360 - \theta & \text{dla } B > G \end{cases} \quad \theta = \arccos \frac{1}{2} \frac{(R - G) + (R - B)}{\sqrt{(R - G)^2 + (R - B)(G - B)}} \quad (9.4)$$

$$S = 1 - 3 \frac{\min(R, G, B)}{R + G + B}$$

$$I = \frac{1}{3} (R + G + B)$$

Wadą modelu HSI jest fakt istnienia osobliwości dla koloru czarnego, białego oraz wszystkich odcieni szarości, czyli sytuacji, w których $R=G=B$. Aby uniknąć dzielenia przez zero, w algorytmy można wbudować odpowiednie korekty, ale nie zmienia to faktu, że składowa H dla tych kolorów wynosi 0, tak samo jak dla składowej czerwonej. Pomimo tej niedogodności, dzięki zerowemu nasyceniu i odpowiedniej jasności, dostajemy w obrazie czarny, szary lub biały kolor.



Il. 9.14. Składowe HSI obrazu 9.12a: a) barwa, b) nasycenie, c) intensywność

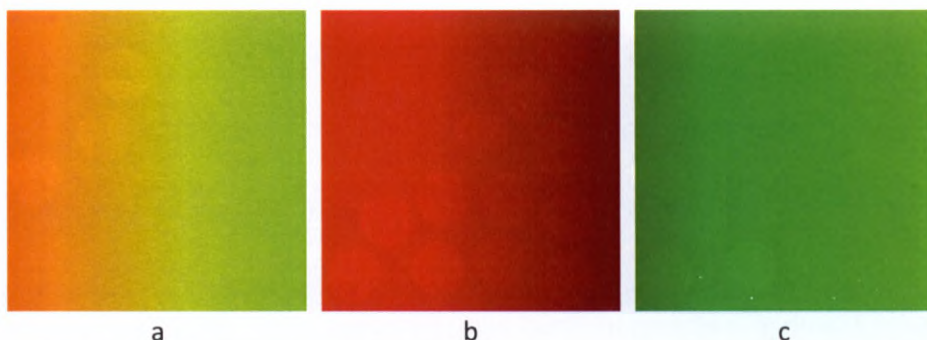
Il. 9.14 przedstawia składowe HSI obrazu z il. 9.12a. Wykrycie czerwonych elementów nie nastręcza teraz trudności. Trzeba jednak pamiętać, że barwa czerwona jest pokazywana tak samo jak czerni, biel i wszystkie odcienie szarości. Do ich eliminacji można jednak łatwo wykorzystać składową S , gdyż wszystkie wymienione wyżej kolory, oprócz czerwonego, wykazują nasycenie o wartości zerowej.

Zastosowanie LUT, odpowiadającego modelowi RGB pokazanemu na il. 9.2, pozwala na pokazanie składowej H w postaci obrazu kolorowego, na którym można zaobserwować rozmieszczenie poszczególnych barw. Widoczne jest wyraźnie, że niezależnie od tego, czy kolor był jasny, czy ciemny, składowa H dobrze rozpoznała poszczególne barwy (il. 9.15). Trzeba pamiętać, że wizualizacja kolorów, pokazana na il. 9.15, ma jedynie znaczenie poglądowe. Przy korzystaniu z modelu HSI należy znać wartości kątów odpowiadających poszczególnym barwom lub obserwować wykrywane obszary podczas interaktywnej binaryzacji składowej H .

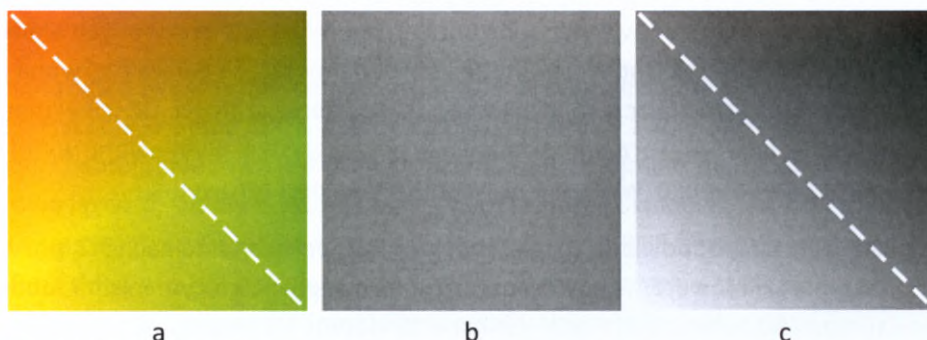
Model HSI został zbudowany tak, aby umożliwić opisywanie kolorów w sposób zbliżony do używanego w życiu codziennym. Na przykład kolor różowy możemy opisać



Il. 9.15. Składowa H obrazu 9.12a po konwersji na model HSI, zamieniona na obraz kolorowy w celu wizualizacji zakodowanych barw



Il. 9.16. a) przykład zakresu kolorów trudnego do wykrycia z wykorzystaniem modelu RGB, b) składowa R , c) składowa G . Opis w tekście



Il. 9.17. a) kolory, które byłyby wykryte przy wykorzystaniu modelu RGB, b) obraz z il. 9.16a zamieniony na obraz szary, c) obraz z il. 9.17a zamieniony na obraz szary. Szczegóły w tekście

jako jasny (intensywny) czerwony o niewielkim nasyceniu. Jest to na pewno bardziej intuicyjne niż np. 255R200G200B, czyli maksymalna (255) wartość składowej czerwonej oraz wartości 200 dla składowej zielonej i niebieskiej.

Innym przykładem problemów z modelem RGB jest próba detekcji zakresu kolorów, widocznego na il. 9.16a. Obraz ten stanowi przejście od koloru czerwonego do zielonego i jest sumą obrazów składowych, pokazanych na il. 9.16b i 9.16c.

Gdybyśmy chcieli dokonać detekcji takiego zakresu kolorów z dowolnego obrazu RGB, to należałoby uwzględnić wszystkie kombinacje kolorów z obrazów 9.16b i 9.16c. W ten sposób bralibyśmy pod uwagę wszystkie kolory pokazane na il. 9.17a, podczas gdy potrzebujemy wykryć jedynie te, które leżą wzdłuż przekątnej obrazu, zaznaczonej białą przerywaną linią na il. 9.17a. Ponieważ natężenie kolorów w składowych obrazach z il. 9.16b i 9.16c zmienia się liniowo w takim samym zakresie wartości, to suma tych składowych ma stałą wartość dla wszystkich punktów obrazu (il. 9.17b), która występuje wzdłuż przekątnej il. 9.17c. Obraz na il. 9.17c można również uzyskać, zamieniając obraz RGB (il. 9.17a) na HSI i wybierając trzecią składową (*Intensity*).

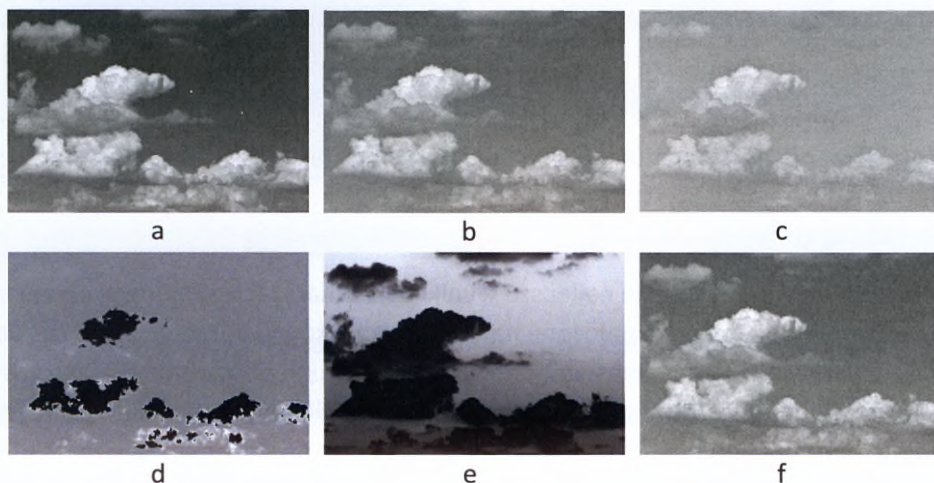


Il. 9.18. Obraz chmur

Sprawdźmy teraz, jak dotychczasowe doświadczenia z detekcją obiektów na obrazach kolorowych sprawdzą się w przypadku problemu wykrywania chmur na niebie (il. 9.18).

Widać wyraźnie, że w przypadku rozbicia obrazu RGB na składowe żadna z nich nie pozwala na skuteczne wykrycie chmur (il. 9.19a-c). Najlepsze wyniki dałaby składowa czerwona (il. 9.19a), gdyż wtedy niebo jest najciemniejsze i uzyskujemy największy kontrast z chmurami. Ale nawet w tym przypadku rozróżnienie chmur i nieba w dolnej części zdjęcia okazuje się niemożliwe.

Konwersja na model HSI (il. 9.19d-g) nie daje radykalnego rozwiązania, gdyż barwa dużej części chmur nie różni się istotnie od zamglonego nieba, zwłaszcza w dolnej



Il. 9.19. Składowe RGB (a, b oraz c) oraz składowe HSI (d, e oraz f) obrazu chmur



Il. 9.20. Wynik detekcji chmur, przeprowadzonej na podstawie składowej nasycenia, nałożony na obraz początkowy



Il. 9.21. Fragment murów kościoła z elementami pozostającymi częściowo w pełnym słońcu i częściowo zacienionymi

części obrazu. Można jednak wykorzystać zmiany nasycenia, które w obszarze chmur jest dość niskie w porównaniu z niebem (il. 9.19e). Końcowy wynik detekcji (il. 9.20) daje wprawdzie błędy w obrębie części delikatnych, pierzastych obłoków w górnej części obrazu, ale należy uznać go za akceptowalny.

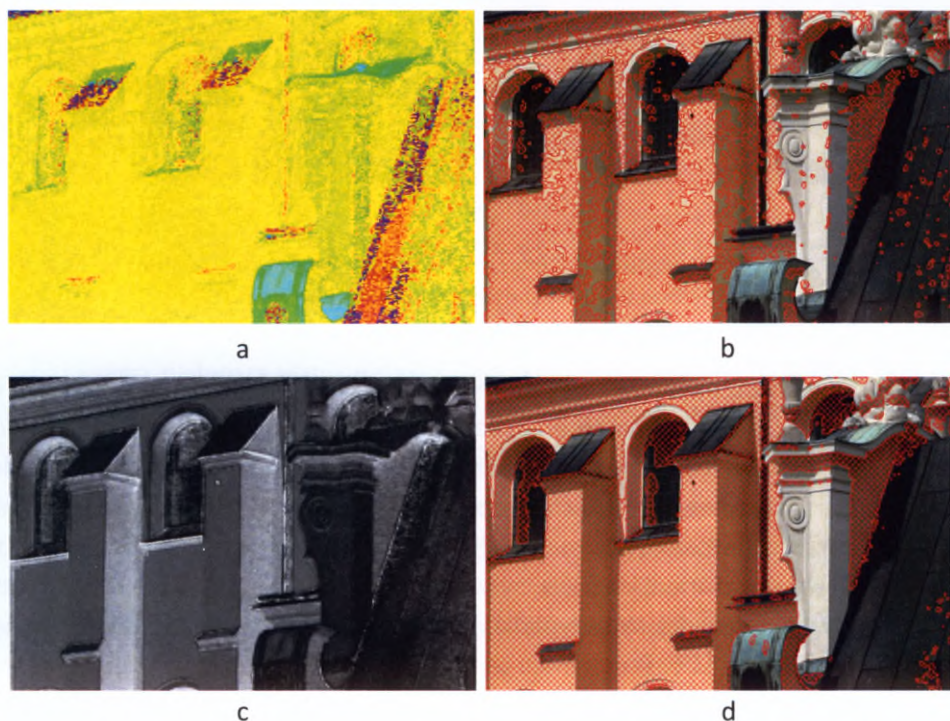
Poważnym problemem modelu RGB są trudności w detekcji obiektu oświetlonego w zróżnicowany sposób. Na przykład miejsca nasłonecznione i zacienione różnią się tak bardzo, że nie można ich rozpoznać jako jeden obiekt (il. 9.21).

W analizowanym przypadku najlepsze wyniki daje binaryzacja składowej czerwonej (il. 9.22a). Wynik detekcji, nałożony na obraz początkowy, pokazuje il. 9.22b. Udało się wykryć jedynie jasno oświetlone ściany, w tym niepotrzebnie dodany został jaśniejszy element ozdobny dachu z prawej strony. Gdybyśmy chcieli wykryć jeszcze elementy zacienione, to jednocześnie zaliczylibyśmy liczne elementy pokrycia dachowego. Zatem model RGB nie pozwala na dobrą detekcję elementów oświetlonych w różnym stopniu.



Il. 9.22. Składowa czerwona (a) oraz wynik detekcji muru nałożony na obraz początkowy (b)

Sprawdźmy teraz, czy konwersja na model HSI pozwoli na lepsze wyniki detekcji.



Il. 9.23. a) składowa H modelu HSI, b) wynik detekcji nałożony na obraz początkowy, c) składowa S , d) wynik detekcji nałożony na obraz początkowy

Składowa barwy jest tak mało zróżnicowana, że bardzo trudno ją wykorzystać do detekcji. Aby pozwolić Czytelnikowi na zobaczenie zmian w tej składowej, została ona pokazana w wersji zabarwionej (il. 9.23a). Niewątpliwie interesujące jest to,

że składowa barwy rzeczywiście nie jest czuła na zmiany oświetlenia. Obszary zacięnione wyglądają tak samo na obrazie składowej H , jak obszary bezpośrednio oświetlone światłem słonecznym. Wynik detekcji na podstawie binaryzacji składowej H jest znacznie lepszy, ale obserwujemy liczne braki w wykrytej powierzchni (il. 9.23b). Idea jest zatem bardzo dobra, ale występujące w praktyce wahania koloru powodują, że składowa H nie jest najlepszym rozwiązaniem przy tak subtelnych zmianach barwy.

Najlepiej sprawdza się składowa nasycenia (il. 9.23c). Zakres jej wartości został rozszerzony w celu lepszego pokazania zmian nasycenia. Wykorzystanie składowej S obrazu HSI pozwoliło na poprawne wykrycie jasnożółtych ścian, zarówno w słońcu, jak i w cieniu. Jedyne błęd polega na niepotrzebnym zaliczeniu fragmentów okien oraz niewielkich fragmentów dachu u dołu obrazu z prawej strony.

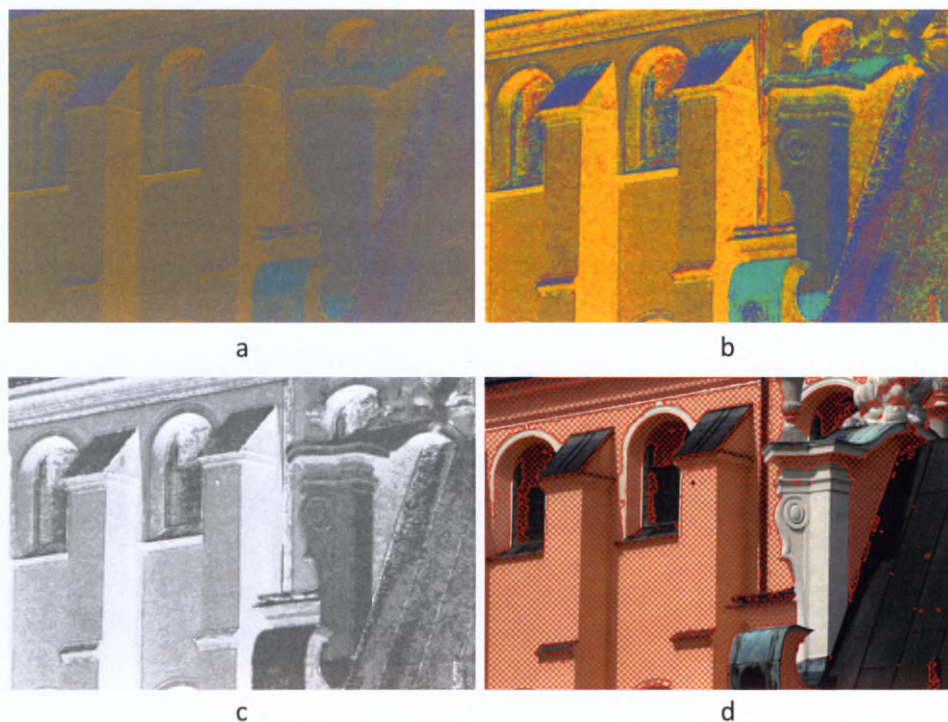
Wykrywanie kolorów można często łatwiej przeprowadzić, jeżeli użyje się znormalizowanego obrazu RGB. Odpowiednie składowe wyznacza się zgodnie z bardzo prostym równaniem (9.5):

$$\begin{aligned}r &= R/(R + G + B) \\g &= G/(R + G + B) \\b &= B/(R + G + B)\end{aligned}\tag{9.5}$$

gdzie r , g oraz b oznaczają znormalizowane wartości składowych RGB.

Wartości r , g oraz b mieszczą się w zakresie 0-1. Jeżeli pomnożymy je przez 255, to uzyskamy obraz, który można obserwować w dowolnym programie graficznym. Idea normalizacji obrazu RGB polega na założeniu, że w przypadku słabszego oświetlenia wszystkie składowe obniżają się proporcjonalnie. W takim przypadku znormalizowane wartości pozostają niezmiennie. W przypadku obrazu z il. 9.21 założenie to jest częściowo spełnione. Składowa czerwona w części zacięnionej stanowi 65% tej składowej w części jasnej. W przypadku składowej zielonej i niebieskiej jest to odpowiednio 62% oraz 51%. Problemem jest niestety część jasna, gdyż wtedy znacznie silniej rosną składowe RGB, odpowiedzialne za rozjaśnianie obrazu, a znormalizowane wartości składowych r , g , b maleją. Powoduje to, paradoksalnie, przyciemnienie normalizowanego obrazu RGB w obszarach jasnych w porównaniu do obszarów ciemnych. Widać to wyraźnie na il. 9.24a i b.

Znormalizowane obrazy RGB są często mało czytelne (il. 9.24a). Dlatego w celu poprawy czytelności na il. 9.24b pokazano ten sam obraz po wyrównaniu histogramu, co nie ma wpływu na wyniki manualnej detekcji. Podobnie jak w przypadku obrazu



Il. 9.24. a) znormalizowany obraz RGB, b) obraz (a) po wyrównaniu histogramu, c) składowa R oraz d) wynik detekcji nałożony na obraz początkowy

wyjściowego, najlepszy wynik detekcji daje składowa czerwona (il. 9.24c). Sam wynik należy ocenić jako co najmniej dobry (il. 9.24d). Jest on bardzo zbliżony do wyniku uzyskanego przy wykorzystaniu nasycenia z modelu HSI.

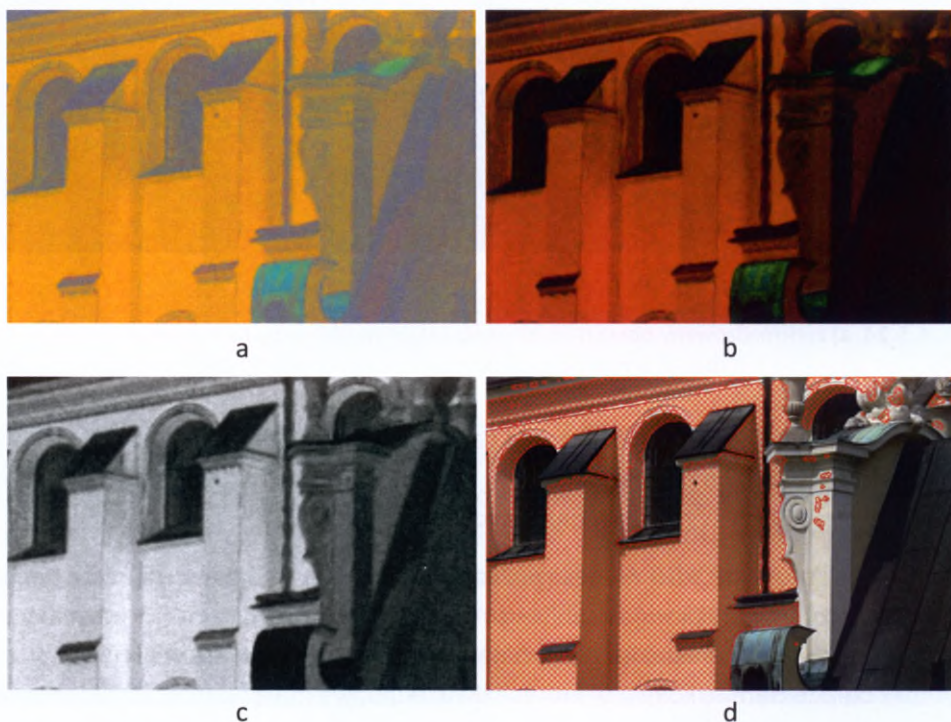
Normalizacja obrazu RGB nie jest pozbawiona wad, zwłaszcza w zakresie bardzo ciemnych kolorów. Wykazuje osobliwość dla czerni, gdy $R = G = B = 0$. Można zapobiec temu błędowi, zmieniając minimalną wartość dowolnej składowej z 0 na 1. Dalej jednak bardzo ciemne kolory, w których dwie składowe mają wartość 0, doprowadzą do dziwnej sytuacji. Załóżmy, że R przyjmuje wartość nieco większą od 0, $a = B = 0$. Wtedy $r = 1$, choć praktycznie R niewiele różni się od G i B , natomiast $g = b = 0$, co nie budzi już wątpliwości. Podobnie, jeżeli przyciemnianie kolorów nie zachodzi proporcjonalnie we wszystkich składowych, znormalizowany obraz RGB daje nieco zaskakujące wyniki. Problem ten jest szerzej omawiany w dalszej części rozdziału.

Kolejnym przekształceniem, pozwalającym na łatwiejsze wykrywanie niektórych kolorów, jest wyznaczenie „nadmiarowego” obrazu RGB (ang. *Excess RGB*), który wzmacnia różnice pomiędzy dominującym kolorem i pozostałymi składowymi:

$$\begin{aligned}
 excessR &= 2R - G - B \\
 excessG &= 2G - B - R \\
 excessB &= 2B - R - G
 \end{aligned}
 \tag{9.6}$$

gdzie: $excessR$, $excessG$ i $excessB$ oznaczają odpowiednio składowe RGB obrazu po przekształceniu.

Jak łatwo zauważyć, składowe nowego obrazu mogą przybierać zarówno wartości dodatnie, jak i ujemne. Dodatnie wartości otrzymuje się wtedy, gdy dana składowa ma wartość co najmniej dwukrotnie większą od sumy dwóch pozostałych. Sprawdzimy teraz, jaki wynik detekcji otrzymamy na podstawie binaryzacji składowej obrazu Excess RGB.



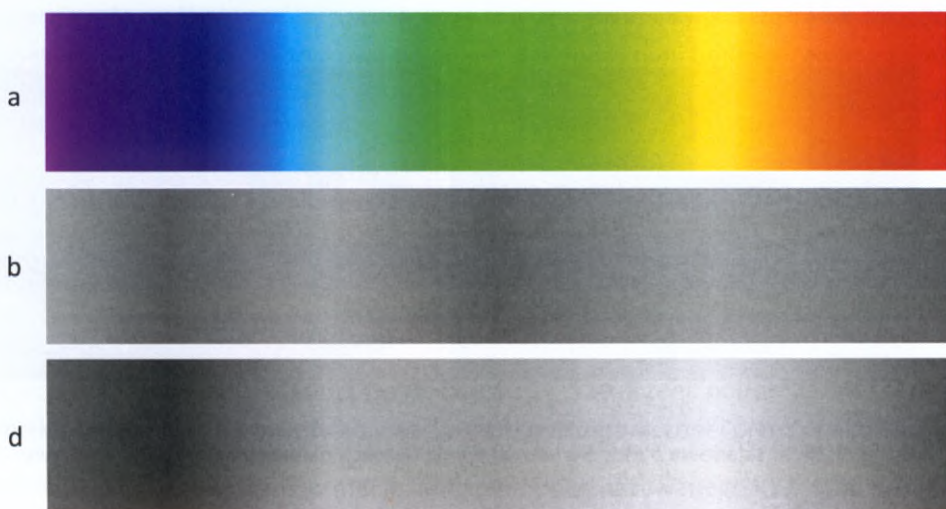
Il. 9.25. a) obraz Excess RGB, b) dodatnie wartości obrazu Excess RGB, c) składowa $excessR$, d) wynik detekcji nałożony na obraz wyjściowy

Ze względu na obecność ujemnych wartości obraz Excess RGB jest wyświetlany tak, że każda składowa przyjmuje 50% swojego zakresu dla wartości 0. Wszystko jaśniejsze niż 50% odpowiada wartościom dodatnim, natomiast ciemniejsze – wartościom ujemnym (il. 9.25a). Ponieważ interesuje nas przede wszystkim wykrycie tych kolorów, które mają dominujący udział w tworzeniu obrazu, możemy brać pod uwagę

tylko dodatnie wartości obrazu Excess RGB. Wszystkim wartościom ujemnym można przypisać wartość 0. Tak przygotowany obraz ma oczywiście nieco zmienione kolory i pokazuje go il. 9.25b. Podobnie jak w poprzednich przypadkach, najlepszy wynik detekcji daje składowa czerwona (tutaj *excessR*, il. 9.25c). Wynik detekcji nałożony na obraz wyjściowy przedstawia il. 9.25d. Wynik ten jest niemal perfekcyjny. Brakuje jedynie wykrycia trójkątnego obszaru z prawej strony obrazu.

Kolejnym modelem barw, który można wykorzystać w analizie obrazu, jest model YUV. Składowa *Y* odpowiada za jasność obrazu i nazywa się ją luminacją. Stanowi odpowiednik intensywności lub jasności znanej z modelu HSI, która jest średnią arytmetyczną składowych RGB. Luminacja jest jednak wyznaczana nieco inaczej. Zamiast średniej arytmetycznej składowych RGB mamy średnią ważoną, w której najważniejszą rolę pełni składowa *G*, natomiast składowa *B* jest brana pod uwagę jedynie w niewielkim stopniu. Odpowiednią formułę wyznaczania wartości *Y* zawiera wzór (9.7). Składowe *U* i *V* są kanałami chrominacji, czyli odpowiedzialnymi za barwy. *U* to przeskalowana składowa niebieska, natomiast *V* to przeskalowana składowa czerwona. Obie te składowe mogą przyjmować wartości zarówno dodatnie, jak i ujemne.

$$\begin{aligned} Y &= 0,299 \cdot R + 0,587 \cdot G + 0,114 \cdot B \\ U &= 0,492 \cdot (B - Y) \\ V &= 0,877 \cdot (R - Y) \end{aligned} \quad (9.7)$$



Il. 9.26. a) tęcza wygenerowana za pomocą modelu RGB, b) intensywność z modelu HSI, c) luminacja z modelu YUV

Il. 9.26 pokazuje, że luminacja z modelu YUV lepiej oddaje odczucia naszego zmysłu wzroku w zakresie jasności barw niż intensywność. Fragment obrazu luminacji (il. 9.26c) odpowiadający barwie niebieskiej jest najciemniejszy, zaś odpowiadający barwie żółtej – najjaśniejszy, co odpowiada naszym odczuciom.

Model YUV był przez pewien czas wykorzystywany w transmisji obrazu telewizyjnego. Posiadacze starych, czarno-białych odbiorników mogli ich nadal używać, gdyż odbierali jedynie składową Y , dającą dobry obraz monochromatyczny. Natomiast telewizory kolorowe korzystały z trzech kanałów YUV. Najlepszy wynik detekcji dała binaryzacja składowej U (il. 9.27a). Końcowy wynik, pokazany na il. 9.27b, jest bardzo dobry, ale również niepozbawiony błędów. Podstawowy błąd to niepotrzebna detekcja jasnego filara na attyce dachu z prawej strony obrazu.



Il. 9.27. a) składowa U modelu YUV, b) wynik detekcji nałożony na obraz początkowy



Il. 9.28. a) składowa b modelu Lab, b) wynik detekcji nałożony na obraz początkowy

Ostatnim modelem barw, który krótko omówimy, jest model Lab. Konwersja z modelu RGB przebiega dwuetapowo. Najpierw należy wyznaczyć wartości dla modelu XYZ. Macierzowy zapis tej operacji zawiera równanie (9.8):

$$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = \begin{bmatrix} 0,49000 & 0,31000 & 0,20000 \\ 0,17697 & 0,81240 & 0,01063 \\ 0,00000 & 0,01000 & 0,99000 \end{bmatrix} \quad (9.8)$$

Następnie wyznacza się wartości L , a i b :

$$\begin{aligned} L &= 116 \sqrt[3]{\frac{Y}{Y_0}} - 16 \\ a &= 500 \left(\sqrt[3]{\frac{X}{X_0}} - \sqrt[3]{\frac{Y}{Y_0}} \right) \\ b &= 200 \left(\sqrt[3]{\frac{Y}{Y_0}} - \sqrt[3]{\frac{Z}{Z_0}} \right) \end{aligned} \quad (9.9)$$

gdzie: $X_0 = 94,81$; $Y_0 = 100,0$; $Z_0 = 107,3$.

Wspominamy o tym skomplikowanym obliczeniowo i pojęciowo modelu między innymi dlatego, że detekcja wykonana na składowej b modelu Lab (il. 9.28a) daje niemal perfekcyjny wynik (il. 9.28b). Oczywiście nie oznacza to, że model Lab jest zdecydowanie najlepszy i należy zrezygnować z używania innych modeli.

Model Lab, a właściwie CIELab (CIE – *Comission Internationale de l'Eclairage*), został opracowany w 1976 r. w wyniku prac nad postrzeganiem kolorów przez człowieka, w tym przede wszystkim różnic pomiędzy kolorami. Celem było stworzenie takiego modelu barw, w którym jednakowa odległość pomiędzy nimi będzie oceniana przez człowieka jako identyczna różnica pomiędzy nimi. Ma to służyć m.in. do obiektywnej oceny jakości wydruków. W modelu Lab występują trzy składowe: luminancja lub jasność, zmieniająca się w zakresie od 0 do 100 (L), barwa zmieniająca się od zielonej do purpury w zakresie od -150 do 100 (a) oraz barwa zmieniająca się od niebieskiej do żółtej w zakresie od -100 do 150 (b).

Jest to zatem przestrzenny model barw. W odróżnieniu jednak od sześcianu RGB niektóre obszary punktów w przestrzeni Lab nie opisują żadnego koloru występującego w rzeczywistości. Kolory zajmują jedynie część tej przestrzeni, podobnie jak np. pojemnik z perfumami zajmuje tylko część przestrzeni ograniczonej przez opakowanie.

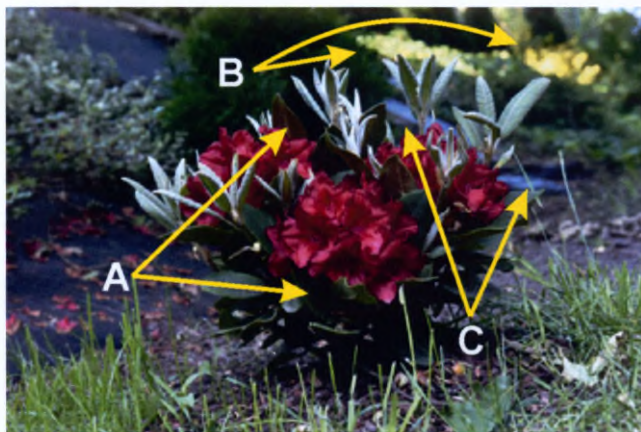
Model CIELab stanowi transformację innego modelu, nazywanego XYZ. Stąd do wyznaczenia składowych CIELab potrzebne jest wyznaczenie wartości składowych modelu XYZ, co opisuje równanie (9.8). Modelem XYZ nie będziemy zajmowali się bliżej w ramach niniejszego podręcznika.

Model CIELab został opracowany tak, aby w możliwie największym stopniu wyeliminować poważną wadę modelu RGB, przedstawioną na il. 9.10, a polegającą na tym, że ta sama odległość dwóch kolorów w sześcianie RGB nie oznacza takiej samej wizualnej różnicy tych kolorów. Można zatem wykorzystać przestrzeń Lab do wykrywania podobnie wyglądających kolorów na podstawie analizy ich odległości euklidesowej w trójwymiarowej przestrzeni Lab. Odległość kolorów w tej przestrzeni określa wzór:

$$\Delta E = \sqrt{(\Delta L)^2 + (\Delta a)^2 + (\Delta b)^2} \quad (9.10)$$

gdzie: symbol Δ oznacza odległość lub różnicę.

Poznaliśmy kilka narzędzi, które mogą pomóc w detekcji obiektów w kolorowych obrazach. Wydaje się, że ze względu na problemy z odpowiednią interpretacją składowych U i V modelu YUV oraz składowych a i b modelu CIELab, częściej będziemy korzystali z konwersji RGB na HSI oraz prostych przekształceń obrazu RGB: jego normalizacji oraz operacji Excess RGB. Dokonamy teraz porównania tych narzędzi na nowym obrazie testowym, którym będzie zdjęcie kwitnącego krzewu rododendrona (il. 9.29).



Il. 9.29. Obraz kwitnącego rododendrona z zaznaczonymi zagadnieniami do rozwiązania metodami analizy obrazu. Opis w tekście

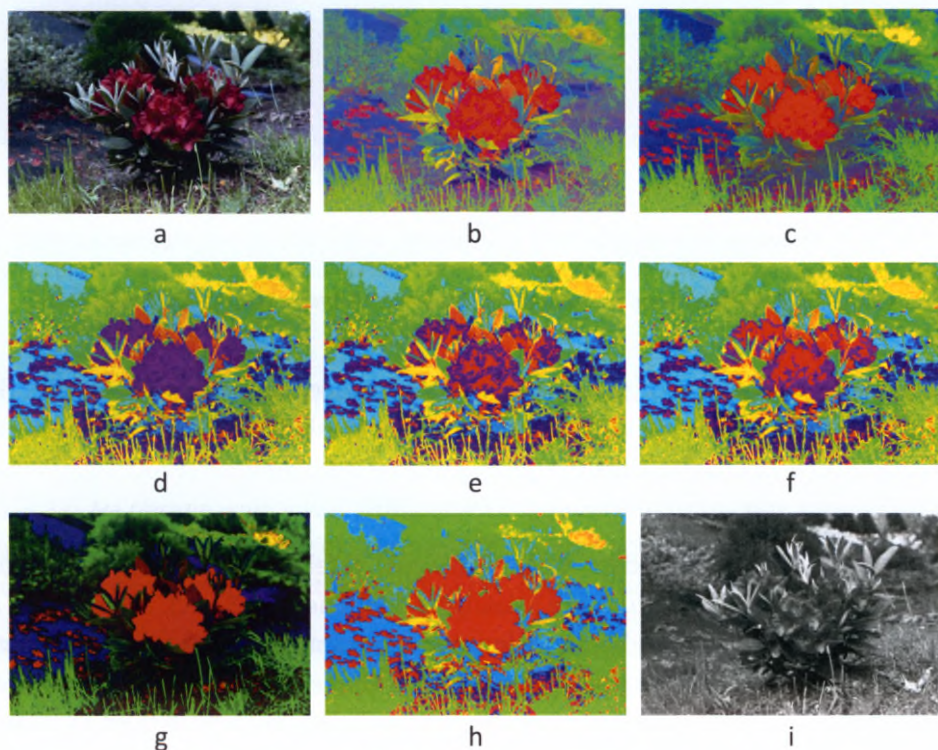
Omawiany obraz może posłużyć do prób rozwiązania metodami analizy obrazu m.in. następujących zagadnień:

- wykrycia kwitnącego kwiatu,
- wykrycia zielonych elementów: trawy, liści i krzewów, które różnią się oświetleniem oraz odcieniem,
- wykrycia bardzo ciemnych liści (elementy oznaczone literą A na il. 9.29),
- wykrycia obszaru żółtych kwiatów (elementy oznaczone literą B na il. 9.29),

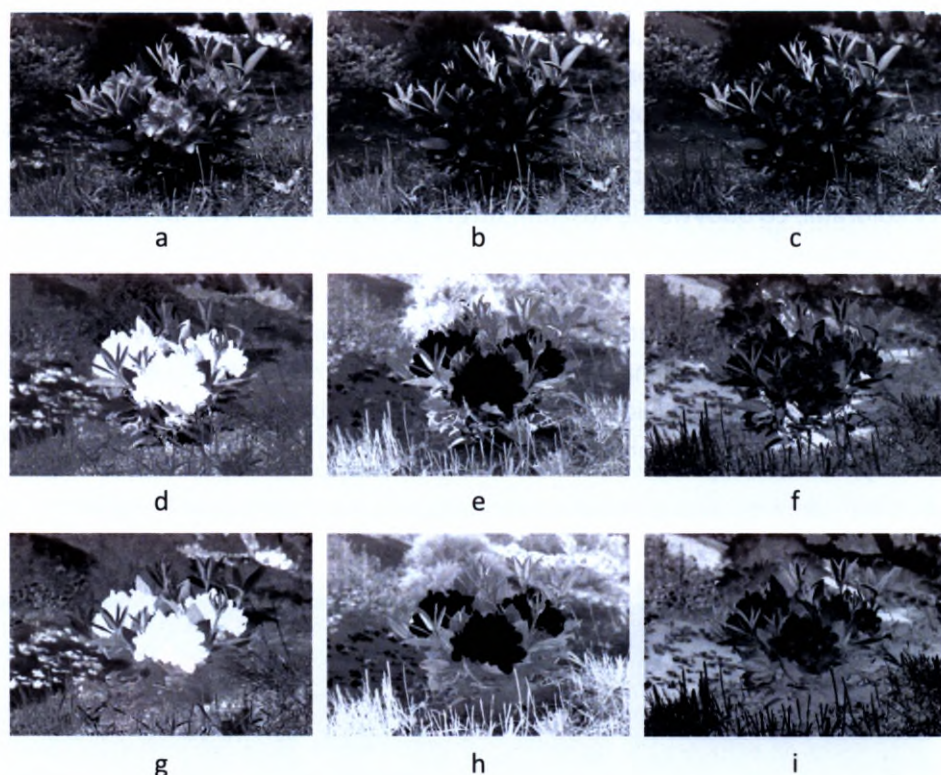
- wykrycia jasnoniebieskiego elementu, będącego mocno oświetlonym kawałkiem folii (elementy oznaczone literą C na il. 9.29).

Il. 9.30a-c przedstawiają kolejno: wyjściowy i znormalizowany obraz RGB oraz Excess RGB. Widać, że przekształcenia wyjściowego obrazu wzmacniają różnice pomiędzy kolorami, co odbywa się kosztem utraty szczegółów. W tym momencie jednak najważniejsze są właśnie obszary poszczególnych kolorów. Szczegóły mogą zostać zachowane i analizowane np. na obrazie szarym, odpowiadającym intensywności (il. 9.30i). Wizualny efekt normalizacji oraz Excess RGB jest bardzo wyraźny, ale nie wpływa w istotny sposób na barwy wykrywane po przekształceniu w obraz HSI. Jedynie wyraźne różnice są widoczne w postaci zmiany obszarów rozpoznawanych jako czerwone i purpurowe w obrębie kwiatów (il. 9.30d-f).

Rodzi się zatem pytanie o celowość stosowania normalizacji oraz Excess RGB. Używanie tych przekształceń jest jak najbardziej celowe z co najmniej dwóch powodów.



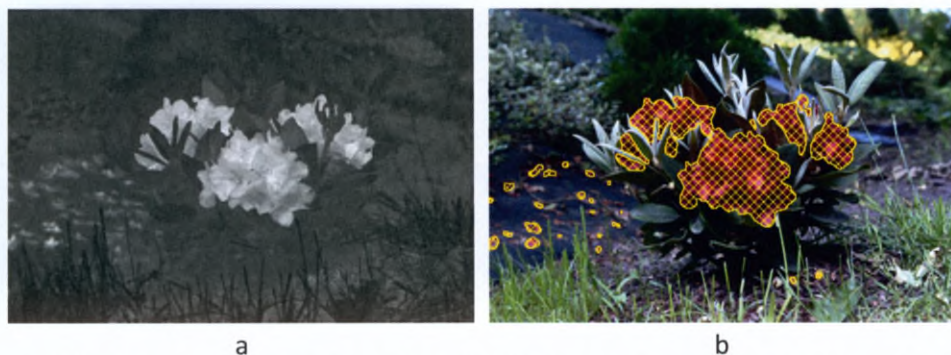
Il. 9.30. a) wyjściowy obraz do analizy, b) znormalizowany obraz RGB, c) obraz Excess RGB, d-f) zabarwione składowe *Hue* (barwa) obrazów a-c, g) obraz Excess RGB ograniczony do wartości dodatnich, h) zabarwiona składowa *Hue* obrazu g, i) obraz szary (*Intensity*). W celu poprawy czytelności zakresy barw w obrazach b, c oraz g zostały rozszerzone do zakresu 0-255



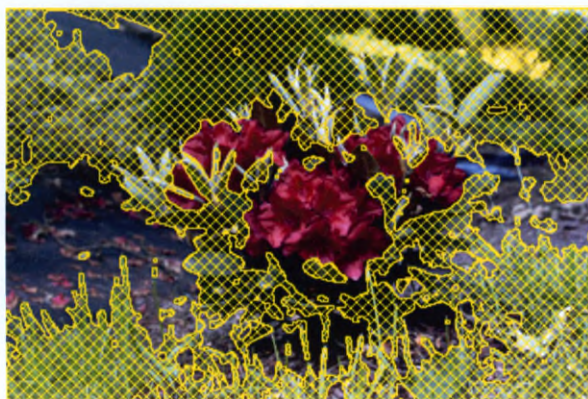
Il. 9.31. Składowe RGB obrazów: a-c) wyjściowego, d-f) znormalizowanego obrazu RGB, g-i) obrazu Excess RGB. W celu poprawy czytelności zakresy barw w obrazie znormalizowanym i Excess RGB zostały przed podziałem na składowe RGB rozszerzone do zakresu 0-255

Po pierwsze, mogą pomóc we wstępnej, wizualnej ocenie obrazów. Po drugie, okazują się przydatne podczas detekcji wykorzystującej składowe RGB, a nie model HSI. Jest to dobrze widoczne na il. 9.31, gdzie można porównać składowe RGB obrazu wyjściowego oraz znormalizowanego i Excess RGB.

Wyjściowy obraz RGB nie daje wielu możliwości detekcji poszczególnych kolorów. Same składowe RGB są wizualnie zaskakująco podobne (il. 9.31a-c). Jeżeli jednak od składowej czerwonej odejmiemy zieloną, to uzyskamy wyraźnie widoczny obszar dominacji koloru czerwonego (il. 9.32a). Binaryzacja tego obrazu pozwala na detekcję czerwonych kwiatów oraz opadłych płatków (il. 9.32b). Użycie tylko składowej czerwonej do wykrycia czerwonych elementów było niewystarczające, gdyż buduje ona również inne kolory występujące w obrazie, np. żółty lub jasnozielony. Tak więc wyjściowy obraz RGB bez dodatkowych przekształceń pozwala jedynie na zrealizowanie pierwszego zadania, czyli wykrycie kwiatu.



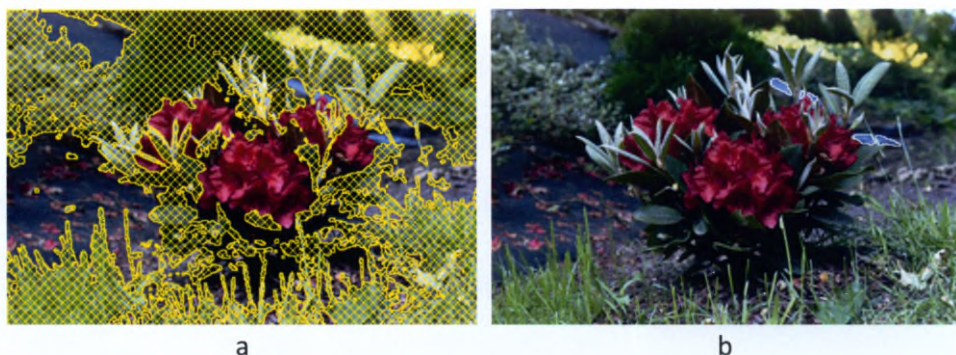
Il. 9.32. a) różnica składowych R i G obrazu RGB, b) detekcja czerwonego kwiatu i opadłych płatków na podstawie binaryzacji obrazu (a)



Il. 9.33. Detekcja obszarów zielonych na podstawie składowej zielonej znormalizowanego obrazu RGB

Znormalizowany obraz RGB pozwala na podobną detekcję czerwonego kwiatu jak wyjściowy obraz RGB. Dodatkowo jednak, dzięki uwydatnieniu niektórych kolorów, można na podstawie składowej zielonej dokonać skutecznej detekcji zielonych obiektów (il. 9.33). Na obrazie wyjściowym mają one różną jasność, która jednak została w pewnym stopniu ujednoliconą dzięki normalizacji. Fakt ten pokazuje, że normalizacja obrazu RGB może w pewnych sytuacjach być przydatna w analizie kolorowych obrazów.

Przekształcenie Excess RGB jeszcze silniej rozdziela poszczególne kolory, a przede wszystkim wzmacnia dominujące składowe. Dzięki temu detekcja elementów zielonych jest jeszcze łatwiejsza niż w znormalizowanym obrazie RGB (il. 9.34a). Dodatkowo, przez wzmocnienie składowej niebieskiej, możliwa jest również bezpośrednia detekcja jasnoniebieskiego obszaru na podstawie binaryzacji składowej niebieskiej (il. 9.34b). W ten sposób udało się wykonać kolejne zadanie. Powyższe wyniki potwierdzają zatem przydatność przekształcenia Excess RGB.



Il. 9.34. a) detekcja na podstawie składowej *G* obrazu Excess RGB, b) detekcja na podstawie składowej *B* obrazu dodatnich wartości Excess RGB. Opis w tekście

Model HSI wykazał ograniczoną skuteczność w przypadku detekcji elementów ścian kościoła (il. 9.21). Sprawdźmy teraz, jak zachowa się w przypadku detekcji obiektów na obrazie kwitnącego rododendrona i czy wykaże jakąś przewagę nad znormalizowanym obrazem RGB i obrazem Excess RGB. Składowa barwy (*Hue*) umożliwia bardzo łatwą detekcję obszaru zielonego (il. 9.35a). Wynik jest bardzo podobny do uzyskanego na podstawie obrazu Excess RGB (il. 9.34a). Dodatkowo, niewielka zmiana zakresu zmian barwy w składowej *Hue* pozwala na wyeliminowanie obszaru żółtych kwiatów (il. 9.35b), co nie było do tej pory możliwe, gdyż składowa zielona buduje również kolor czerwony w modelu RGB.

Eliminacja żółtych kwiatów powoduje jednocześnie eliminację części ciemnych liści w pobliżu kwiatu (il. 9.35b). Bliższa analiza wyjściowego obrazu (patrz fragment na il. 9.35c) wskazuje, że widoczna spodnia część liści jest w niektórych miejscach bardziej czerwona niż zielona. Składowa zielona (*G*) ma w tym miejscu wartości wynoszące ok. 1/3 wartości dla składowej czerwonej (*R*). Dlatego też ten element został poprawnie zakwalifikowany jako żółty.

Komputer miał w tym przypadku rację. Nasze oko, co prawda, widzi lekko brązowy kolor spodniej części liścia, ale informacja ta jest korygowana przez mózg, który wie, że liść powinien być w tym przypadku zielony.

Próba detekcji obszaru żółtych kwiatów na podstawie składowej barwy (*Hue*) prowadzi do wyniku, pokazanego na il. 9.35d w postaci żółtych plam. Obejmują one zarówno żółte kwiaty, jak i brązowe spodnie części liści. Niestety, model HSI nie przewiduje barwy brązowej. Ten kolor można opisać jako żółty o niskim nasyceniu oraz intensywności, co tłumaczy jego detekcję w kanale *Hue*. Jeżeli jednak dodatkowo wykorzystamy składową intensywności, wykryjemy za jej pomocą interesujące nas



Il. 9.35. a-b) detekcje na podstawie składowej *Hue* z różnym zakresem zmian barwy, c) fragment obrazu wyjściowego, d) detekcja ze składowej *Hue* (żółty obszar) oraz z *H AND I* (czerwony obszar) e) detekcja ze składowej *Hue* (czerwony zakresowany obszar) oraz z *H AND I* (zielony obszar), f) detekcja obszarów o nasyceniu powyżej 50%. Szczegółowy opis w tekście

żółte obszary, gdyż wykazują one dużą intensywność. Oczywiście wykryjemy również inne obszary o wysokiej intensywności, ale wspólna część (logiczne AND) obszarów o barwie żółtej oraz obszarów o dużej intensywności da nam poszukiwany obszar żółtych kwiatów. Ten końcowy wynik detekcji pokazano na il. 9.35d w postaci czerwonych plamek.

Jeżeli chcielibyśmy wykryć ciemne liście, które są brązowe, a nie zielone, to wystarczy wykorzystać detekcję obiektów o barwie żółtej i odjąć od wyniku omówiony wyżej obszar żółtych kwiatów.

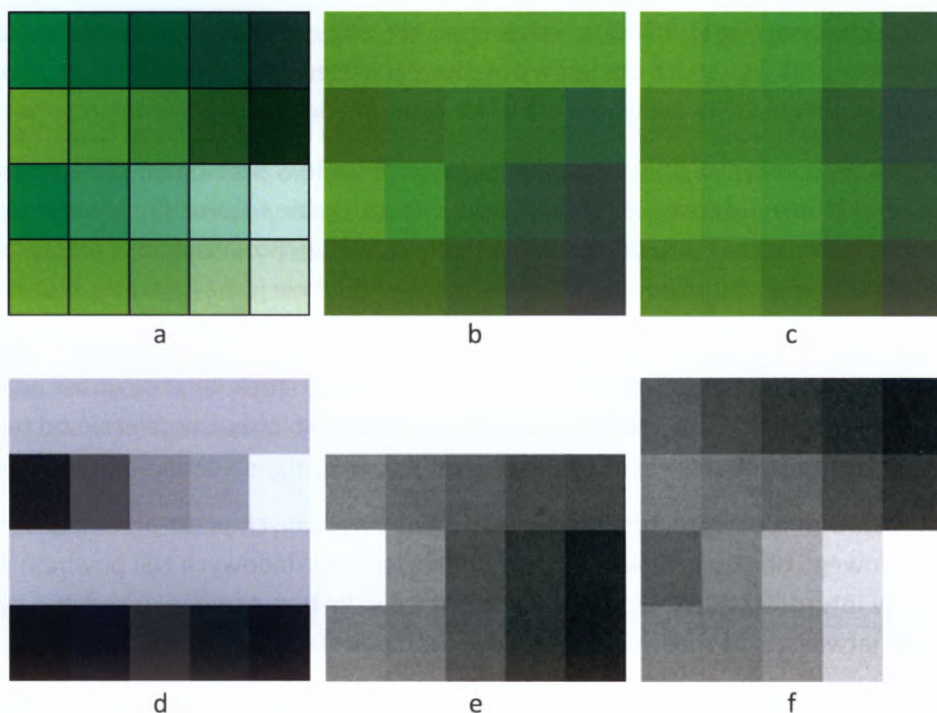
W przestrzeni HSI możemy wykryć również jasnoniebieskie elementy, co udało się poprzednio zrobić dopiero po przekształceniu Excess RGB. W tym celu wybieramy w składowej *Hue* barwę niebieską. W analizowanym przez nas obrazie barwę tę posiada znaczna część jego powierzchni, co pokazuje zakreskowany na czerwono obszar il. 9.35e. Podobnie jak w przypadku obszaru żółtych kwiatów, poszukiwany jasnoniebieski obszar wykrywamy, wykorzystując składową intensywności. Końcowy wynik detekcji pokazują zielone plamki na il. 9.35e.

W przedstawionych przykładach nie była wykorzystywana składowa nasycenia. Aby pokazać, że trudno jest wyobrazić sobie, co w istocie oznacza nasycenie koloru, na obrazie z il. 9.35f zakreskowano na czerwono obszary, w których nasycenie wynosi co najmniej 50%. Jest to niemal wyłącznie obszar czerwonego kwiatu. Pozostałe miejsca, zarówno jasne, jak i ciemne, są słabo nasycone. Dla naszego zmysłu wzroku stanowi to zazwyczaj niespodziankę.

Przeanalizujemy jeszcze jeden testowy obraz, ilustrujący wiele problemów występujących w analizie obrazów kolorowych. Testowy obraz (il. 9.36a) składa się z 20 prostokątów wypełnionych różnymi odcieniami koloru zielonego. Niektóre kolory są dla ludzkiego oka bardzo podobne, zwłaszcza w trzecim wierszu, dlatego na il. 9.36a dodano czarne linie, rozgraniczające poszczególne prostokąty. Pozostałe obrazy z il. 9.36 powstały już bez udziału tych dodatkowych linii.

Pierwsze od lewej prostokąty w pierwszym i trzecim (licząc od góry) wierszu są wypełnione kolorem zielonym, czyli czystą składową *G* modelu RGB. Kolejne prostokąty w pierwszym wierszu są coraz ciemniejsze z powodu malejącej wartości składowej *G*, przy zachowaniu zerowych wartości składowych *R* i *B*. Natomiast kolejne prostokąty w trzecim wierszu są coraz jaśniejsze, gdyż do składowej *G* dodawane są coraz większe wartości składowych *R* i *B*, przy czym obie te składowe mają zawsze takie same wartości.

Wiersze drugi i czwarty powstały w wyniku przetworzenia wierszy pierwszego oraz trzeciego. Pierwotna przestrzeń RGB została zamieniona na przestrzeń CMYK, w której nie mamy dokładnych odpowiedników kolorów RGB, co powoduje widoczne dla oka różnice. Następnie przestrzeń CMYK została zamieniona powtórnie na przestrzeń RGB. Zatem kolory w drugim i czwartym wierszu stanowią przybliżoną wizualizację możliwie dobrego dopasowania przestrzeni CMYK.



Il. 9.36. a) testowy obraz RGB omówiony w tekście, b) znormalizowany obraz RGB, c) obraz Excess RGB, d) składowa *Hue* z wartościami proporcjonalnie rozciągniętymi w celu uwidocznienia różnic, e) *Saturation*, f) *Intensity*

Normalizacja obrazu testowego (il. 9.36b) prowadzi do ujednolicenia koloru w pierwszym wierszu, co jest bezpośrednią konsekwencją sposobu prowadzenia normalizacji. Z kolei wartości w trzecim wierszu maleją, choć odpowiednie kolory w obrazie testowym są coraz jaśniejsze (jest to widoczne na obrazie intensywności na il. 9.36f). Kolory w drugim i czwartym wierszu są zbliżone do siebie, co wynika z udziału składowych *R* i *B* w kolorach prostokątów z tych rzędów.

Ciekawie wygląda pierwszy i trzeci wiersz w obrazie Excess RGB (il. 9.36c). W obu przypadkach zmiany są takie same. Wynikają one z faktu, że zarówno podczas przyciemniania, jak i rozjaśniania składowa zielona stawiała się mniej dominująca. W przypadku przyciemniania jest to wynikiem zmniejszania wartości składowej *G* przy pozostałych składowych na poziomie zerowym, a w przypadku rozjaśniania dominację składowej *G* osłabiają rosnące wartości składowych *R* i *B*. Drugi i czwarty wiersz wygląda podobnie jak w znormalizowanym obrazie RGB.

Przechodzimy teraz do przestrzeni HSI. Barwa każdego z prostokątów pozostaje zielona i różnice w składowej *Hue* są niemal niedostrzegalne dla ludzkiego oka. Dlatego

obraz składowej *Hue* (il. 9.36d) przeskalowano tak, aby je uwidocznili. Barwa pozostaje niezmienna dla wszystkich obszarów z pierwszego i trzeciego rzędu, natomiast zamiana na przestrzeń CMYK spowodowała lekkie przesunięcie barwy w kierunku czerwieni.

Analiza nasycenia (il. 9.36e) wskazuje, że przyciemnianie w przestrzeni RGB nie zmienia tej składowej. Podobnie jest w drugim wierszu, gdzie różnice są niewielkie, ale nasycenie mniejsze z powodu udziału składowych *R* i *B*. Rozjaśnianie w przestrzeni RGB (trzeci wiersz) powoduje spadek nasycenia. Podobnie jest z kolorami, które powstały przez symulację przestrzeni CMYK (czwarty wiersz na il. 9.36e).

Analiza intensywności (il. 9.36f) nie przynosi niespodzianek. Intensywność spada podczas przyciemniania i rośnie podczas rozjaśniania kolorów, niezależnie od tego, jak były generowane.

Przedstawione powyżej omówienie zmian wprowadzanych przez przekształcenie wyjściowego obrazu oraz wartości występujących w składowych HSI powinno pomóc w interpretacji różnych badanych obrazów. W tym przypadku zadanie było o tyle łatwiejsze, że znaleźliśmy dokładnie sposób powstania testowego obrazu.

Przedstawione w tym rozdziale narzędzia, służące do wydobywania informacji z obrazów kolorowych, pokazują, jak wiele możliwości pojawia się wraz z dodaniem koloru do analizowanych obrazów. Dość trudno jest uzyskać takie efekty analizy jak w przypadku zmysłu wzroku, ale dysponujemy różnymi narzędziami i wydobywanie informacji z kolorowych obrazów może być naprawdę ciekawe.

Kończąc ten rozdział il. 9.37 pokazuje, jak wiele tracimy, zamieniając obraz kolorowy na monochromatyczny. Na il. 9.37a możemy ocenić pewne elementy związane z techniką jazdy na nartach, ale na temat kolorystyki stroju nie wiemy niczego.



a



b

Il. 9.37. Obraz w skali szarości uniemożliwia jakąkolwiek ocenę kolorystyki stroju narciarza

Nawet nie możemy stwierdzić, że dominują w nim dwa kolory, bo czerwony i niebieski mają na obrazie szarym taki sam poziom szarości.

Możliwości detekcji różnych elementów obrazów kolorowych przedstawiano w tym rozdziale w postaci zarysów obiektów (najczęściej z zakreskowanym wnętrzem), nałożonych na analizowany obraz zamiast obrazów binarnych. Pozwoliło to na lepsze pokazanie zgodności wyników detekcji z naszymi oczekiwaniami. Nie zmienia to jednak faktu, że do wyznaczenia obszarów tych obiektów niezbędna była binaryzacja. Za każdym razem była ona wykonywana w sposób manualny.

Automatyczna binaryzacja w przypadku obrazów kolorowych jest oczywiście możliwa, ale wymaga wcześniejszego przeprowadzenia wielu testów i powinna być używana w odniesieniu do obrazów o zbliżonym charakterze i zakresie kolorów.

10. Zaawansowane metody analizy

Tytuł tego rozdziału może być nieco mylący, gdyż część omawianych do tej pory operacji, jak na przykład detekcja działów wodnych, z całą pewnością zasługuje na nazwanie ich zaawansowanymi. Z drugiej strony narzędzia omawiane w tym rozdziale odbiegają nieco od głównego nurtu tej książki, która opisuje klasyczny ciąg operacji analizy obrazu, obejmujących akwizycję, korekty geometryczne i LUT, filtrowanie, operacje morfologiczne, binaryzację oraz pomiary. Zostawmy zatem dyskusję o terminologii na boku jako mniej istotną i przejdźmy do przedstawienia kilku ciekawych narzędzi. Ponieważ niniejsza książka dotyczy przede wszystkim podstaw analizy obrazu, to nie stanowi przeglądu wszystkich istniejących operacji, lecz przedstawia te, które wydają się być najważniejsze lub, w niektórych przypadkach, po prostu interesujące.

Zacniemy od posiadającej różne odmiany transformaty Hougha, przy czym skupimy się na jednej z nich, a mianowicie na przekształceniu, służącym wykrywaniu elementów liniowych obrazu. Jest to operacja oparta na bardzo sprytnym pomysśle, mająca kilka wartych podkreślenia cech:

- przetwarza cały obraz, a nie jedynie wybrane obiekty,
- nie wymaga, aby występowały ciągłe linie lub krawędzie; wystarczy, aby obiekty leżały wzdłuż jednej linii,
- jest mało wrażliwa na obecność szumu w obrazie.

Omawianą transformatę opracował w 1962 r. Paul Hough i uzyskał na nią w USA patent nr 3.069.654. Jest to o tyle ciekawe, że w Europie nie ma możliwości patentowania algorytmów.

Wyobraźmy sobie, że w prostokątnym układzie współrzędnych x - y (il. 10.1) znajduje się prosta, która przecina osie x oraz y , odpowiednio w punktach o współrzędnych x_0 oraz y_0 . Prosta ta może być opisana równaniem:

$$y = a \cdot x + b \quad (10.1)$$

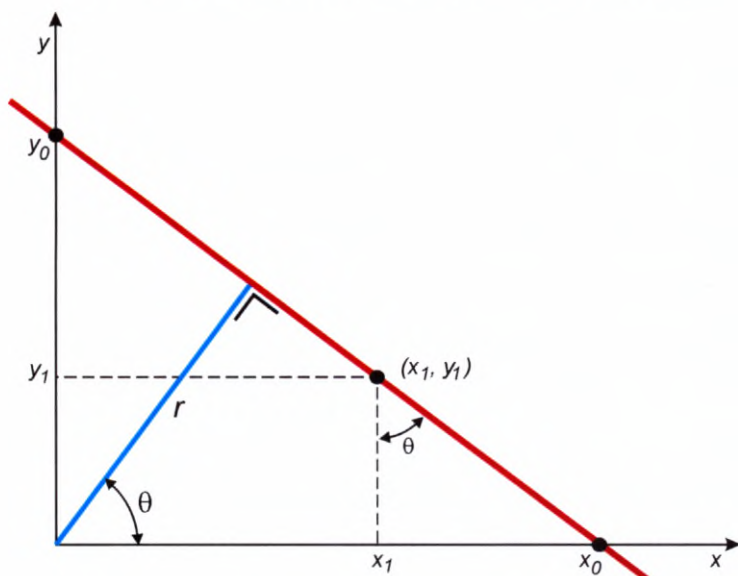
gdzie a i b są stałymi.

Zazwyczaj wartości obu stałych a i b można łatwo wyznaczyć na podstawie współrzędnych x_0 oraz y_0 . W przypadku il. 10.1 wynoszą one:

$$a = -\frac{y_0}{x_0} \quad b = y_0 \quad (10.2)$$

Gdy prosta jest równoległa do jednej z osi lub przechodzi przez początek układu współrzędnych, co powoduje, że $x_0 = y_0 = 0$, to równania (10.2) nie mogą być wykorzystane. Nie będziemy głębiej wchodzić w ten problem. Istotne jest to, że sposób opisu prostej za pomocą równania (10.1) jest powszechnie znany oraz stosowany.

Można jednak opisać prostą również w inny sposób, za pomocą promienia r i kąta θ . Promień jest odcinkiem o długości równej odległości naszej prostej od początku układu współrzędnych. Jeden jego koniec pokrywa się z początkiem układu współrzędnych, a drugi styka się z prostą. Oczywiście promień r jest prostopadły do prostej. Z kolei kąt θ jest kątem pomiędzy osią x i promieniem (il. 10.1).



Il. 10.1. Parametry używane w opisie transformaty Hougha

Promień i kąt jednoznacznie opisują każdą prostą. Załóżmy, że prosta przechodzi przez punkt o współrzędnych x_1 oraz y_1 , zaś promień r tworzy z osią x kąt θ (il. 10.1). Zajmiemy się wyznaczeniem wartości promienia r . Z definicji funkcji cosinus wynika, że:

$$\cos \theta = \frac{r}{x_0}, \quad \text{czyli} \quad r = x_0 \cdot \cos \theta \quad (10.3)$$

Jednocześnie:

$$x_0 = x_1 + (x_0 - x_1) \quad (10.4)$$

Równanie (10.4) opisuje zupełnie oczywistą sytuację, ale zaraz okaże się, że ten zapis jest w pełni uzasadniony. Zwróćmy bowiem uwagę, że:

$$\frac{x_0 - x_1}{y_1} = \operatorname{tg} \theta = \frac{\sin \theta}{\cos \theta} \quad (10.5)$$

Zatem:

$$x_0 - x_1 = y_1 \frac{\sin \theta}{\cos \theta} \quad (10.6)$$

Jeżeli w równaniu (10.3) w miejsce x_0 wstawimy równanie (10.4), to podstawiając w miejsce różnicy $x_0 - x_1$ równanie (10.6), otrzymujemy:

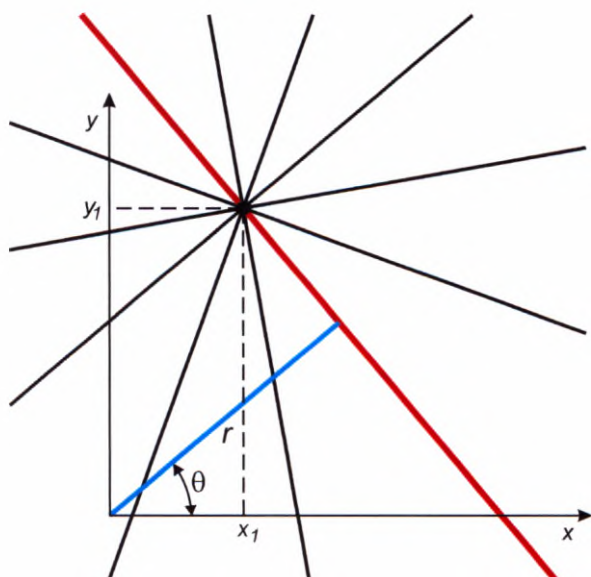
$$r = \cos \theta \cdot \left(x_1 + y_1 \frac{\sin \theta}{\cos \theta} \right) \quad (10.7)$$

W wyniku tej analizy dochodzimy do ostatecznej formy równania, opisującego prostą w układzie r - θ :

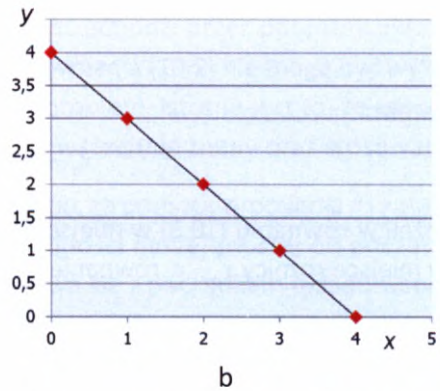
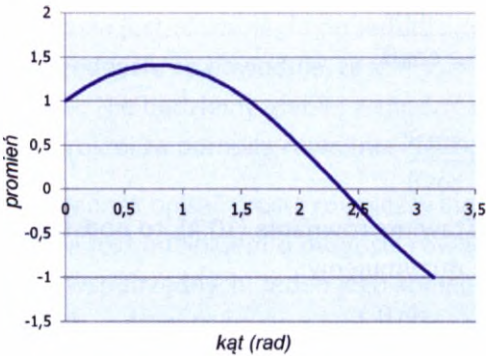
$$r = x \cdot \cos \theta + y \cdot \sin \theta \quad (10.8)$$

Dzięki temu prosta może być reprezentowana w przestrzeni r - θ przez jeden punkt. Z kolei przez jeden punkt przestrzeni x - y może przechodzić nieskończenie wiele różnych prostych. Na il. 10.2 pokazano przykładowe sześć linii, przechodzących przez punkt o współrzędnych (x_1, y_1) . Kąty θ tych linii mogą przyjmować dowolne wartości z zakresu $0-\pi$. Odpowiadające im wartości promienia r można wyznaczyć, wykorzystując równanie (10.8), jako:

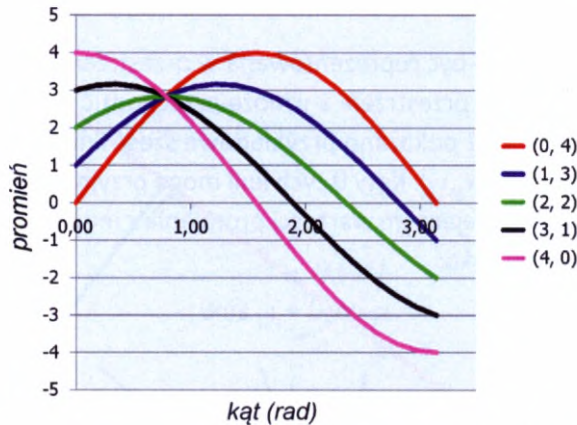
$$r = x_1 \cos \theta + y_1 \sin \theta \quad (10.9)$$



Il. 10.2. Proste przechodzące przez jeden punkt o współrzędnych (x_1, y_1) . Opis w tekście



Il. 10.3. a) reprezentacja wszystkich prostych przechodzących przez punkt o współrzędnych (1, 1), b) pięć punktów leżących na jednej prostej. Opis w tekście



Il. 10.4. Reprezentacja wszystkich prostych przechodzących przez punkty pokazane na il. 10.3b. Kolory krzywych odpowiadają poszczególnym punktom

Transformata Hougha zamienia proste w przestrzeni x - y na punkty w przestrzeni r - θ . Wszystkie proste, przechodzące przez punkt o współrzędnych (x_1, y_1) , można przedstawić w formie graficznej jako krzywą w układzie r - θ , opisaną równaniem (10.9). Na il. 10.3a pokazano taką krzywą dla punktu o współrzędnych (1, 1).

Z kolei il. 10.3b przedstawia prostą w układzie x - y z zaznaczonymi pięcioma punktami. Dla każdego z tych punktów można wyznaczyć krzywą analogiczną do pokazanej na il. 10.3a, ale o nieco innym przebiegu. Z dotychczasowych rozważań wynika, że skoro wszystkie punkty z il. 10.3b leżą na jednej prostej, to krzywe r - θ , reprezentujące wszystkie proste przechodzące przez te punkty, powinny przecinać się w jednym miejscu. Il. 10.4 pokazuje, że tak właśnie się dzieje. Wszystkie krzywe na il. 10.4 przecinają się w punkcie o współrzędnych $0,87 \text{ rad} = 45^\circ$ oraz 2, co odpowiada prostej z il. 10.3b.

W celu przygotowania punktów, na podstawie których przeprowadza się transformatę Hougha, pierwszym krokiem podczas analizy jest zazwyczaj detekcja krawędzi. Jest to uzasadnione faktem, że właśnie krawędzie są tym elementem obrazu, w którym można spodziewać się występowania liniowych elementów.

Pomimo bardzo ciekawego pomysłu transformatę Hougha wykorzystuje się dość rzadko. Głównym powodem jest bardzo długi czas obliczeń oraz problemy z dopasowaniem parametrów filtrujących wstępne wyniki.

Transformatę Hougha można rozszerzyć na inne kształty. Najpopularniejszym z nich jest koło. Warto jednak pamiętać, że do opisanego prostej wystarczą dwa parametry, natomiast opisanie koła wymaga już trzech parametrów, którymi są współrzędne środka ciężkości oraz promień. Dlatego w przypadku detekcji kół za pomocą transformaty Hougha czas analizy wydłuża się jeszcze bardziej.

Na il. 10.5 pokazano zbiór odcinków linii prostej, wykrytych za pomocą transformaty Hougha, nałożony na wyjściowy obraz budynku. Liczne elementy liniowe zostały wykryte w koronach drzew oraz krzewów, co dla ludzkiego zmysłu wzroku jest zaskakujące. Można wpływać na wynik detekcji elementów liniowych przez wybór odpowiednich wartości krawędzi lub liczby punktów, które powinny być uwzględnione podczas wyznaczania elementu liniowego. Przefiltrowane w ten sposób elementy liniowe pokazano na il. 10.6. Usunięte zostały elementy liniowe z koron drzew i krzewów, ale jednocześnie zniknęły też niektóre odcinki widoczne na il. 10.5 na granicy dachu i nieba. Stało się to w miejscach o niższym kontraście. Transformata Hougha okazuje się zatem narzędziem bardzo interesującym, ale jednocześnie dość trudnym do sterowania. Dodatkowo długi czas obliczeń nie zachęca do licznych eksperymentów ze zmianą parametrów detekcji.

Kolejna zaawansowana metoda, którą jest transformata Fouriera, w pełni zasługuje na to określenie ze względu na skomplikowany aparat matematyczny. Ponieważ niniejsza książka skupia się na podstawach analizy obrazu, ograniczymy się do krótkiego omówienia idei tego przekształcenia, bez wnikania w szczegóły jego implementacji.

Każdy sygnał (funkcję) y można przedstawić za pomocą liniowej kombinacji ze współczynnikami a_i różnych funkcji f_i , które nazywamy funkcjami bazowymi:

$$y(x) = a_1 f_1(x) + a_2 f_2(x) + \dots + a_n f_n(x) + \dots \quad (10.10)$$

Suma, opisana równaniem (10.10), może mieć w zależności od funkcji skończoną lub nieskończoną liczbę składników. Szczególnym przypadkiem jest suma, w skład której wchodzi wyłącznie funkcje sinus i cosinus o wzrastającej częstotliwości. Mówimy



Il. 10.5. Obraz budynku z naniesionymi odcinkami prostych, wyznaczonymi za pomocą transformaty Hougha. Opis w tekście



Il. 10.6. Przechwytowane elementy liniowe z il. 10.5. Opis w tekście

w takim przypadku, że przechodzimy z dziedziny czasu lub odległości (jeżeli zmienna x opisuje czas lub odległość) na dziedzinę częstotliwości. Opisywane przekształcenie jest odwracalne, czyli możemy na podstawie funkcji bazowych odtworzyć pierwotną funkcję. Najbardziej popularną transformatą częstotliwościową jest transformata Fouriera, którą opisuje równanie (10.11):

$$y'(x) = A_0 + \sum_{k=1}^{\infty} A_k \cos\left(\frac{2\pi k}{\lambda} x\right) + B_k \sin\left(\frac{2\pi k}{\lambda} x\right) \quad (10.11)$$

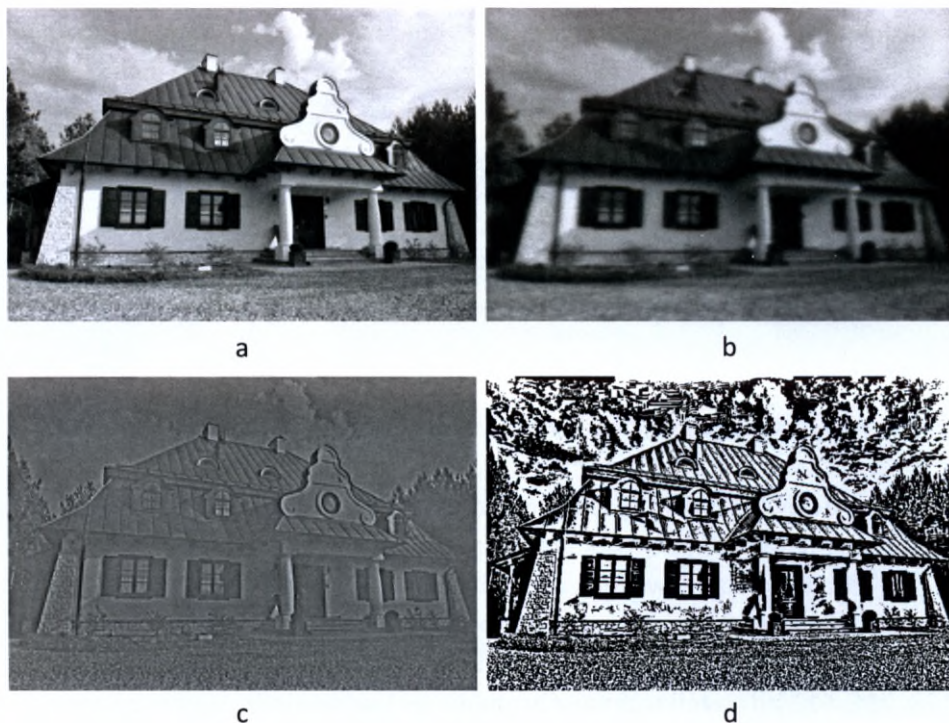
gdzie: λ – okres funkcji y , natomiast A i B są stałymi w dziedzinie liczb zespolonych.

Nie dyskutujemy teraz, w jaki sposób wyznacza się współczynniki równania (10.11), gdyż jest to złożony proces obliczeniowy. Transformatę Fouriera można rozszerzyć na zbiory dwuwymiarowe i zastosować do obrazów. Ze względu na dużą złożoność obliczeniową transformaty Fouriera, która dla N elementów wynosi N^2 , zwykle używa się zmodyfikowanej wersji, tzw. szybkiej transformaty Fouriera (FFT, ang. *Fast Fourier Transformation*), dla której złożoność obliczeniowa wynosi $N \cdot \ln N$. Jest jednak jedno silne ograniczenie – rozmiar obrazu musi być potęgą liczby 2, czyli np. 64, 128, 256, 512 itd. Jeżeli rozmiar obrazu nie jest odpowiedni, to oprogramowanie uzupełnia obraz do rozmiaru będącego najbliższą potęgą liczby 2 i przypisuje dodatkowym pikselom wartość 0.

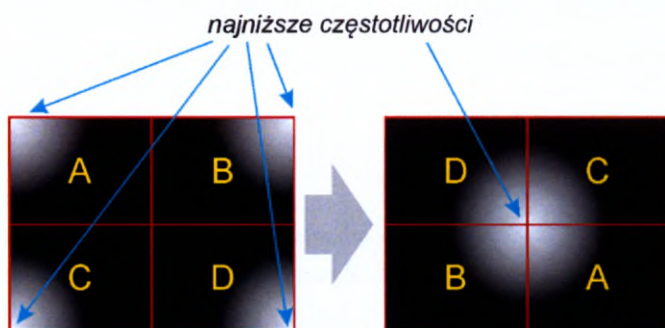
Największą zaletą transformaty Fouriera w analizie obrazu jest możliwość usunięcia lub osłabienia, względnie wzmocnienia, wybranych częstotliwości. Szczególnie skuteczne jest jej zastosowanie do modyfikacji elementów okresowych. Dodatkowo, wykorzystując widmo Fouriera, możemy filtrować obiekty zorientowane pod różnymi kątami. Niestety, w praktyce transformata Fouriera jest dość trudna do praktycznego zastosowania, zwłaszcza jeżeli chcielibyśmy jej działanie zautomatyzować, czyli wykorzystać ją do dużej liczby różnych obrazów.

Il. 10.7 pokazuje, że proste usunięcie wybranych częstotliwości daje wyniki podobne, jak omówione w rozdziale o filtrach filtry nazwane dolnoprzepustowymi i górnoprzepustowymi. Dodanie obrazów 10.7b i 10.7c da obraz wyjściowy 10.7a.

Zazwyczaj najwięcej interesujących nas informacji zawierają elementy obrazu o niskich częstotliwościach i te częstotliwości najczęściej ulegają przetworzeniu. Na obrazie widma Fouriera odpowiadają im miejsca znajdujące się w narożnikach obrazu (lewa strona il. 10.8). Dlatego bez utraty informacji obraz widma Fouriera przetwarza się tak, aby najniższe częstotliwości odpowiadały centralnej części obrazu (prawa strona il. 10.8). Jeżeli interesuje nas ta część obrazu, która zawiera wysokie



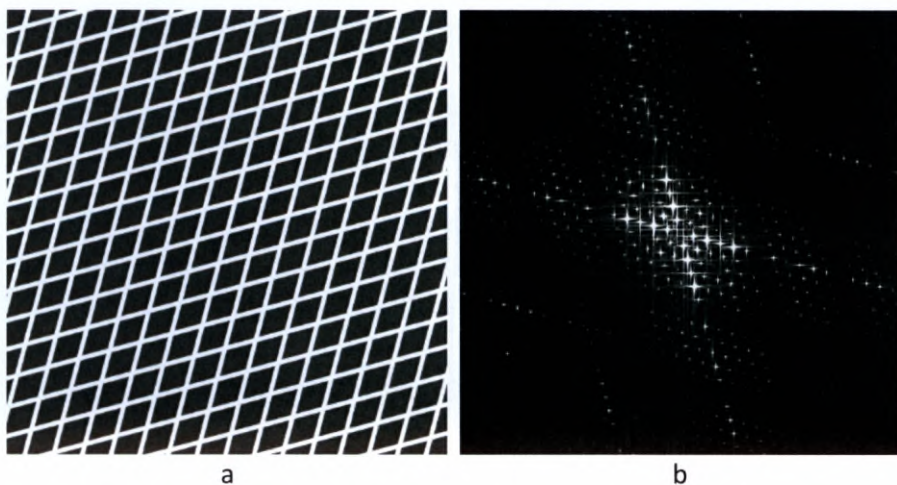
Il. 10.7. Wykorzystanie transformaty Fouriera: a) obraz początkowy, b) usunięcie wysokich częstotliwości, c) usunięcie niskich częstotliwości, d) obraz (c) po binaryzacji; białe miejsca odpowiadają wyższym amplitudom



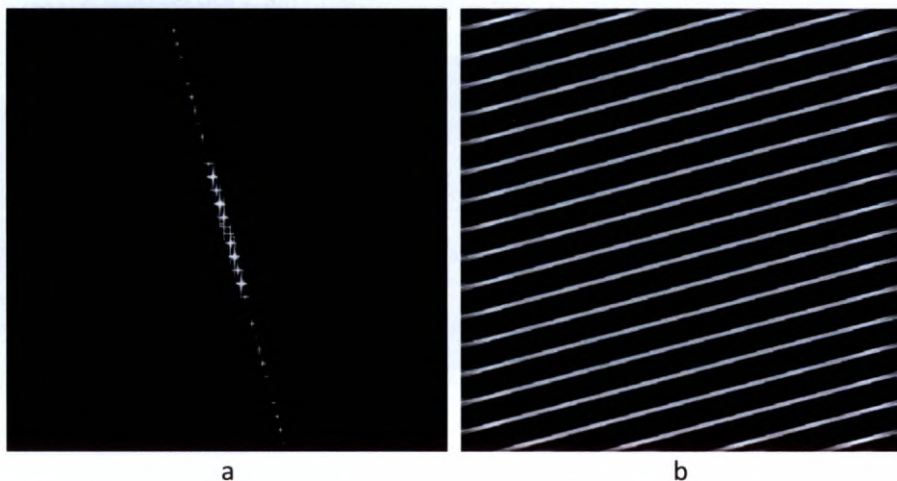
Il. 10.8. Przetworzenie obrazu widma Fouriera w celu lepszego uwidocznienia niskich częstotliwości

częstotliwości, to wystarczy od obrazu początkowego odjąć obraz składający się z niskich częstotliwości lub odpowiednio zmodyfikować filtr korygujący widmo Fouriera.

Wynik FFT obrazu siatki (il. 10.9a) wykazuje okresowość pików widocznych w postaci jasnych punktów na il. 10.9b. Dodatkowo można zauważyć, że punkty układają się



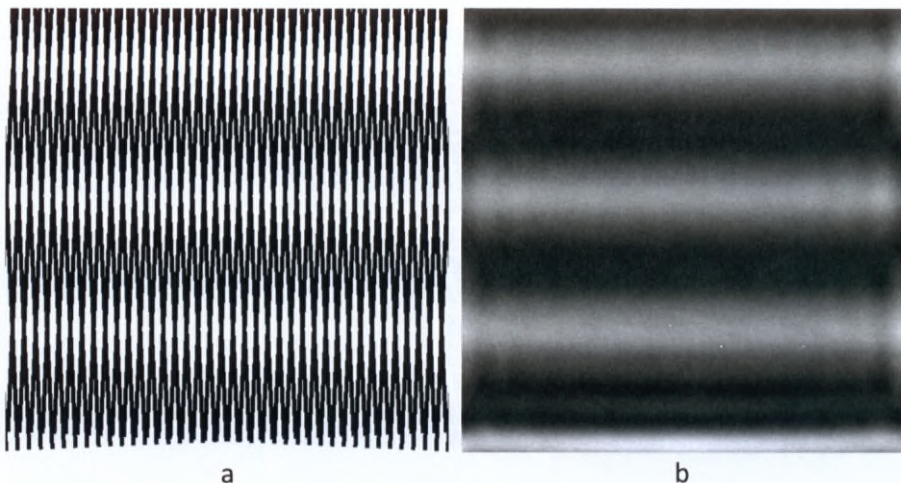
Il. 10.9. Transformata Fouriera: a) obraz początkowy, b) jego transformata



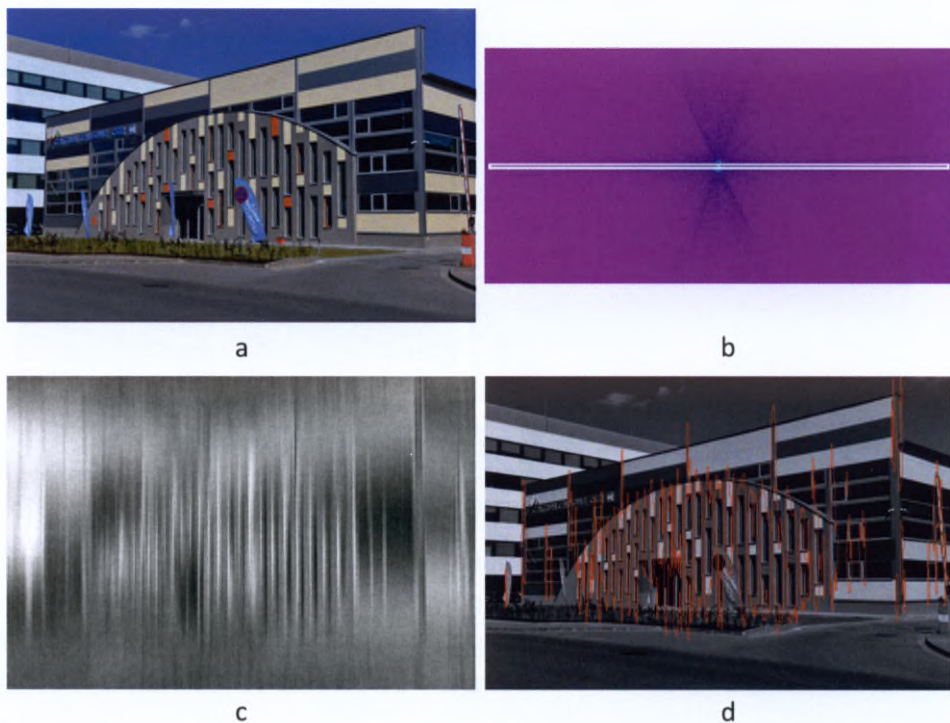
Il. 10.10. Transformata Fouriera obrazu z il. 10.9a: a) ograniczenie zakresu transformaty i b) jej odwrotna transformata (ang. *inverse FFT*)

wzdłuż linii prostopadłych do przetwarzanej siatki. Można to wykorzystać do eliminacji jednej z rodzin linii (il. 10.10a oraz b). Z powodu okresowości silne piki obserwujemy również daleko od centrum obrazu transformaty. Wynika to z faktu okresowości obrazu siatki oraz względnie niewielkich odległości pomiędzy jej poszczególnymi liniami.

Interesujący wynik daje pozostawienie jedynie niskich częstotliwości i odwrotna transformata Fouriera w odniesieniu do rysunku ilustrującego efekt mory (il. 10.11a). Widoczne na il. 10.11b szerokie jasne i ciemne pasy wskazują, że nasz zmysł wzroku nie do końca nas oszukiwał.



II. 10.11. a) ilustracja efektu mory, b) efekt zachowania niskich częstotliwości w widmie Fouriera



II. 10.12. Wykorzystanie transformaty Fouriera do detekcji zorientowanych obiektów:
 a) zdjęcie hali tenisowej, b) widmo Fouriera obrazu (a) zamienionego na szary,
 c) odwrotna transformata wąskiego poziomego paska zaznaczonego na obrazie (b),
 d) elementy pionowe wykryte na podstawie obrazu (c)

Przekształcenie FFT można wykorzystać do wykrywania elementów zorientowanych w pewnym kierunku również na obrazach o mniej wyraźnej obecności takich elementów niż na il. 10.9a. Il. 10.12a przedstawia halę tenisową, a il. 10.12b widmo jej monochromatycznego obrazu. Obraz widma został pokazany jako pseudokolorowy w celu lepszej wizualizacji. Biały prostokąt pokazuje część widma, którą poddano odwrotnej transformacji (il. 10.12c). Na podstawie tego obrazu można wykryć elementy zorientowane pionowo. W tym celu można wykorzystać fakt, że il. 10.12c stanowi silnie rozmyty obraz hali tenisowej z zachowanymi elementami pionowych krawędzi. Wykorzystując filtrację Prewitt i następnie binaryzację otrzymujemy obszary zorientowane pionowo, których granice pokazują czerwone linie na il. 10.12d. Interesujące jest to, że transformata Fouriera wykrywa orientację niezależnie od tego, czy widzimy wyraźnie zdefiniowane obiekty, czy też nie.

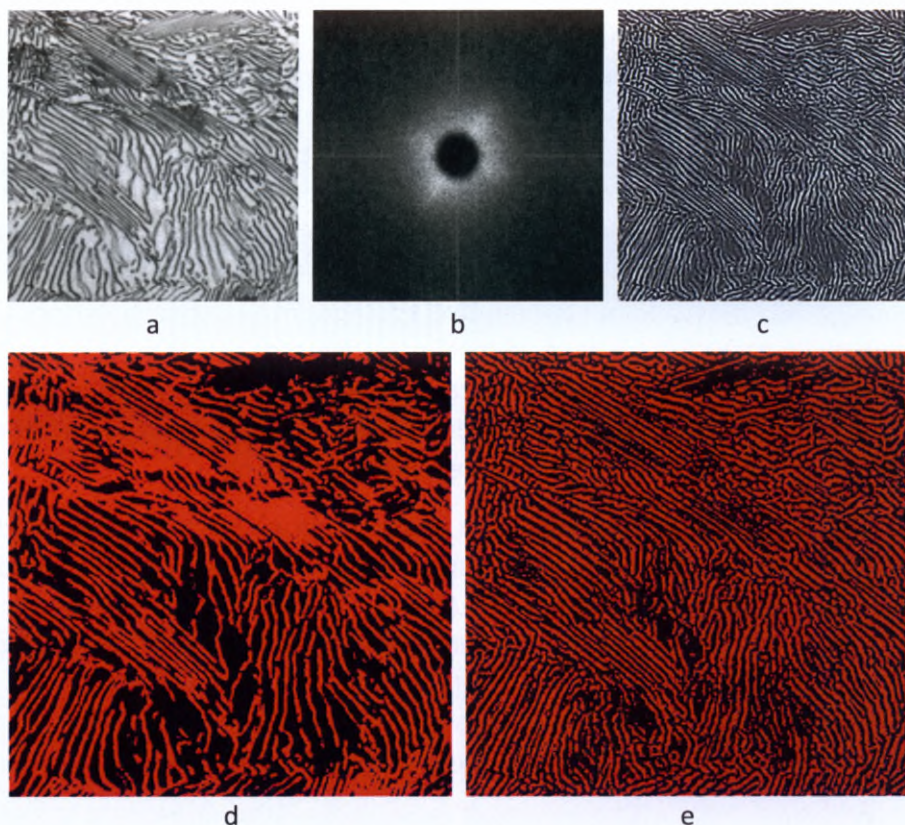
Zdarzają się sytuacje, gdy potrzebne jest usunięcie nie wysokich, jak w przykładach omówionych do tej pory, ale niskich częstotliwości. Próba detekcji płytek perlitu widocznych na il. 10.13a prowadzi do obszarów, w których płytki są sklejone w jeden obszar (lewy górny róg il. 10.13d). Jeżeli wykona się transformatę Fouriera, to po usunięciu z widma niskich częstotliwości (il. 10.13b) uzyskuje się obraz, który sprawia wrażenie bardzo słabo skonstrastowanego (il. 10.13c). Kontrast rzeczywiście jest słaby i na potrzeby wizualizacji il. 10.13c przedstawia obraz po dodatkowym wyrównaniu histogramu. Pozwala on jednak na znacznie lepszą detekcję płytek perlitu (il. 10.13e).

Podsumowując krótkie spotkanie z transformatą Fouriera, można stwierdzić, że jest ona silnym narzędziem w analizie obrazu, ale jej skuteczne wykorzystanie wymaga często wielu prób i zwykle stanowi poważne wyzwanie dla projektanta algorytmu.

Zajmiemy się teraz kolejnym zaawansowanym narzędziem, jakim jest korelacja dwóch obrazów. Schemat korelacji pokazuje il. 10.14.

Korelację wyznacza się dla dwóch obrazów: pierwszym jest obraz, który można nazywać analizowanym, a drugim jest obraz nazywany wzorcem. Na ogół wzorzec jest względnie mały w stosunku do analizowanego obrazu. Sposób wyznaczania wartości korelacji jest bardzo podobny do filtrowania za pomocą filtrów liniowych. Obraz wzorca przykładą się do analizowanego obrazu i nową wartość wyznacza się dla piksela, którego położenie odpowiada centralnemu pikselowi wzorca (il. 10.14). Omówimy dwa warianty korelacji, a mianowicie normalizowaną konwolucję oraz różnicę parami.

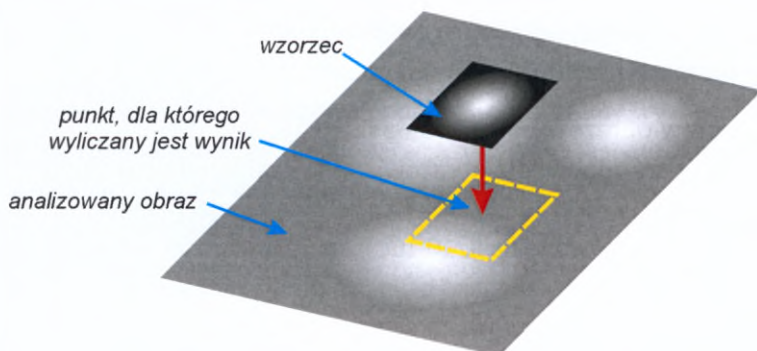
Konwolucja obrazu i wzorca jest bardzo podobna do filtra liniowego, przy czym rolę współczynników filtra pełnią wartości pikseli wzorca. W przypadku normalizowanej konwolucji bierze się pod uwagę w obrazie i we wzorcu różnicę pomiędzy wartością



Il. 10.13. Detekcja płytek w strukturze perlitu w stali węglowej: a) obraz początkowy, b) widmo Fouriera z usuniętymi niskimi częstotliwościami, c) odwrotna transformata Fouriera, d) wynik detekcji na obrazie początkowym, e) wynik detekcji na obrazie po odwrotnej transformacie Fouriera

danego piksela i wartością średnią w obrębie wzorca lub pokrytego przez wzorec obszaru analizowanego obrazu (obszar ograniczony żółtą przerywaną linią na il. 10.14). W wyniku takiej operacji wartości zarówno wzorca, jak i fragmentu analizowanego obrazu przyjmują wartości zarówno dodatnie, jak i ujemne. Suma tych wartości wynosi oczywiście 0. Jest to bardzo podobne do matryc filtrów do detekcji krawędzi. Konsekwentnie, wynik konwolucji również może być zarówno dodatni, jak i ujemny. Jeżeli wynik podzielimy przez pierwiastki z sum kwadratów odchyłeń od wartości średniej dla wzorca i analizowanego fragmentu obrazu, to otrzymamy wynik znormalizowany. Wartości korelacji będą się zmieniały w takim przypadku w zakresie od -1 do 1.

W przypadku gdy wzorec składa się z pikseli o jednakowej wartości, to korelacja z użyciem normalizowanej konwolucji da zawsze wynik 0. Taki sam wynik da korelacja z dowolnym wzorcem w przypadku miejsc, w których poziomy szarości są stałe.

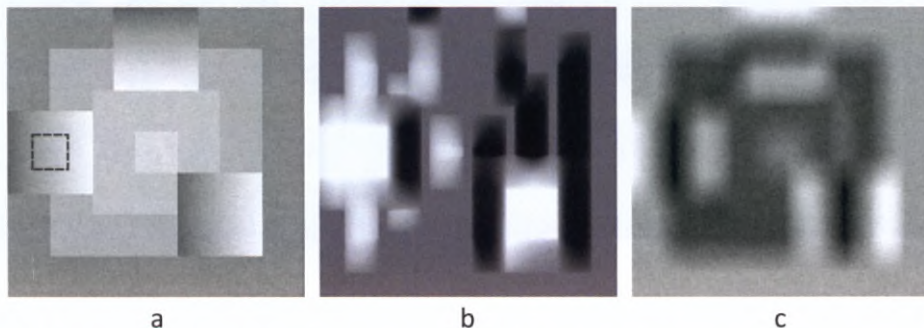


Il. 10.14. Schemat wyznaczania korelacji obrazów

W przypadku różnic parami obliczenia są znacznie prostsze. Sumuje się jedynie bezwzględne różnice wartości odpowiednich pikseli analizowanego obrazu oraz wzorca. Jeżeli wzorzec dokładnie odpowiada obrazowi, to wynik wynosi 0. Wyższe wartości mogą oznaczać zarówno różnice w konfiguracji pikseli, czyli inny wzór lub obiekt, jak i zmiany jasności lub kontrastu dla identycznego obiektu. Zatem również i ta wersja korelacji nie zawsze jest łatwa do interpretacji. Wyjątek stanowi dokładna zgodność wzorca i obrazu.

Poważnym problemem przy wyznaczaniu korelacji obrazów jest bardzo duża liczba obliczeń, która powoduje, że zajmują one dużo czasu i silnie zależą m.in. od zasobów systemu komputerowego.

Wyniki obliczania korelacji dla obrazu testowego (il. 10.15a) i wzorca stanowiącego jego fragment pokazują, że w przypadku metody konwolucji ważna jest raczej tendencja zmian niż rzeczywista zgodność (il. 10.15b). Jest to pełna analogia z działaniem filtrów do detekcji krawędzi. W przypadku metody różnic łatwo jest wyznaczyć miejsca dokładnej zgodności obrazu ze wzorcem (czarne elementy na il. 10.15c), natomiast pozostałe miejsca są zdecydowanie trudniejsze do interpretacji. Przede

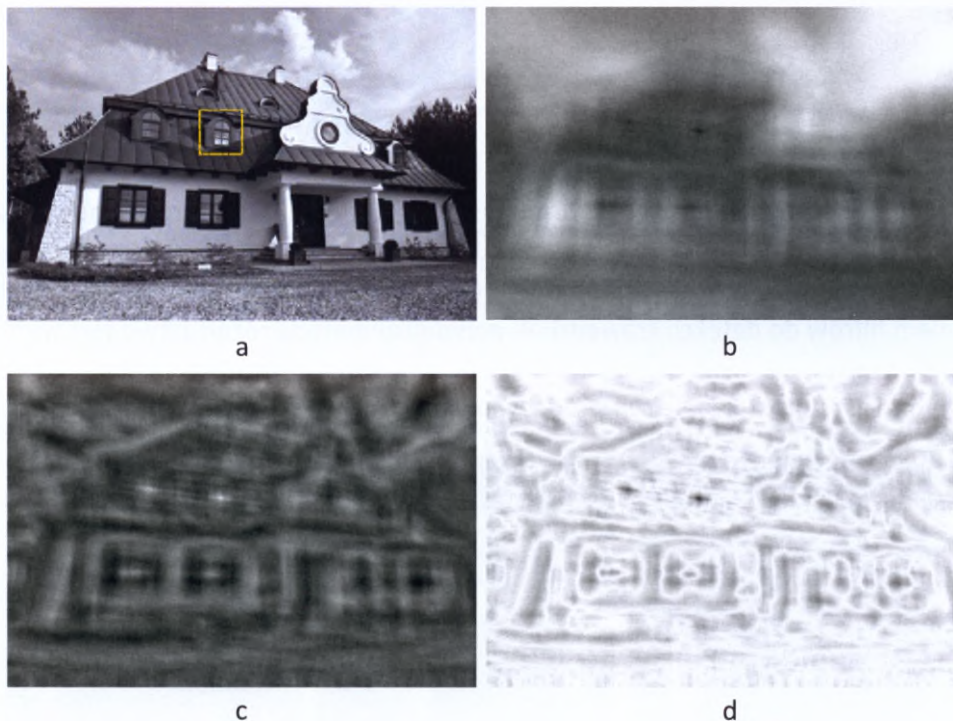


Il. 10.15. Korelacja obrazu i wzorca: a) obraz testowy z zaznaczonym fragmentem pełniącym rolę wzorca, b) korelacja metodą normalizowanej konwolucji, c) korelacja metodą różnic parami

wszystkim nie wiadomo, czy różnica wynika z lokalnie innego wzorca, czy po prostu z różnic w poziomach szarości.

Podobny zestaw obrazów pokazano na il. 10.16. Najłatwiej jest znaleźć miejsca, w których wzorzec odpowiada obrazowi. Są to czarne miejsca na il. 10.16b i d, jak również białe punkty na il. 10.16c. Miejsca zerowe korelacji metodą normalizowanej konwolucji są odwzorowane na il. 10.16c jako 50% szarości. Łatwiej jest je dostrzec na il. 10.16d, gdzie są pokazane jako białe. Na tym rysunku widać, że miejsca zerowe korelacji, czyli najgorsze z możliwych dopasowanie wzorca, odpowiadają w przybliżeniu krawędziom na il. 10.16a.

Il. 10.17a przedstawia fragment ściany budynku z oknami oraz zaznaczony żółtą przerywaną linią zarys wzorca. Wyniki korelacji wyznaczonej metodami normalizowanej konwolucji i różnic parami pokazano na il. 10.17b i c. Z kolei na il. 10.17e oraz f widać wyniki korelacji dla wzorca będącego negatywem wzorca użytego poprzednio. W przypadku obrazów 10.17b oraz e jeden jest po prostu negatywem drugiego.

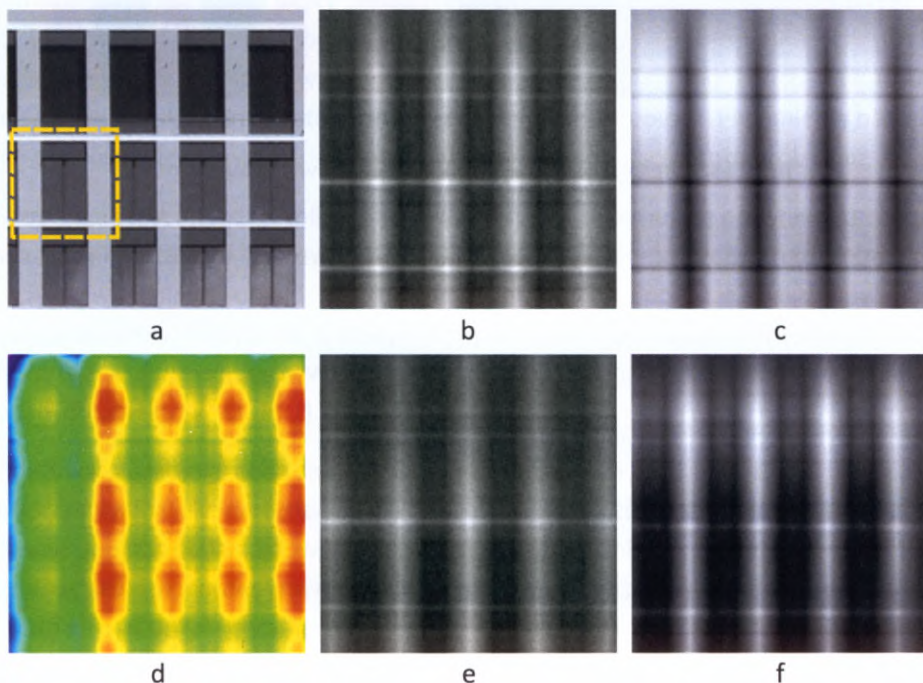


Il. 10.16. Korelacja obrazu i wzorca: a) obraz testowy z zaznaczonym fragmentem pełniącym rolę wzorca, b) korelacja metodą różnic parami, c) korelacja metodą normalizowanej konwolucji, d) negatyw wartości bezwzględnej obrazu (c)

Korelacja metodą normalizowanej konwolucji z negatywem wzorca nazywana jest również korelacją wzajemną (ang. *crosscorrelation*). Jeżeli w tej odmianie korelacji jako wzorzec wykorzystana się analizowany obraz, to będziemy wyznaczali korelację wzajemną obrazu z samym sobą. Taka operacja nazywana jest autokorelacją. Wykorzystuje się ją przede wszystkim do oceny ewentualnego występowania okresowości w obrazie. Przykład takiej autokorelacji pokazuje il. 10.17d. Został on pokazany w wersji pseudokolorowej w celu lepszej czytelności obrazu. Autokorelacja potwierdza okresowość obrazu ściany z oknami. Warto przy tym zwrócić uwagę na podobieństwo obrazów 10.17 d i e.

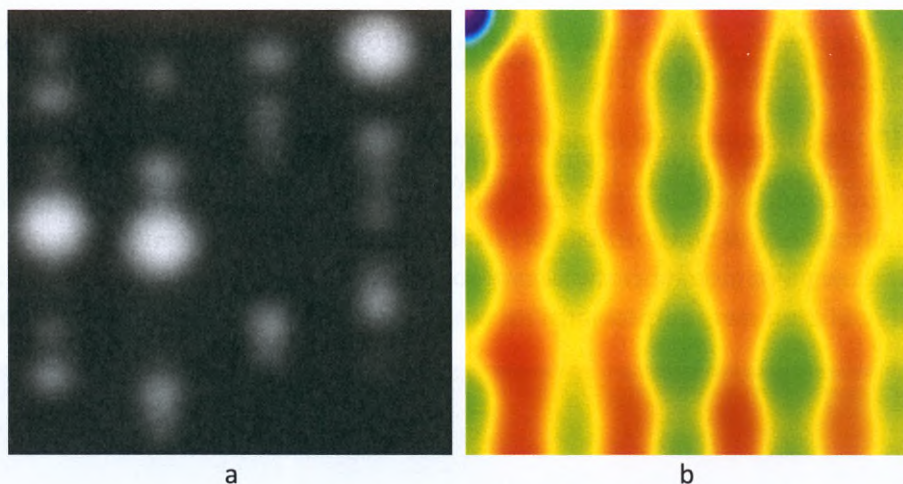
Il. 10.18a pokazuje przykład mniej oczywistej okresowości obiektów w obrazie, ale wynik autokorelacji (il. 10.18b) zdecydowanie potwierdza jej istnienie.

Przedstawione w tym rozdziale przykłady stanowią jedynie niewielki wycinek zaawansowanych metod przetwarzania obrazu. Można wykorzystywać rozmaite metody obliczeniowe, co prowadzi do praktycznie nieskończonej liczby możliwych



Il. 10.17. Korelacje obrazów: a) obraz wyjściowy z zaznaczonym obszarem wzorca, b) korelacja metodą normalizowanej konwolucji, c) korelacja metodą różnic parami, d) autokorelacja, e) korelacja metodą normalizowanej konwolucji z negatywem wzorca, f) korelacja metodą różnic parami z negatywem wzorca

algorytmów. Bardzo duże pole do popisu daje wykorzystanie metod logiki rozmytej oraz sieci neuronowych, przy czym ich zastosowanie ogranicza się przede wszystkim do rozpoznawania ustalonych wcześniej wzorców. Omawianie większej liczby zaawansowanych metod spowodowałoby znaczne zwiększenie objętości książki, a jednocześnie stałoby się, przynajmniej w tej części, trudne do zrozumienia dla osób niemających doświadczenia ze złożonymi matematycznie operacjami.



Il. 10.18. Korelacje obrazów: a) obraz początkowy, b) autokorelacja

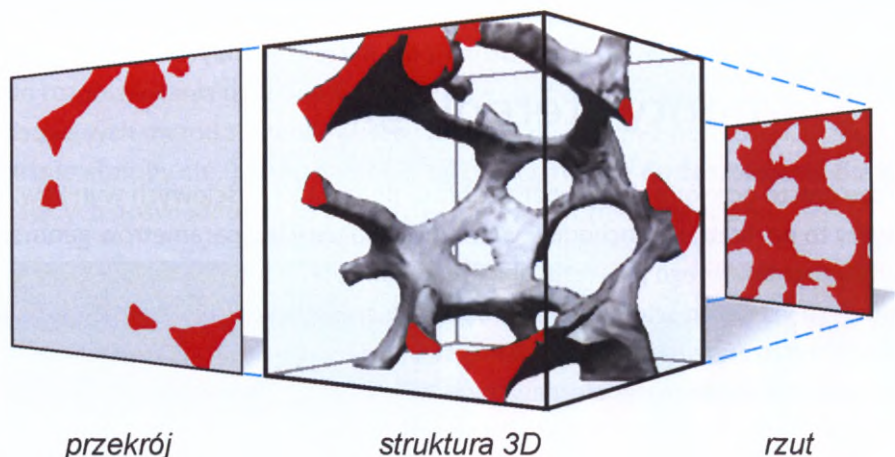
11. Elementy stereologii

Bardzo często oczekuje się od analizy obrazu podania jej ilościowych wyników. Jeżeli dotyczy to np. liczby samochodów widocznych na ulicy lub parametrów geometrycznych sfotografowanego płaskiego obiektu, to taka ocena nie stanowi większego problemu. Pojawia się on wtedy, gdy na podstawie analizy przekroju mamy scharakteryzować przestrzenne parametry badanej struktury. Takie sytuacje zdarzają się często m.in. podczas badań materiałoznawczych, biologicznych i medycznych. Istotę problemu wyjaśnia bardzo prosty przykład: mamy do dyspozycji plaster żółtego sera z dziurami i na podstawie jego analizy powinniśmy ocenić, jaka jest objętość i powierzchnia porów w serze. Rozwiązanie zadania nie jest oczywiste, a stosowne narzędzia oferuje stereologia. Jej podstawy w zakresie przydatnym w analizie obrazu opisuje ten rozdział.

Termin stereologia został użyty po raz pierwszy w 1961 r. w wyniku poszukiwania odpowiedniego słowa, które mogłoby opisać metody badania trójwymiarowych cech badanych obiektów na podstawie ich płaskich przekrojów. Pomimo upływu blisko 60 lat wciąż trwają dyskusje nad ostateczną wersją definicji stereologii. Przyjmiemy następującą definicję, która wydaje się być możliwie pełna i oddająca istotę problemu:

„Stereologia jest interdyscyplinarną dziedziną nauki, obejmującą podstawy teoretyczne, zasady i procedury ilościowego charakteryzowania geometrycznych cech rzeczywistych lub wirtualnych obiektów, które mogą być wykryte podczas analizy dowolnego rodzaju obrazów. Najstarszym, tradycyjnym zadaniem stereologii jest trójwymiarowa interpretacja dwuwymiarowych, płaskich obrazów, będących przekrojami lub rzutami badanych obiektów” (il. 11.1).

W definicji stereologii pada słowo „obraz”, zatem analiza obrazu może być i jest często wykorzystywana jako narzędzie oceny stereologicznej. Metody stereologii mogą być stosowane podczas badania dowolnych przestrzennych obiektów, np. cząstek różnych minerałów, włókien, porów, krawędzi, powierzchni pęknięć itp. Należy jednak podkreślić, że w odróżnieniu od tomografii, która pozwala na rekonstrukcję trójwymiarowej struktury badanych obiektów, celem stereologii nie jest rekonstrukcja, lecz uzyskanie obiektywnej, ilościowej i możliwie prostej oceny analizowanej struktury. Otrzymywane parametry powinny pozwolić na klasyfikację i porównywanie struktur. Niekiedy, np. w zastosowaniach medycznych, analiza stereologiczna bywa używana równolegle z tomografią.



przekrój

struktura 3D

rzut

Il. 11.1. Do badania geometrycznych cech trójwymiarowych struktur możemy wykorzystać dwa rodzaje obrazów: rzuty oraz przekroje

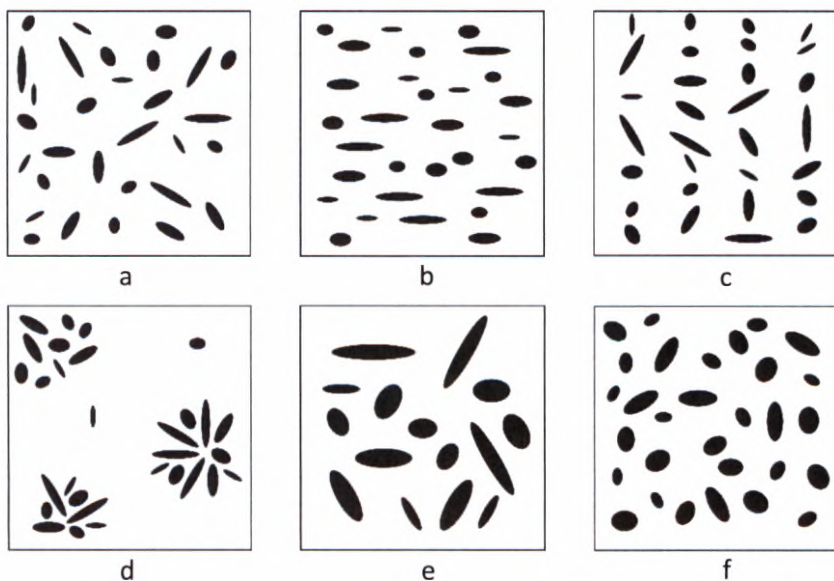
Badanie dowolnej przestrzennej struktury prowadzi do różnych zbiorów danych, które można zaliczyć do trzech kategorii, a mianowicie: rzutów, przekrojów oraz rekonstrukcji 3D (por. il. 11.1):

- Rzuty są najbardziej rozpowszechnioną i dostępną formą informacji wizualnej. Zdjęcia wykonywane kamerami lub smartfonami są właśnie rzutami sfotografowanych obiektów. Nasz zmysł wzroku potrafi na podstawie analizy dwóch nieco odmiennych rzutów odtworzyć częściowo charakterystykę 3D obserwowanych obiektów. Przezroczystość niektórych obiektów (np. wody w akwarium) lub badanie rentgenowskie pozwalają na częściowe poznanie przestrzennej budowy badanej struktury, ale nadal mamy do czynienia z rzutami.
- Przekroje mogą być obrazami fizycznych cięć materiału lub wirtualnymi podzbiorami informacji z obrazów tomograficznych. Pojedynczy przekrój zwykle nie zawiera informacji o charakterze 3D.
- Rekonstrukcje 3D zapewniają jako jedyne potencjalnie kompletną informację o przestrzennej strukturze. Niestety wymagają specjalistycznej aparatury, która jest bardzo kosztowna i wprowadza wiele ograniczeń w zakresie rozdzielczości. W dodatku tomografia komputerowa oparta na promieniowaniu rentgenowskim nie rozróżnia obiektów o zbliżonej masie atomowej, a rezonans magnetyczny wymaga dla rozróżnienia struktur różnicy w zawartości atomów wodoru.

Il. 11.1 ilustruje podstawowe ograniczenia metod stereologicznych, za pomocą których nie da się charakteryzować ani przestrzennych połączeń, ani kształtów trójwymiarowych obiektów. Te informacje można efektywnie pozyskać jedynie za pomocą

badzeń tomograficznych. Taki stan rzeczy może sugerować, że stereologia nie jest potrzebna. Z drugiej strony tomografia też ma swoje ograniczenia. Przede wszystkim mocno ograniczona jest rozdzielczość obrazów. Poza tym, o ile tomografia dobrze identyfikuje np. porowatość, o tyle pozostaje bezradna choćby przy badaniu granic ziaren w materiałach polikrystalicznych. Przekrojów i rzutów dostarczają takie narzędzia jak: mikroskopia świetlna, mikroskopia konfokalna, mikroskopia elektronowa transmisyjna oraz skaningowa (TEM i SEM), dyfrakcja elektronów wstecznie rozproszonych (EBSD), tomografia elektronowa FIB-SEM (*Focused Ion Beam*), radiografia, ultrasonografia (USG), mikroskopia sił atomowych (AFM), radar, sonar, tomografia komputerowa (CT) z jej odmianą mikrotomograficzną (μ CT), rezonans magnetyczny (MRI) lub pozytonowa tomografia emisyjna (PET). Ogromne zróżnicowanie wymienionych powyżej narzędzi zapewnia duży obszar zastosowań dla analizy stereologicznej.

Stereologia opiera się na statystycznej analizie przestrzennych własności obiektów i dostarcza statystycznie dokładne estymatory różnych parametrów. Trzeba pamiętać, że podstawowe formuły mogą wymagać pewnych założeń dotyczących badanych struktur. Najważniejsze z nich mówi, że przestrzenne rozmieszczenie obiektów powinno być UIR (*Uniform, Independent and Random*), czyli równomierne, niezależne i przypadkowe. Powyższe założenie gwarantuje, że poszczególne przekroje są statystycznie jednakowe, i pozwala na wyprowadzenie licznych zależności. W rzeczywistości (por. il. 11.2) wiele struktur nie spełnia założeń dotyczących równomiernego,



Il. 11.2. Modelowe struktury ilustrujące zmienność wybranych cech struktur

niezależnego i przypadkowego rozmieszczenia obiektów, ale dzięki odpowiednim metodom również i one mogą być skutecznie analizowane. Wybrane metody takiej analizy podaje niniejszy rozdział.

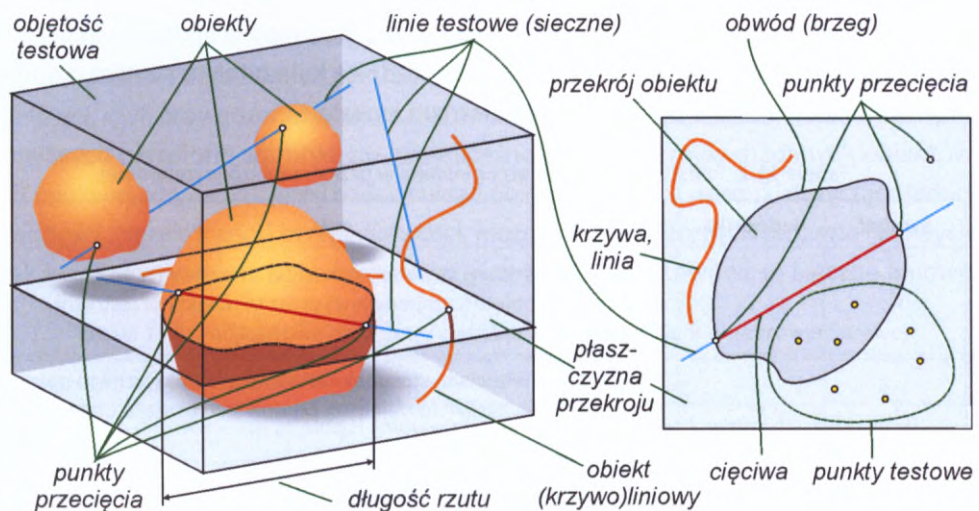
Przyjmuje się, że do pełnego opisu grupy obiektów potrzeba i jednocześnie wystarczy podać ilościową ocenę czterech parametrów: ilości, wielkości, kształtu i sposobu rozmieszczenia (por. il. 11.2).

Ilość obiektów jest najprostsza do oceny i najczęściej oceniana. Miarą ilości obiektów może być ich sumaryczna objętość w przestrzeni lub powierzchnia na przekroju. Należy pamiętać, że liczebność obiektów nie jest miarą ich ilości. Ilość czarnych obiektów we wszystkich sześciu strukturach il. 11.2 jest jednakowa.

Wielkość pojedynczego obiektu, której najbardziej oczywistą miarę stanowi jego objętość, jest trudna lub wręcz niemożliwa do oceny na przekroju struktury. Można oczywiście charakteryzować wielkość przekrojów obiektów. Zwykle obiekty oceniane jako większe na przekroju okazują się również większymi w przestrzeni, ale nie ma jednoznacznej i ścisłej zależności pomiędzy tymi charakterystykami. Na il. 11.2 mamy obiekty o tej samej wielkości, z wyjątkiem struktury na il. 11.2e, w której obiekty są większe. Dzięki zmniejszeniu liczebności ich ilość pozostaje taka sama jak w pozostałych strukturach il. 11.2.

Kształt jest bardzo ważnym, ale jednocześnie bardzo trudnym do oceny parametrem. W przypadku il. 11.2 wszystkie obiekty wykazują ten sam kształt lub jego zróżnicowanie, z wyjątkiem il. 11.2f. Więcej na temat kształtu oraz możliwości jego ilościowej oceny Czytelnik znajdzie w rozdziale poświęconym pomiarom.

Rozmieszczenie jest najtrudniejsze do opisu i nie dysponujemy uniwersalnymi miarami tego parametru. Można przyjąć, że równomierne, niezależne oraz przypadkowe rozmieszczenie wykazują obiekty na il. 11.2a, e oraz f. Obiekty na il. 11.2b są zorientowane w jednym kierunku, na il. 11.2c są rozmieszczone periodycznie (okresowo), a na il. 11.2d są zgrupowane w skupiskach, czyli ich rozmieszczenie jest niejednorodne. Warto przy tym zwrócić uwagę, że każda struktura, badana w odpowiednio dużym powiększeniu, staje się niejednorodna. Przykładem takiej struktury może być naturalnie rosnący las. Obserwowany z wysoko lecącego samolotu wygląda na jednorodną zieloną plamę. Jeżeli jednak chodzimy po tym lesie, to widzimy, że tylko w niektórych miejscach są pnie drzew, a większość powierzchni jest od nich wolna. Jeżeli miarą jednorodności byłaby liczba drzew na jednostkę powierzchni, to przy wyborze 1 m^2 struktura lasu będzie bardzo niejednorodna, ale okaże się prawdopodobnie jednorodna, jeżeli jako miarę odniesienia wybierzemy 1 ha lub 1 km^2 .



Il. 11.3. Podstawowe pojęcia używane w stereologii

Il. 11.3 przedstawia podstawowe pojęcia używane w analizie stereologicznej. Nigdy nie mamy do dyspozycji całej przestrzeni, tylko pewną objętość testową, która na potrzeby wyprowadzania różnych zależności jest najczęściej sześcianem. Obiekty w tej objętości, które mogą być bryłami, powierzchniami lub liniami, są testowane za pomocą siecznych, płaszczyzn przekroju lub położonych na tych płaszczyznach punktów testowych. Jeżeli sieczna przecina bryłę lub figurę, to jej odcinek leżący wewnątrz bryły lub figury nazywamy cięciwą. Podstawowe symbole i wielkości używane w stereologii zawiera tabela 11.1.

Tabela 11.1. Podstawowe wielkości i symbole używane w stereologii

symbol	jednostka	opis
A, a	m^2	Area; pole powierzchni płaskiej (obektu, przekroju, powierzchni testowej itp.)
B, b	m	Border; długość brzegu (obwodu) obiektu lub obiektów
D, d	m	Diameter; średnica koła, kuli, obiektu itp.
L, l	m	Length; długość elementu liniowego; mogą to być sieczne, cięciwy, rzuty itp., ale również elementy krzywoliniowe, jak obwody, granice obiektów itp.
N, n	-	Number; liczba obiektów, siecznych itp.
P, p	-	Point; liczba punktów przecięcia, testowych lub węzłowych
S, s	m^2	Surface; pole powierzchni krzywoliniowej, np. pęknięcia, obiektu itp.
V, v	m^3	Volume; objętość obiektu, grupy obiektów lub testowa

Symbole wielkości używanych w stereologii nie zostały formalnie znormalizowane. Tabela 11.1 zawiera sugerowane, najczęściej używane symbole. Należy jednak pamiętać, że w niektórych obszarach, np. w medycynie, powszechnie używane są

inne oznaczenia dla tych samych wielkości. Dość często można spotkać się z terminem „średnica równoważna obiektu”. Jest to średnica koła o takim samym polu powierzchni jak przekrój obiektu o nieregularnym kształcie.

Tabela 11.2. Podstawowe parametry globalne w przestrzeni trójwymiarowej

symbol	jednostka	opis
V_V	-	Udział objętościowy; stosunek całkowitej sumarycznej objętości obiektów do objętości zajmowanej przez te obiekty przestrzeni testowej
S_V	m^{-1}	Powierzchnia właściwa; stosunek całkowitego sumarycznego pola powierzchni obiektów do objętości zajmowanej przez te obiekty przestrzeni testowej
L_V	m^{-2}	Całkowita długość obiektów liniowych w jednostkowej objętości testowej
N_V	m^{-3}	Liczba obiektów na jednostkę objętości testowej

Na podstawie symboli zaprezentowanych w tabeli 11.1 można zdefiniować 12 parametrów stereologicznych nazywanych globalnymi, gdyż odnoszą się do całej przestrzeni odniesienia, która może być trój-, dwu- lub jednowymiarowa. Symbole takich parametrów są budowane według schematu X_Y , gdzie X oznacza analizowaną wielkość, a indeks Y oznacza przestrzeń odniesienia. Tabela 11.2 zawiera parametry globalne, zdefiniowane dla przestrzeni trójwymiarowej.

Tabela 11.3. Podstawowe parametry globalne w przestrzeni dwuwymiarowej

symbol	jednostka	opis
A_A	-	Udział powierzchniowy; stosunek całkowitej sumarycznej powierzchni przekrojów obiektów do pola zajmowanej przez te obiekty powierzchni
L_A	m^{-1}	Długość właściwa; stosunek całkowitej sumarycznej długości obiektów do pola zajmowanej przez te obiekty powierzchni
N_A	m^{-2}	Liczba obiektów przypadająca na jednostkowe pole powierzchni testowej
P_A	m^{-2}	Liczba punktów przecięcia lub punktów testowych, przypadająca na jednostkowe pole powierzchni testowej zawierającej te punkty

Tabela 11.3. przedstawia parametry globalne zdefiniowane dla przestrzeni dwuwymiarowej, czyli wyznaczone na płaszczyźnie. Za względu na niższy wymiar przestrzeni można było spodziewać się mniejszej o jeden liczby parametrów w porównaniu z przestrzenią trójwymiarową, ale mamy rozróżnienie pomiędzy liczbą obiektów i liczbą punktów, co powoduje powstanie dodatkowego parametru.

Kolejna tabela, o numerze 11.4, zawiera zestawienie parametrów globalnych dla przestrzeni jedno- i zerowymiarowej. Są to parametry wyznaczane na podstawie analizy liniowej, czyli prowadzonej wzdłuż siecznych lub polegające na zliczaniu punktów spełniających określone warunki, np. przynależność do przekrojów badanych obiektów. W ogólnym przypadku, jeżeli badany układ obiektów spełnia warunki dotyczące jednorodności, niezależności i przypadkowości, można używać siecznych zarówno liniowych, jak i nieliniowych, np. kołowych. Najczęściej jednak stosowane są sieczne liniowe.

Tabela 11.4. Podstawowe parametry globalne w przestrzeni jedno- i zerowymiarowej

symbol	jednostka	opis
L_L	-	Udział liniowy; stosunek całkowitej sumarycznej długości cięciw obiektów do całkowitej długości linii testowych (siecznych)
N_L	m^{-1}	Stosunek liczby przeciętych obiektów do całkowitej długości linii testowych (siecznych)
P_L	m^{-1}	Stosunek liczby punktów przecięcia liniowych obiektów (np. obwodów) do całkowitej długości linii testowych (siecznych)
P_p	-	Stosunek liczby punktów trafiających w analizowane obiekty do liczby wszystkich punktów testowych trafiających w badany obszar; parametr ten można też interpretować jako prawdopodobieństwo warunkowe trafienia w obiekt, pod warunkiem że analizowany punkt znajduje się w obszarze testowym

Oprócz parametrów globalnych występują też parametry lokalne, które najczęściej są po prostu średnimi wartościami różnych parametrów. Cieszą się one dużą popularnością, gdyż łatwiej jest sobie wyobrazić, co oznacza parametr lokalny niż parametr globalny. Z drugiej strony, używanie parametrów lokalnych może być mylące, gdyż niekiedy opisują one elementy, które nie istnieją. Problem ten jest często ilustrowany żartobliwym przykładem człowieka idącego na spacer z psem. Średnio mają bowiem po trzy nogi, co oczywiście nie zachodzi w praktyce. To, że zwyczajowo nogi psa nazywamy łapami, nie ma tutaj znaczenia.

Tabela 11.5 zawiera zestawienie najczęściej używanych parametrów lokalnych. Najbardziej miarodajne jest podawanie rozkładów poszczególnych parametrów, np. w postaci histogramów, lub wyznaczanie analitycznych funkcji, opisujących te rozkłady. Metody komputerowej analizy obrazu pozwalają na łatwe ustalanie wartości różnych parametrów dla indywidualnych obiektów, co umożliwia późniejszą obróbkę statystyczną wyników pomiarów. Należy jednak pamiętać, że nadmiar danych zwykle znacząco utrudnia ich analizę. Jeżeli będziemy dysponowali rozkładami kilku wielkości dla kilku grup obiektów, to najprawdopodobniej interpretacja tego zestawu wyników będzie bardzo trudna.

Tabela 11.5. Wybrane parametry lokalne używane w stereologii

symbol	jednostka	opis
$\bar{v}$	m^3	Średnia objętość obiektu
$\bar{a}$	m^2	Średnie pole powierzchni płaskiego obiektu lub jego przekroju
$\bar{s}$	m^2	Średnie pole powierzchni krzywoliniowej
$\bar{l}$	m	Średnia długość liniowego obiektu, jego rzutu lub cięciwy
$\bar{d}$	m	Średnia średnica
λ	m	Średnia droga swobodna; jest to średnia odległość między obiektami, mierzona wzdłuż wybranego kierunku
$\bar{h}$	m	Średnia wysokość lub odległość obiektów

Przejsie od parametrów globalnych do lokalnych jest bardzo łatwe i do wyznaczenia poszczególnych zależności wystarczy znajomość definicji parametrów, które zostały opisane w tabelach powyżej. Zależności pomiędzy parametrami lokalnymi i globalnymi ilustrują zależności od (11.1) do (11.5). Wyznaczenie średniej objętości grupy obiektów wydaje się bardzo proste:

$$\bar{v} = \frac{V_V}{N_V} \quad (11.1)$$

Równanie (11.1) ma niewielkie znaczenie praktyczne, gdyż poza wyjątkowymi sytuacjami nie ma możliwości wyznaczenia wartości względnej liczebności N_V . Z tego samego powodu wyznaczenie średniego pola powierzchni również może być niemożliwe:

$$\bar{s} = \frac{S_V}{N_V} \quad (11.2)$$

Wartość powierzchni względnej S_V potrafimy skutecznie wyznaczyć, co zostanie pokazane w dalszej części rozdziału, ale podobnie jak w przypadku wzoru (11.1), brakuje nam dobrej metody wyznaczania N_V .

W przypadku pozostałych parametrów lokalnych nie będzie już takich poważnych problemów z wyznaczeniem ich wartości.

$$\bar{a} = \frac{A_A}{N_A} \quad (11.3)$$

Wyznaczanie średniego pola powierzchni jest używane bardzo często, a do jego obliczenia wystarczą parametry łatwe do uzyskania na podstawie nawet pojedynczego obrazu. Staje się to szczególnie łatwe w przypadku materiałów jednofazowych,

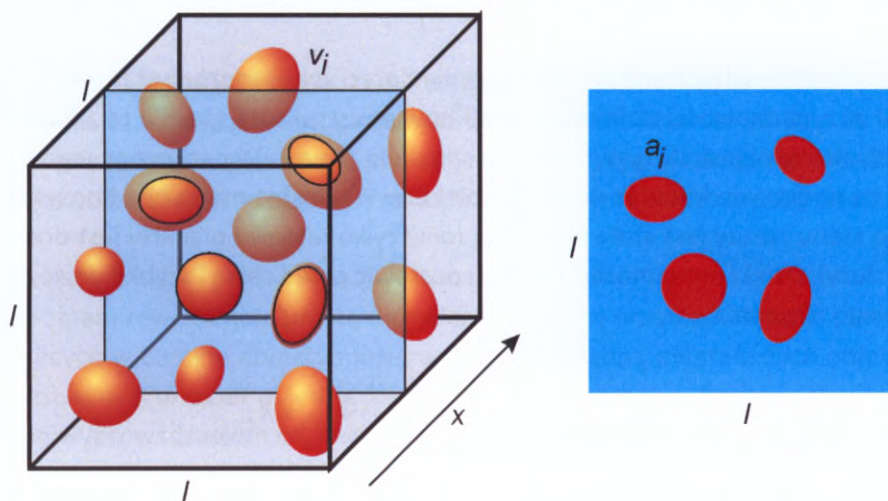
czyli obrazów zawierających tylko jeden rodzaj obiektów, które wypełniają całą powierzchnię i $A_A = 1$. Niewielkie problemy mogą towarzyszyć poprawnemu wyznaczeniu wartości N_A , ale odpowiednie narzędzia do rozwiązania tego zagadnienia są opisane w rozdziale dotyczącym pomiarów.

$$\bar{l} = \frac{L_L}{N_L} \quad (11.4)$$

Wyznaczenie średniej cięciwy również okazuje się bardzo łatwe. Podobnie jest ze średnią drogą swobodną:

$$\lambda = \frac{1 - L_L}{N_L} \quad (11.5)$$

Trzeba jedynie pamiętać, że średnia droga swobodna jest parametrem kierunkowym i należy go wyznaczać dla każdego kierunku osobno. Jest to zatem jeden z parametrów przydatnych do charakteryzowania obiektów, których rozmieszczenie w przestrzeni nie spełnia założeń dotyczących jednorodności, niezależności i przypadkowości.



Il. 11.4. Model do wyprowadzenia zależności do oceny udziału objętościowego. Opis w tekście

Omówimy teraz nieco dokładniej problem wyznaczania udziału objętościowego, co pozwoli nam zbliżyć się do poznania metodyki analiz stereologicznych. Mamy wyznaczyć udział objętościowy obiektów zajmujących testowy sześcian o boku długości l (il. 11.4). W objętości testowej mamy N obiektów (w przypadku il. 11.4 $N = 13$). Każdy z obiektów ma swoją objętość v_i . Ich suma daje całkowitą objętość obiektów v . Udział objętościowy tych obiektów można więc zapisać jako:

$$V_V = \frac{\sum_{i=1}^N v_i}{l^3} = \frac{v}{l^3} \quad (11.6)$$

Zakładamy następnie, że badaną objętość testujemy za pomocą przekrojów równoległych do jednej ze ścian testowego sześcianu, a położenie przekroju opisuje zmienna x . Przekrój (prawa strona il. 11.4) jest kwadratem o boku długości l , który przecina część obiektów. Każdy przekrój przeciętego obiektu ma swoje pole powierzchni a_i . Suma tych pól daje całkowitą powierzchnię przekrojów obiektów, którą oznaczmy a .

Dowolną bryłę możemy zamodelować w postaci zestawu bardzo cienkich przekrojów, podobnie jak robią to drukarki 3D. Zatem objętość dowolnej bryły można obliczyć, całkując zmienne pola jej przekrojów wzdłuż linii prostopadłej do tych przekrojów. Na potrzeby naszego wyprowadzenia funkcję opisującą zmiany sumarycznego pola powierzchni przekrojów obiektów oznaczmy $a(x)$. Możemy teraz zapisać wzór na udział objętościowy w następujący sposób:

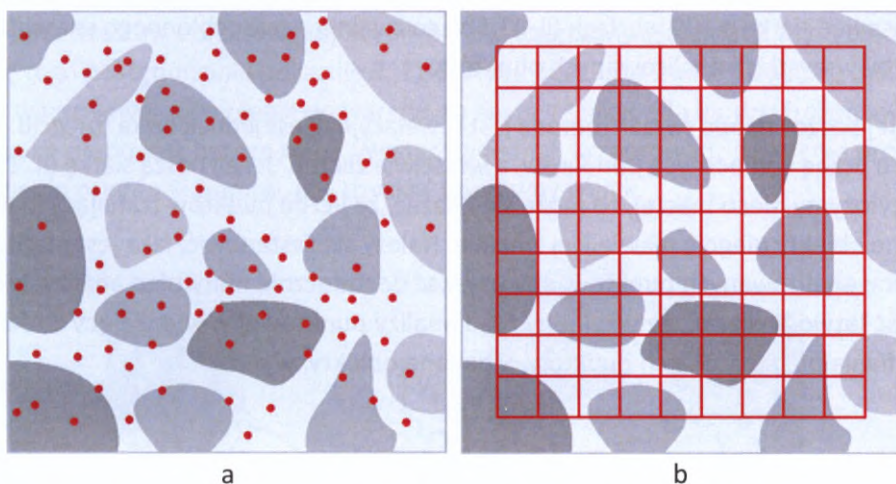
$$V_V = \frac{v}{l^3} = \frac{\int_0^l a(x) dx}{l^3} \quad (11.7)$$

W tym momencie przydadzą się założenia dotyczące jednorodności, niezależności oraz przypadkowości obiektów. Jeżeli bowiem struktura spełnia te założenia, to każdy przekrój jest statystycznie taki sam i nie ma możliwości oceny jego położenia na podstawie badania przecięć obiektów w obrębie przekroju. Konsekwencją tego stanu rzeczy jest stała wartość a (o ile tylko testowy przekrój jest dostatecznie duży). Dzięki temu możemy stałą a postawić przed całką i szybko dojść do końcowego równania:

$$V_V = \frac{v}{l^3} = \frac{\int_0^l a(x) dx}{l^3} = \frac{a \int_0^l dx}{l^3} = \frac{a}{l^2} = A_A \quad (11.8)$$

Jak widać, do oceny udziału objętościowego wystarczy znajomość udziału powierzchniowego. Ponieważ dziury we wspomnianym wcześniej żółtym serze dość dobrze spełniają założenia dotyczące ich przypadkowego rozmieszczenia, to wiemy już, że zajmują taki ułamek objętości jak dziury na plasterku. Zazwyczaj, ze względu na proces dojrzewania, najsmaczniejsze są sery z dużymi dziurami. Na nasze szczęście sery są sprzedawane na wagę, a nie na objętość.

Pozostaje wyznaczyć udział powierzchniowy. Zadanie to nie jest banalne, choć w praktyce niezbyt trudne. W przeszłości rysowano koła lub prostokąty o polu powierzchni w przybliżeniu równym polu powierzchni obiektów. Czasami wycinano



Il. 11.5. Wyznaczanie udziału objętościowego. Opis w tekście

obiekty ze zdjęcia i ważono. Jeżeli wycięto je dokładnie, a papier miał na całej powierzchni taką samą gramaturę, to taki pomiar był nawet dość dokładny, choć bardzo pracochłonny.

Metody stereologiczne przychodzą jednak z pomocą i proponują bardzo proste rozwiązanie (il. 11.5a). Na przekrój badanej struktury rzucamy losowo punkty. Im większa jest powierzchnia zajęta przez przekroje obiektów, tym większe jest prawdopodobieństwo trafienia w te obiekty. Prowadzi to do kolejnej bardzo prostej zależności:

$$A_A = P_P \quad (11.9)$$

Mamy zatem rewelacyjnie prostą metodę. Za pomocą prostego zliczania punktów trafiających w badane obiekty możemy ocenić względną objętość tych obiektów w przestrzeni. A przecież objętość jest trudna do wyznaczenia. Nie będziemy zajmowali się wyprowadzaniem odpowiednich zależności dla metody liniowej; ale wynik jest równie prosty jak w przypadku równań (11.8) i (11.9):

$$V_V = A_A = L_L = P_P \quad (11.10)$$

Największym problemem podczas ręcznej oceny udziału objętościowego (do metod analizy obrazu przejdziemy nieco później) są punkty, które trafiają w granicę między obiektem i jego otoczeniem. W takiej sytuacji należy albo zaliczyć połowę trafienia, albo zaliczać co drugi taki wątpliwy przypadek. Jeżeli nasza struktura jest przypadkowa, to nie musimy rzucać punktów w sposób przypadkowy i nie będzie to miało wpływu na wynik. Można na obraz badanego zbioru przekrojów obiektów po prostu nałożyć siatkę i jako punkty testowe traktować węzły tej siatki. Najwygodniej jest

stosować siatkę o 100 węzłach (il. 11.5b), co zwalnia z czasochłonnego sprawdzania liczby wszystkich analizowanych punktów.

W tej beczce miodu (wszak metoda jest rewelacyjna) jest jednak łyżka dziegciu. Chodzi o liczbę koniecznych pomiarów, a właściwie zliczeń. Jeżeli naszą siatkę (il. 11.5b) przyłożymy nieco inaczej, to może się okazać, że liczba punktów trafiających w badane obiekty ulegnie niewielkiej zmianie. Należy się zastanowić, jaka jest niezbędna liczba analizowanych punktów, aby uzyskać dostatecznie mały błąd analizy. Można dość łatwo wykazać, że względny błąd analizy punktowej γ , wynikający z różnego trafiaania poszczególnych punktów w badane obiekty, wynosi:

$$\gamma = U_{\alpha} \sqrt{\frac{1 - V_v}{n \cdot V_v}} \quad (11.11)$$

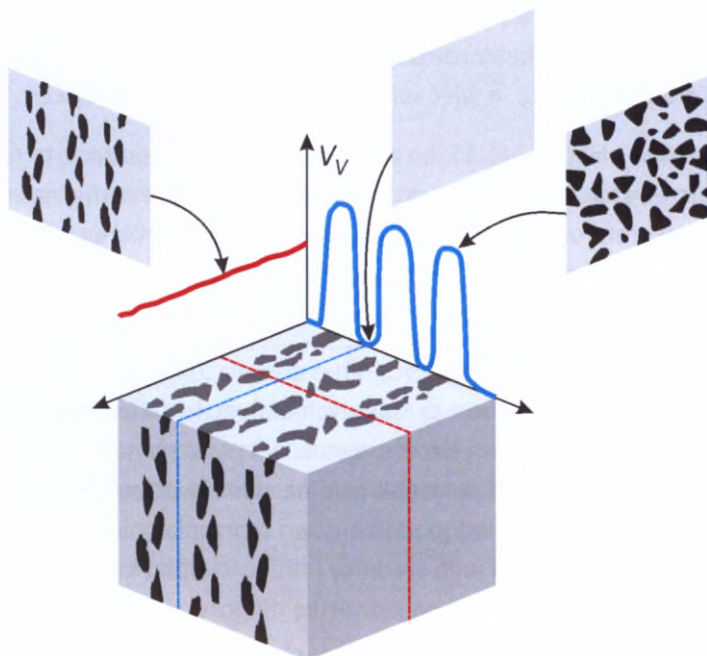
gdzie: α – poziom istotności (ryzyka błędu), wynoszący zwykle 0,05; U_{α} – statystyka rozkładu normalnego, której wartość możemy przyjąć dla szacunkowych obliczeń jako 2; V_v – oceniany udział objętościowy oraz n – liczba rzuconych punktów.

Wnioski z analizy wzoru (11.11) są trochę deprymujące. Jeżeli badany udział objętościowy wynosi ok. 10%, czyli 0,1, to w celu uzyskania względnego błędu analizy nieprzekraczającego 5% potrzeba przebadac ok. 14 350 punktów. Jeżeli będziemy bardzo sprawni i w ciągu sekundy będziemy analizowali cztery punkty, to na jeden pomiar potrzebna będzie godzina. Oczywiście nie jest to akceptowalne rozwiązanie i zwracamy się o pomoc do komputerów. Komputer odpowiednią analizę wykona w ułamku sekundy i tutaj widzimy ogromną korzyść z wykorzystania metod analizy obrazu. Niestety, do analizy komputerowej nie można wykorzystać wzoru (11.11), gdyż komputer wykorzysta wszystkie piksele obrazu i tym samym pojedynczy pomiar wydaje się bezbłędny. I faktycznie, z punktu widzenia wzoru (11.11), który podaje względny błąd metody pomiaru, metoda pomiaru komputerowego jest bezbłędna.

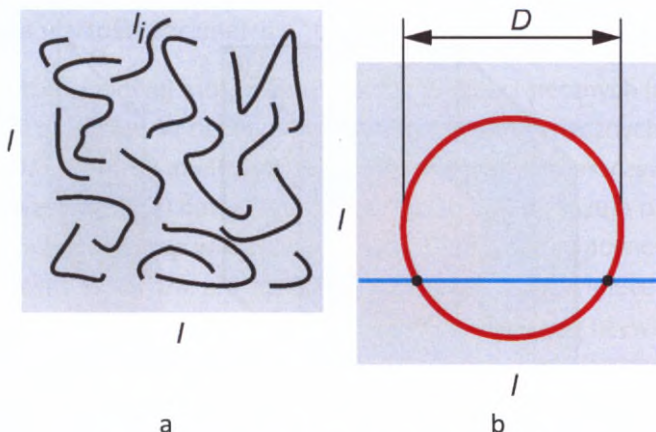
Nie oznacza to jednak pewności, że uzyskany wynik jest poprawny. Wszystko zależy od wyników detekcji obiektów. Błędy detekcji mogą spowodować duże błędy wyników oceny udziału powierzchniowego, choć sama metoda pomiaru jest wolna od błędów. Więcej informacji na temat możliwych błędów oceny powierzchni obiektów oraz sposobów ich minimalizacji znajduje się w rozdziale poświęconym pomiarom. Ewentualne błędy oceny powierzchni obiektów nie mają na szczęście wpływu na bardzo wygodną i prostą zależność (11.10).

Pozostaje pytanie, co zrobić, jeżeli badana struktura nie spełnia założeń metody. Należy w takiej sytuacji przeanalizować, jakie założenia nie są spełnione, i starać się

tak badać zestaw obiektów, aby wynik nie był obarczony systematycznym błędem. Ideę tego postępowania pokazuje il. 11.6. Jeżeli będziemy badali strukturę o budowie warstwowej równoległe do warstw, to nasze wyniki będą bardzo zróżnicowane i ryzyko błędu jest bardzo duże. Wystarczy jednak przeprowadzić badanie prostopadle do orientacji warstw, aby uzyskać poprawny wynik.



Il. 11.6. Unikanie błędów w ocenie udziału objętościowego niejednorodnych struktur.
Opis w tekście



Il. 11.7. Wyznaczanie wartości parametru L_A

Ogólnie możemy badać dowolną strukturę bez jakichkolwiek ograniczeń dotyczących kształtu i rozmieszczenia obiektów, pod warunkiem że nasze badanie będzie spełniało warunki jednorodności, niezależności i przypadkowości. W dalszej części rozdziału zostanie pokazane, że takie przeprowadzenie badań jest trudne, ale jednak możliwe.

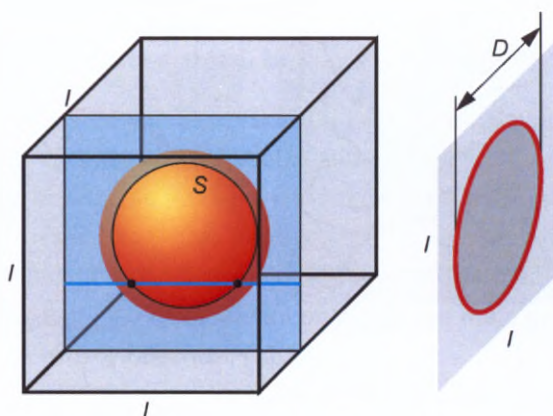
Zajmiemy się teraz wyznaczeniem długości układu przypadkowo zorientowanych linii na płaszczyźnie (il. 11.7a), a konkretnie parametru L_A . Znając L_A oraz N_A , którego wyznaczenie nie jest trudne (patrz rozdział poświęcony pomiarom), możemy łatwo określić średnią długość obiektu, a więc scharakteryzować jego wielkość.

Jeżeli analizowany układ linii (il. 11.7a) nie wykazuje żadnej orientacji, to możemy zastąpić go na potrzeby naszej analizy po prostu okręgiem, który jest idealną krzywą, niewykazującą żadnej preferowanej orientacji (11.7b). Mamy zatem okrąg o średnicy D , leżący na płaszczyźnie kwadratu o boku l . Przy tak prostym modelu od razu znamy wynik, gdyż:

$$L_A = \frac{\pi D}{l^2} \quad (11.12)$$

Naszą strukturę badamy za pomocą siecznych o długości l , równoległych do jednego z boków kwadratu. Przykładowa sieczna została narysowana na il. 11.7b jako niebieska pozioma linia. W zależności od położenia sieczna albo nie przecina okręgu, albo przecina go w dwóch miejscach. Prawdopodobieństwo zdarzenia polegającego na przecięciu okręgu jest ilorazem długości jego średnicy i długości boku kwadratu. Można zatem zapisać, że średnia liczba przecięć na jednostkę długości siecznej wynosi:

$$P_L = \frac{2}{l} \cdot \frac{D}{l} = \frac{2D}{l^2} \quad (11.13)$$



Il. 11.8. Wyznaczanie powierzchni właściwej S_v

W obu równaniach, czyli (11.12) i (11.13), występuje ten sam czynnik, a mianowicie średnica D , podzielona przez kwadrat długości boku kwadratu. Dzięki temu możemy łatwo połączyć te dwa równania i otrzymać kolejną zależność stereologiczną:

$$L_A = \frac{\pi}{2} \cdot P_L \quad (11.14)$$

W podobny sposób możemy wyprowadzić wzór na wyznaczanie powierzchni właściwej S_V , przypadkowo zorientowanego przestrzennie układu krzywoliniowych powierzchni (il. 11.8).

Podobnie jak w przypadku układu linii na płaszczyźnie, modelem układu powierzchni niewykazującego orientacji jest powierzchnia kuli o średnicy D , znajdująca się w sześcianie o boku l . Tym samym:

$$S_V = \frac{\pi D^2}{l^3} \quad (11.15)$$

Analogicznie jak poprzednio, sieczna o długości l , równoległa do jednej z krawędzi sześcianu, albo nie trafia w kulę, albo przecina ją dwa razy. Ponieważ prawdopodobieństwo trafienia w kulę jest równe stosunkowi pola powierzchni jej rzutu (czyli koła wielkiego) do pola powierzchni ściany kwadratu (patrz prawa strona il. 11.), to:

$$P_L = \frac{2}{l} \cdot \frac{\pi D^2}{4} \cdot \frac{1}{l^2} = \frac{1}{2} \cdot \frac{\pi D^2}{l^3} \quad (11.16)$$

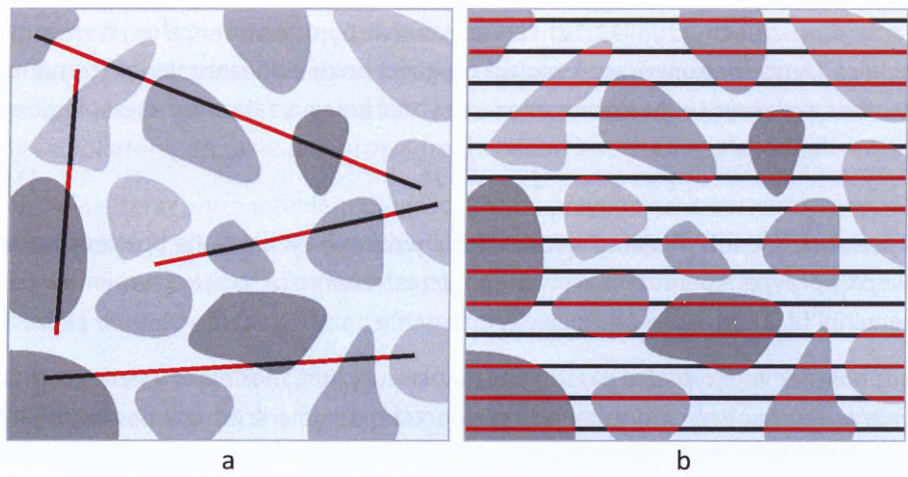
Połączenie wzorów (11.15) i (11.16) daje kolejną bardzo prostą zależność stereologiczną:

$$S_V = 2P_L \quad (11.17)$$

Do wyznaczenia wartości parametru S_V używamy metody liniowej (il. 11.9).

Za pomocą metody liniowej możemy wyznaczać długości siecznych lub liczby przecięć z granicami obiektów. W ogólnym przypadku orientacja siecznych może być dowolna (il. 11.19a). Gdyśmy mieli wyznaczyć długość linii na płaszczyźnie, to można skorzystać ze wzoru (11.14) nawet wtedy, gdy układ linii wykazuje orientację. Wystarczy wtedy przyjąć sieczne w kształcie okręgu. W przypadku powierzchni względnej S_V sprawa jest trudniejsza, ale ten problem przeanalizujemy nieco później. Jeżeli układ powierzchni nie wykazuje orientacji i jest przypadkowy, to używamy zależności (11.17). Możemy wtedy przyjąć układ poziomych, równoległych siecznych (il. 11.19b).

W przypadku analizy komputerowej zdarza się, że badacze modelują układ siecznych, ale nie ma takiej potrzeby. Piksele można uznać za punkty układu poziomo



Il. 11.9. Metoda liniowa. Opis w tekście

zorientowanych siecznych, z których część zaznaczono na il. 11.9b. Sumaryczna długość tych siecznych będzie równa iloczynowi liczby pikseli obrazu oraz jego rozdzielczości, czyli fizycznej odległości, odpowiadającej przejściu od jednego piksela do następnego. Można też modelować w ten sposób sieczne pionowe. Ich sumaryczna długość będzie identyczna z poziomymi. Dodatkowo można analizować sieczne zorientowane pod kątem 45° lub 135°. Ich całkowita długość będzie większa od pionowych lub poziomych $\sqrt{2}$ razy. Innych orientacji siecznych nie można modelować w tak prosty sposób. Największą zaletą opisanego powyżej modelowania siecznych jest łatwość wyznaczania liczby przecięć P lub liczby obiektów N .

Tabela 11.6. Konfiguracje pikseli do wyznaczania liczby przecięć lub liczby obiektów w metodzie liniowej. Opis w tekście

Orientacja siecznych																			
0°	45°	90°	135°																
<table><tr><td>0</td><td>1</td></tr><tr><td>X</td><td>X</td></tr></table>	0	1	X	X	<table><tr><td>X</td><td>1</td></tr><tr><td>0</td><td>X</td></tr></table>	X	1	0	X	<table><tr><td>1</td><td>X</td></tr><tr><td>0</td><td>X</td></tr></table>	1	X	0	X	<table><tr><td>1</td><td>X</td></tr><tr><td>X</td><td>0</td></tr></table>	1	X	X	0
0	1																		
X	X																		
X	1																		
0	X																		
1	X																		
0	X																		
1	X																		
X	0																		
<table><tr><td>1</td><td>0</td></tr><tr><td>X</td><td>X</td></tr></table>	1	0	X	X	<table><tr><td>X</td><td>0</td></tr><tr><td>1</td><td>X</td></tr></table>	X	0	1	X	<table><tr><td>0</td><td>X</td></tr><tr><td>1</td><td>X</td></tr></table>	0	X	1	X	<table><tr><td>0</td><td>X</td></tr><tr><td>X</td><td>1</td></tr></table>	0	X	X	1
1	0																		
X	X																		
X	0																		
1	X																		
0	X																		
1	X																		
0	X																		
X	1																		

Zakładamy, że obiekty przeznaczone do analizy są zapisane w obrazie binarnym. Tło obrazu stanowią piksele o wartości 0, natomiast obiekty – piksele o wartości 1. Symbol X w tabeli 11.6 oznacza piksele, których wartość pomijamy. Aby wyznaczyć liczbę przecięć P dla jednej z czterech omawianych orientacji siecznych, wystarczy policzyć liczbę konfiguracji pikseli pokazanych w odpowiedniej kolumnie tabeli 11.6. Natomiast w celu wyznaczenia liczby obiektów N , wystarczy policzyć liczbę konfiguracji pikseli zawartych w odpowiedniej kolumnie, ale tylko w jednym wierszu. Można też wyznaczyć liczbę przecięć i skorzystać z bardzo prostej relacji:

$$N = \frac{1}{2}P, \quad \text{czyli} \quad N_L = \frac{1}{2}P_L \quad (11.18)$$

Spośród wielu zależności stereologicznych wskażemy jeszcze dwie. Pierwsza pozwala na ocenę długości właściwej niezorientowanego układu przypadkowo rozmieszczonych krzywoliniowych obiektów na podstawie pomiaru liczby punktów ich przecięć z płaszczyzną:

$$L_V = 2P_A \quad (11.19)$$

Pewne nadzieje na wyznaczenie względnej liczby obiektów na jednostkę objętości daje zależność znana jako wzór Mirkina:

$$N_V = \frac{N_A}{h} \quad (11.20)$$

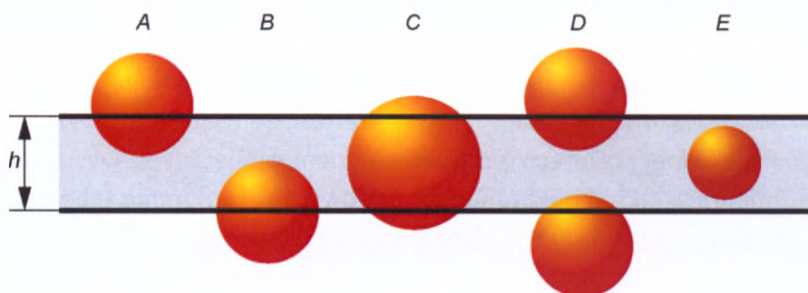
gdzie oznacza średnią wysokość obiektu, czyli długość jego rzutu pionowego. Problem w tym, że w praktyce zwykle nie możemy wyznaczyć tej wielkości. Sprytnie rozwiązanie problemu wyznaczania wartości N_V proponuje metoda znana pod nazwą disector (czyli podwójne cięcie). Wymaga ona analizy dwóch położonych blisko siebie przekrojów (il. 11.10).

Zakładamy, że wykonujemy dwa przekroje badanej struktury (zaznaczone na il. 11.10 czarną linią), oddalone od siebie o znaną odległość h . W oryginalnej wersji metody rozpatruje się trzy możliwe położenia obiektu:

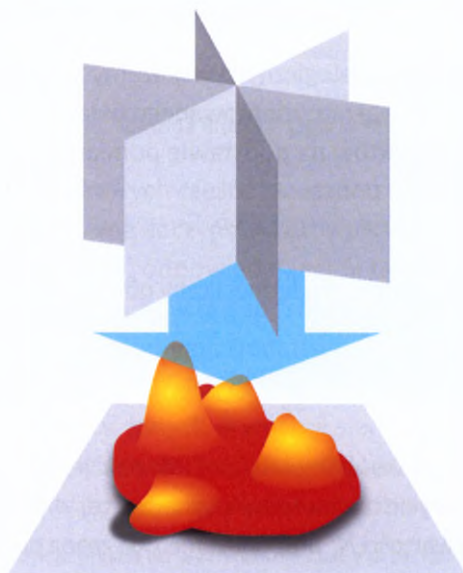
- A – obiekt przecięty tylko przed górnym przekrojem,
- B – obiekt przecięty tylko przez dolny przekrój,
- C – obiekt przecięty przez oba przekroje.

Porównując oba przekroje, z łatwością wydzielimy wspomniane trzy grupy obiektów. Do objętości pomiędzy przekrojami zaliczamy tylko obiekty z grupy A. Znając pole powierzchni badanego przekroju, wyznaczamy $N_A(A)$ obiektów grupy A. Możemy teraz wyznaczyć liczbę obiektów na jednostkę objętości:

$$N_V = \frac{N_A(A)}{h} \quad (11.21)$$



Il. 11.10. Disector. Opis w tekście



Il. 11.11. Idea metody pionowych przecięć. Opis w tekście

Metoda jest sprytna, prosta i skuteczna. Ma jednak mocno ograniczony zakres stosowania. Nadaje się praktycznie przede wszystkim do analizy gęstości jąder komórkowych w badaniach biologicznych. Wystarczy wtedy, żeby odległość h była mniejsza od rozmiarów jąder, co można uzyskać podczas cięcia preparatu za pomocą mikrotomu. W innych dziedzinach tej metody nie wykorzystamy, gdyż mogą pojawić się dwa rodzaje błędów:

- D – dwa obiekty mogą być przecięte w taki sposób, że sugeruje to dwukrotne przecięcie jednego obiektu (przypadek C na il. 11.10), przez co zaniżamy liczbę obiektów, bo jeden z nich niepotrzebnie odrzucamy,
- E – obiekt jest tak mały, że mieści się między płaszczyznami cięcia i nie wiemy o jego obecności, co również zaniża wynik pomiaru.

Budowa komórek biologicznych nie dopuszcza do opisanych powyżej sytuacji i dlatego metoda disectora może być skutecznie używana dla tych struktur.

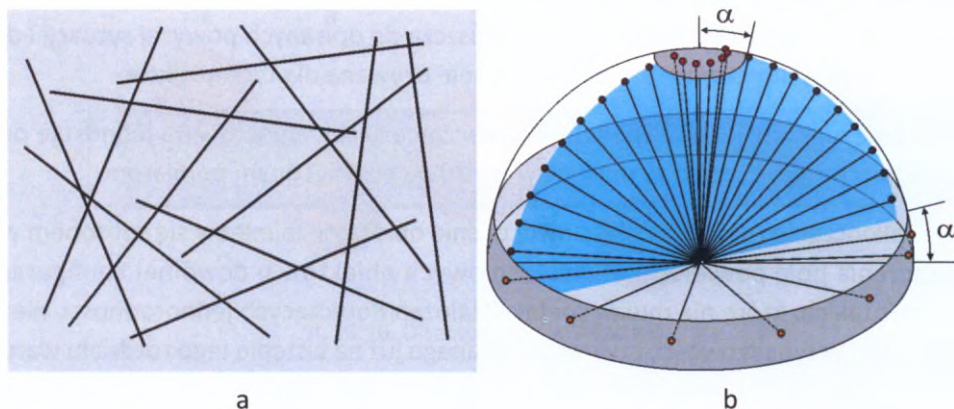
Nie omawialiśmy jeszcze sposobu wyznaczania liczby obiektów na jednostkę pola powierzchni N_A . Zostało to opisane w rozdziale poświęconym pomiarom.

Na zakończenie rozdziału o ilościowej ocenie obiektów zajmiemy się sposobem wyznaczenia pola powierzchni krzywoliniowych obiektów o dowolnej konfiguracji, a więc takich, które nie muszą spełniać założeń dotyczących jednorodności, niezależności i przypadkowości, czyli wspomnianego już na wstępie tego rozdziału warunku UIR. Możemy w takim przypadku wykorzystać metodę siecznych, pod warunkiem że sieczne będą UIR. Sposobem realizacji tego pomysłu jest użycie metody pionowych przecięć (ang. *vertical sections*). Nazwa pochodzi stąd, że metoda została opracowana na potrzeby badań biologicznych, gdzie preparat leżał na stole, a przekroje były wykonywane za pomocą ustawionego pionowo noża.

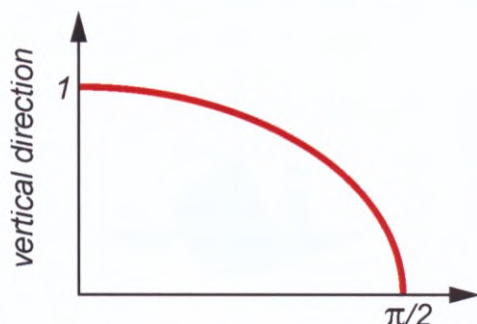
Ideę metody pokazuje il. 11.11. Widać w jego dolnej części nieregularny obiekt, który będzie przecinany szeregiem pokazanych u góry płaszczyzn, zgodnie z pionowym kierunkiem zaznaczonym niebieską strzałką. W ogólnym przypadku płaszczyzny tnące nie muszą przecinać się wzdłuż jednej linii jak na il. 11.11. Powinny jednak pokrywać systematycznie wszystkie możliwe orientacje. Zaleca się, aby tych orientacji było co najmniej trzy, jak na il. 11.11. Narysowany powyżej układ płaszczyzn tnących posiada jedną wspólną linię, wyznaczającą kierunek pionowy. Poszczególne płaszczyzny są obrócone względem siebie wzdłuż tej linii, która stanowi oś obrotu. Dlatego w metalografii używa się dla tej metody również innego określenia, nazywając ją metodą złądów (przekrojów) współosiowych.

Intuicja podpowiada nam, że skoro pionowe przecięcia są równomiernie rozłożone wokół jednego kierunku, to wystarczy, aby sieczne w każdej z tych płaszczyzn były przypadkowo rozmieszczone i obejmowały wszystkie kierunki z jednakowym prawdopodobieństwem (il. 11.12a), aby spełnić globalny warunek UIR. Niestety, tak się nie dzieje.

W celu analizy równomierności rozłożenia orientacji siecznych układamy je tak, aby wszystkie przecięły się w jednym punkcie. Jeżeli będą siecznymi UIR, to ich punkty przecięcia z powierzchnią kuli o środku w tym punkcie, który jest wspólny dla wszystkich siecznych, będą równomiernie rozłożone na powierzchni kuli. Na il. 11.12b mamy zaznaczoną na niebiesko jedną płaszczyznę z równomiernie rozłożonymi siecznymi. Taka sama liczba tych siecznych przecina pas wokół równika o szerokości określonej kątem α , jak i niewielkie koło wokół bieguna o powierzchni określonej tym samym kątem. Jeżeli wybierzemy jeszcze inne orientacje płaszczyzn tnących,



Il. 11.12. Metoda pionowych przecięć. Opis w tekście



Il. 11.13. Krzywoliniowa sieczna w kształcie cykloidy używana w metodzie pionowych przecięć. Długość krzywej pokazanej na rysunku wynosi dokładnie 2

to sytuacja powtórzy się. Mamy zatem nierównomierne rozmieszczenie siecznych. Zbyt dużo ich ma orientację zbliżoną do kierunku pionowego, a zbyt mało ma orientację zbliżoną do poziomej.

Można wykazać, że używając krzywoliniowych siecznych w kształcie odpowiednio zorientowanej cykloidy (il. 11.13), da się wyznaczyć powierzchnię właściwą dowolnego układu powierzchni w przestrzeni, stosując znany już nam wzór (11.17). Tylko tyle i aż tyle. Bardzo sprytne rozwiązanie.

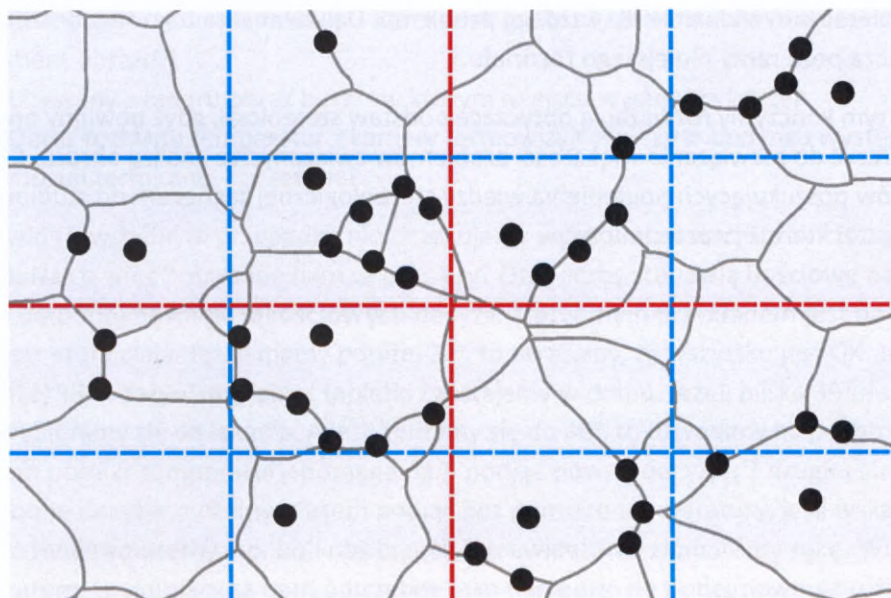
W tym miejscu warto zrobić małą dygresję. Przedstawione powyżej rozwiązanie zaproponowała w 1987 r. międzynarodowa grupa badaczy: Adrian Baddeley (Australijczyk), Hans-Jurgen Gundersen (Duńczyk) oraz Luis Manuel Cruz-Orive (Hiszpan pracujący w Szwajcarii). Pokazuje to, jak ważna jest międzynarodowa współpraca. Obecnie, dzięki rozwojowi Internetu, możemy być ogniwem takiej współpracy i warto z tego korzystać. Jest jeszcze jedna ciekawostka. Po raz pierwszy zależność (11.17)

pozwalającą na szacowanie wartości S_V , zaproponował w 1945 r. Sarkis Sałtykov, uznany na całym świecie Gruzin, którego książki wydawano m.in. w Niemczech, ale któremu władze nigdy nie pozwoliły wyjechać z Rosji. Trzeba było czekać aż 42 lata, zanim ktoś wymyślił tak proste, ale bardzo skuteczne rozszerzenie metody.

W przypadku metod analizy obrazu nie możemy równie skutecznie modelować cykloid, jak poziome i pionowe sieczne. Można jednak przygotować cyfrowe przybliżenie zestawu cykloid o grubości jednego piksela i przekroje powierzchni, których pole chcemy wyznaczyć, również sprowadzone do linii o grubości pojedynczego piksela. Wspólna część obu zestawów linii da nam poszukiwaną liczbę przecięć.

Najbardziej kłopotliwa w analizie stereologicznej jest ocena rozmieszczenia obiektów lub niejednorodności struktury. Często te dwie cechy są mylone ze sobą. Wprawdzie, jeżeli obiekty są rozmieszczone nierównomiernie, to struktura jest niejednorodna, ale z faktu jej niejednorodności nie płyną żadne konkretne informacje dotyczące rozmieszczenia obiektów.

Wybrane problemy, związane z oceną rozmieszczenia obiektów, można prześledzić, analizując il. 11.14. Widzimy tam strukturę komórkową oraz szereg czarnych obiektów. Jednym ze sposobów oceny ich rozmieszczenia jest podanie względnej liczby obiektów znajdujących się na granicach komórek. W tym przypadku wynosi ona 0,8,



Il. 11.14. Schemat struktury wyjaśniający problemy oceny rozmieszczenia obiektów.
Opis w tekście

bo na granicach komórek mamy 32 obiekty, a poza nimi tylko 8. W praktyce może okazać się, że łatwiej będzie ocenić względny udział powierzchniowy niż liczebność. Obydwa sposoby są dobre.

Zdecydowanie trudniej będzie ocenić jednorodność. Przedyskutujemy to na przykładzie struktury komórkowej (il. 11.14). Do oceny niejednorodności możemy wykorzystać współczynnik zmienności W , który jest stosunkiem odchylenia standardowego do wartości średniej. Jeżeli podzielimy obraz z il. 11.14 na cztery obszary (czerwone linie podziałowe), to dla liczebności ziaren $W = 0,077$. Dzieląc obszar obrazu na 16 segmentów (niebieskie i czerwone linie podziałowe brane łącznie), dostaniemy $W = 0,29$. Jest to, co prawda, czterokrotnie więcej niż poprzednio, ale trzeba wziąć od uwagę, że powierzchnia ocenianego obszaru w drugim przypadku jest zaledwie dwukrotnie większa od średniego pola powierzchni komórki. Zatem można przyjąć, że komórki są rozmieszczone jednorodnie.

W przypadku czarnych obiektów i podziału na cztery obszary $W = 0$, a więc mamy idealną jednorodność, ale przejście na 16 segmentów powoduje wzrost współczynnika zmienności do wartości $W = 1,00$. Z tego punktu widzenia obiekty są rozmieszczone bardzo niejednorodnie. Zwróćmy uwagę, że te bardzo duże różnice w wynikach dotyczą tego samego obrazu. Dobór wielkości obszaru analizy do konkretnego problemu jest najważniejszym elementem w ocenie niejednorodności i trzeba go dobierać indywidualnie dla każdego problemu. Dalsza analiza tego zagadnienia wykracza poza ramy niniejszego rozdziału.

Na tym kończymy rozważania dotyczące podstaw stereologii, gdyż powinny one wystarczyć do rozwiązania większości problemów związanych z analizą obrazu. Czytelników poszukujących pogłębienia wiedzy stereologicznej zachęcam do studiowania bogatej literatury przedmiotu.

12. Pomiary

Rozważania na temat pomiarów w analizie obrazu zaczniemy od pewnych rozważań ogólnych, można powiedzieć filozoficznych. Filozofię będziemy tutaj rozumieli, zgodnie z definicją zawartą w Encyklopedii PWN, jako najbardziej ogólną, fundamentalną, racjonalną i krytyczną wiedzę o wszystkim, co istnieje, a nie jako naukę o skomplikowanym wyrażaniu prawd oczywistych.

Zwykle jesteśmy przekonywani, że informacja o charakterze ilościowym ma pewną przewagę nad informacją o charakterze jakościowym. Zanim potwierdzimy lub zanegujemy twierdzenie o przewadze informacji ilościowej, zastanówmy się, czego oczekujemy po analizie obrazu w przypadku różnych zakresów naszej aktywności:

- Diagnostyka obrazowa w medycynie – przede wszystkim chcemy dostać odpowiedź na pytanie: czy pacjent jest zdrowy? A jeżeli nie jest zdrowy, to jak mu pomóc?
- Kontrola jakości w przemyśle elektronicznym – podstawową, jeśli nie jedyną informacją, jakiej oczekujemy, jest odpowiedź na pytanie: czy dany element jest poprawny, czy też należy go odrzucić?
- OCR (rozpoznawanie tekstu) – który znak graficzny przedstawia oceniany fragment obrazu?
- Uzyskany z radaru obraz burz – w którym miejscu występują burze?
- Obraz rozkładu temperatur z kamery termowizyjnej – czy w budynku występują mostki termiczne, czy też nie?

Jak widać, w żadnym przypadku nie oczekujemy wyniku w postaci liczby czy zbioru liczb! Na co więc potrzebne nam są pomiary? Otóż liczby stanowią **ilościowe** narzędzie do podejmowania **jakościowych** decyzji. Klasycznym przykładem jest pomiar temperatury ciała. Jeżeli mamy poniżej 37°, to uważamy, że wszystko jest OK. Jeżeli poniżej 38° – zażywamy jakieś tabletki i zostajemy w domu. Jeżeli blisko 39° – szybko wybieramy się do lekarza. A jeśli zbliżamy się do 40°, to dzwoniemy na pogotowie. Zatem pomiar temperatury pomaga nam podjąć pewną decyzję. Z drugiej strony, podobne decyzje możemy czasem podjąć bez pomiaru temperatury, jeśli wskazują na to inne symptomy, np. boli nas brzuch lub ewidentnie złamaliśmy rękę. Wiemy już zatem, że pomiary są nam potrzebne jako narzędzie do podejmowania różnych decyzji. Nie są jednak same w sobie celem naszych działań.

Począwszy od drugiej połowy XX wieku coraz częściej podczas wykonywania pomiarów korzystamy z pomocy komputerów. Dokonywane przez te urządzenia pomiary mają z punktu widzenia użytkownika dwie podstawowe cechy:

- pomiary są powtarzalne (kolejny pomiar tego samego obiektu, wykonywany w tych samych warunkach, da te same lub bardzo zbliżone wyniki),
- pomiary są wykonywane szybko.

Obie cechy są dla nas bardzo korzystne. Sprawia to, że sam fakt korzystania z komputerów bywa ciągle jeszcze używany jako element reklamy. Możemy spotkać warsztaty samochodowe z planszą: komputerowa regulacja zbieżności lub komputerowe wyważanie kół. Ma to przekonać potencjalnych klientów o rzekomo wyższej jakości usług świadczonych przy użyciu komputerowego wspomagania. Tymczasem złe umieszczenie komputerowych czujników może spowodować wręcz popsucie ustawień zawieszenia.

Wygoda w korzystaniu z komputerowych pomiarów spowodowała, że w naszej podświadomości ulokowała się trzecia cecha pomiarów komputerowych:

- pomiary komputerowe są lepsze, bo są bezbłędne.

Niestety, ta powszechna wiara nie ma żadnego uzasadnienia. Ani praktycznego, ani teoretycznego. Ale jest dość powszechna.

W dalszej części poświęcimy sporo miejsca analizie błędów występujących podczas komputerowych pomiarów oraz omówimy niektóre sposoby ich unikania lub przynajmniej ograniczania. Zanim przejdziemy do omawiania pomiarów w komputerowej analizie obrazu, zapamiętajmy, że:

- Nie należy się bać komputerów, gdyż jest to po prostu jedno z narzędzi, którymi dysponujemy. Jak każde narzędzie ma swoje zalety i ograniczenia. Trzeba jedynie umiejętnie korzystać z tego narzędzia.
- Komputerowe pomiary mogą być obciążone bardzo dużymi błędami, niejednokrotnie większymi od mierzonej wartości.
- Znajomość potencjalnych błędów może nam pomóc w takim przeprowadzeniu pomiarów, aby utrzymać błędy na akceptowalnym poziomie.
- Znajomość potencjalnych błędów może nas również uchronić przed niewłaściwą interpretacją wyników.

Istnieje wiele ilościowych charakterystyk obrazów, np. rozmiary obrazu, zakres poziomów szarości, wartość średnia, mediana, moda i wariancja poziomów szarości czy histogram poziomów szarości. Są one ważne i przydatne, ale nie będziemy ich traktować jako pomiary. Nazwę tę zarezerwujemy obecnie do charakterystyk przede

wszystkim geometrycznych, możliwych do przeprowadzenia na obrazach binarnych, ewentualnie etykietowanych lub z użyciem wirtualnych obiektów.

Nie ma to większego wpływu na dokładność naszych analiz, ale trzeba pamiętać, że w analizie obrazu mamy dwie klasy pomiarów: pomiary globalne, przeprowadzane dla całego obrazu, oraz pomiary lokalne, przeprowadzane dla pojedynczych obiektów. W niektórych systemach analizy obrazu te dwa rodzaje pomiarów są od siebie wyraźnie odseparowane.

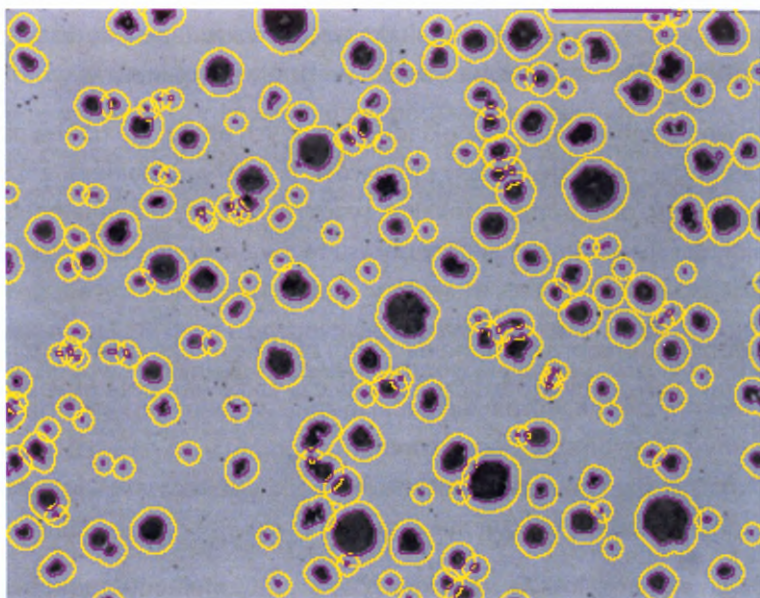
Pomiary globalne, przeprowadzane na całym obrazie, mogą nam dostarczać takich informacji jak: liczba obiektów, liczba cięciw czy całkowite pole powierzchni analizowanych obiektów.

Pomiary lokalne, dla konkretnych obiektów, będą obejmowały wiele parametrów, które zostaną dokładniej omówione w dalszej części.

W niektórych systemach wykryte podczas binaryzacji elementy zamienia się na wirtualne obiekty i dopiero one są analizowane. Takie rozwiązanie, stosowane m.in. w systemie analizy obrazu Aphelion, ma kilka istotnych zalet:

- Granice obiektów można wyświetlać w formie nakładki na analizowany obraz, przez co można łatwo wizualnie określić, na ile wyniki detekcji są poprawne, albo sprawdzić, które elementy obrazu wykazują poszukiwane przez nas cechy.
- Obiekty po przeprowadzeniu pomiarów można łatwo filtrować i np. usunąć obiekty mniejsze od założonego progu wielkości lub wyodrębnić obiekty o założonym kształcie.
- Obiekty można poddawać pewnym modyfikacjom, np. filtracji maksymalnej, która zwiększa ich rozmiary. Tak przekształcone obiekty mogą na siebie zachodzić, ale nie stanowi to żadnej przeszkody dla ich analizy. Gdybyśmy zamiast obiektów filtrowali obraz, to zachodzące na siebie obszary zostałyby sklejone (il. 12.1).

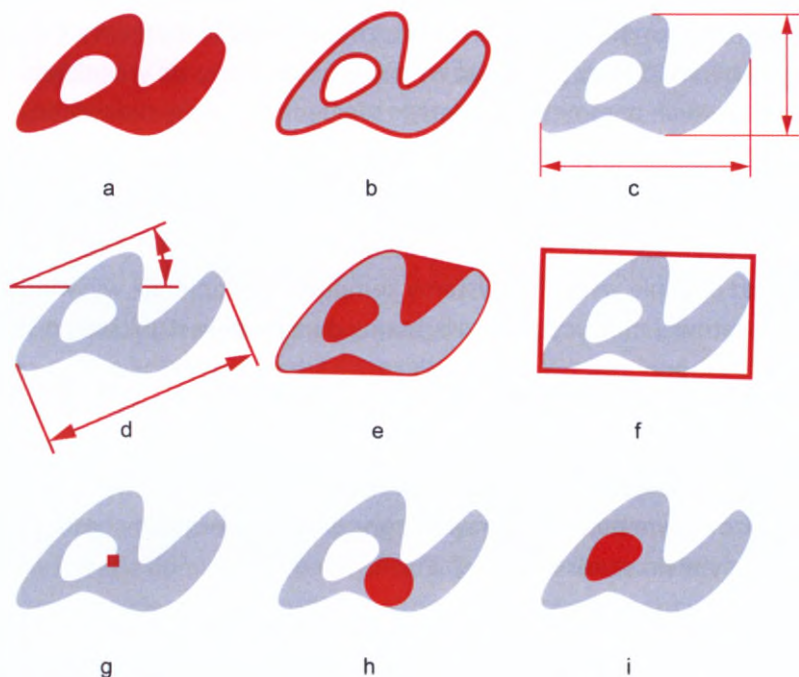
Uwaga: W celu ułatwienia opisu parametrów jednostką wszystkich pomiarów omawianych w tym rozdziale jest piksel, czyli pojedynczy punkt obrazu. Kwadrat składający się z 9 pikseli ma powierzchnię 9 i bok o długości 3. W rzeczywistych pomiarach należy obraz skalibrować, czyli ustalić, jakiej fizycznej odległości w pionie i poziomie odpowiada przejście od jednego piksela do drugiego. Można te wartości albo wpisać, albo ustalić interaktywnie, podając długość rysowanego na obrazie odcinka. W niektórych systemach, zwłaszcza dedykowanych na potrzeby mikroskopii, obrazy są kalibrowane już na etapie ich akwizycji. Odpowiednie dane są zapisywane w nagłówku pliku i mogą być wykorzystane przez system analizy obrazu.



Il. 12.1. Obiekty ilustrujące wykryte podczas binaryzacji elementy (purpurowe linie) oraz te same obiekty po filtracji filtrem maksymalnym (żółte linie). Widoczne jest nakładanie się na siebie obiektów, które jednak wciąż pozostają odrębne

W przypadku pojedynczych obiektów mamy pozornie bardzo dużą liczbę różnych parametrów. W rzeczywistości tych mierzonych parametrów jest znacznie mniej, a pozostałe są wyliczane jako funkcje ustalonych wcześniej podstawowych miar. Większość podstawowych parametrów, które mierzymy na poszczególnych obiektach, ukazuje il. 12.2. Wśród nich znajdują się:

- **pole powierzchni**, które jest po prostu równe liczbie pikseli tworzących obiekt (il. 12.2a);
- **obwód**, którego wyznaczenie okazuje się dużym problemem i zostanie omówione dokładnie później (il. 12.2b);
- **średnice Fereta**, czyli po prostu rzuty. Mamy poziomą średnicę Fereta, czyli rzut na kierunek poziomy, której długość jest równa różnicy maksymalnej oraz minimalnej wartości współrzędnych X punktów obiektu. Analogicznie wyznacza się pionową średnicę Fereta. Nazwa pochodzi od nazwiska francuskiego uczonego, który w roku 1930 użył długości rzutów do charakteryzowania geometrii ziaren w materiałach (il. 12.2c);
- **maksymalna średnica Fereta**, która jest podawana łącznie z kątem jej orientacji, stanowi również miarę wielkości obiektu i jest używana też do oceny orientacji wydłużonych obiektów. Na potrzeby oceny kształtu używa się niekiedy zbioru średnic Fereta, wyznaczonych dla kilku systematycznie zmienianych kierunków (il. 12.2d);



Il. 12.2. Podstawowe miary obiektów: a) pole powierzchni, b) obwód, c) średnice Fereta, d) maksymalna średnica Fereta, e) minimalne wypukłe otoczenie, f) minimalny prostokąt opisany na obiekcie, g) środek ciężkości, h) maksymalna szerokość, i) liczba otworów

- **minimalne wypukłe otoczenie**, którego kształt przyjmie guma naciągnięta na obiekt (il. 12.2e). Porównanie powierzchni obiektu i jego wypukłego otoczenia pozwala na ocenę kształtu obiektu, zwłaszcza jego wypukłości lub wklęsłości;
- **minimalny prostokąt opisany na obiekcie**, który jest prostokątem o najmniejszej możliwej powierzchni spośród wszystkich prostokątów opisanych na obiekcie. Jego zastosowanie jest analogiczne do wypukłego otoczenia i maksymalnej średnicy Fereta (il. 12.2f);
- **środek ciężkości**, który jest unikalnym punktem dla każdego obiektu. Obiekt podparty w tym punkcie pozostaje w równowadze. Współrzędne środka ciężkości mogą być wykorzystywane do oceny położenia obiektu (il. 12.2g). Z podstaw mechaniki (a konkretnie z równowagi momentów) wynika, że współrzędne środka ciężkości to średnie arytmetyczne współrzędnych wszystkich pikseli danego obiektu;
- **maksymalna szerokość**, która odpowiada średnicy największego koła wpisanego w obiekt. Może być użyta do charakteryzowania kształtu lub wielkości obiektów, np. porów w filtrze (il. 12.2h);
- **liczba otworów**, która jest jedną z topologicznych charakterystyk obiektów i może być używana m.in. do charakteryzowania kształtu (il. 12.2i).

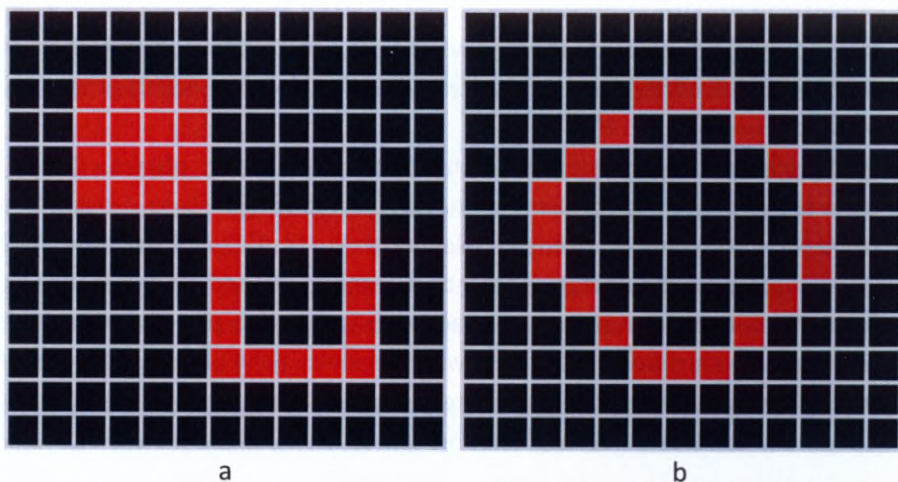
Oprócz wymienionych tu parametrów można wyznaczać momenty bezwładności, kierunki osi bezwładności, szacować wymiar fraktalny itp., ale w praktyce nie korzystamy często z tych parametrów i dlatego nie będą one dalej omawiane.

Jednym z najczęściej wykonywanych pomiarów jest zliczanie obiektów widocznych w obrębie obrazu. Zadanie jest proste tylko pozornie, gdyż musimy najpierw określić zasady spójności, czyli umówić się, kiedy dwa punkty należą do jednego obiektu. Ponieważ obecnie wszystkie systemy komputerowe opierają wyświetlanie obrazu na kwadratowej macierzy punktów, mamy dwa modele najbliższego sąsiedztwa: 4-punktowe i 8-punktowe. W przypadku sąsiedztwa 4-punktowego za najbliższych sąsiadów uznaje się stykające się z analizowanym pikselem inne piksele, położone nad, pod oraz z lewej i z prawej strony (lewa strona il. 12.3). W przypadku 8-punktowego sąsiedztwa do sąsiadów zalicza się dodatkowo piksele stykające się narożnikami z analizowanym punktem (prawa strona il. 12.3). Identyczny problem spójności jest omawiany w przypadku operacji morfologicznych.

Wybór modelu spójności może mieć bardzo istotny wpływ na liczbę analizowanych obiektów (il. 12.4). Jeżeli wybierzemy 4 najbliższych sąsiadów, czyli czterospójne



Il. 12.3. Modele najbliższego sąsiedztwa: 4-punktowe (z lewej) oraz 8-punktowe (z prawej)



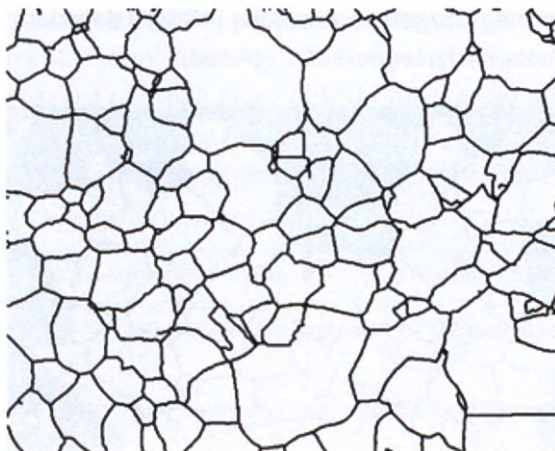
Il. 12.4. Wpływ rodzaju spójności na liczbę obiektów w obrazie. Opis w tekście

sąsiedztwo, to na il. 12.4a będziemy mieli dwa kwadraty, z czego jeden pełny. Jeśli jednak przejdziemy na model 8 najbliższych sąsiadów, czyli ośmiospójne sąsiedztwo, to na il. 12.4a będziemy mieli tylko jeden obiekt. Jeszcze większe różnice zajdą w przypadku il. 12.4b. Widać na nim pojedynczy obiekt, ale pod warunkiem, że mamy 8 najbliższych sąsiadów. Jeżeli przejdziemy na sąsiedztwo czterospójne, to liczba obiektów wzrośnie aż do 12. Jak widać, konsekwencje zmian zasad spójności mogą wręcz drastycznie wpłynąć na ocenę liczebności obiektów.

Z powyższej analizy płyną następujące wnioski:

- Niezależnie od wybranego modelu spójności obiektów, podczas ich zliczania mogą pojawiać się błędy, których nie da się przewidzieć ani uniknąć.
- Jeżeli w obrazie dominują duże obiekty, których najwęższe elementy mają co najmniej grubość kilku pikseli, to lepiej wybrać sąsiedztwo 4-punktowe, aby uniknąć sklejania obiektów stykających się tylko narożnikami.
- Jeżeli w obrazie dominują małe obiekty, których najwęższe elementy mają często grubość zaledwie pojedynczych pikseli (wszelkie obrazy cienkich włókien lub linii, niezależnie od ich krzywizny i orientacji), to lepiej wybrać sąsiedztwo 8-punktowe, aby uniknąć interpretacji cienkiego włókna jako zbioru odizolowanych pikseli.

Na tym nie kończą się problemy ze zliczaniem obiektów. Nawet, jeśli uzgodnimy zasady spójności, to problem stwarza zliczanie obiektów przeciętych przez brzeg obrazu (il. 12.5).



Il. 12.5. Przykład obrazu z obiektami przeciętymi przez brzeg

Można w takim przypadku zastosować metodę Jeffriesa, zgodnie z którą liczba obiektów N wynosi:

$$N = N_w + \frac{1}{2} N_B + \frac{1}{4} N_N \quad (12.1)$$

gdzie:

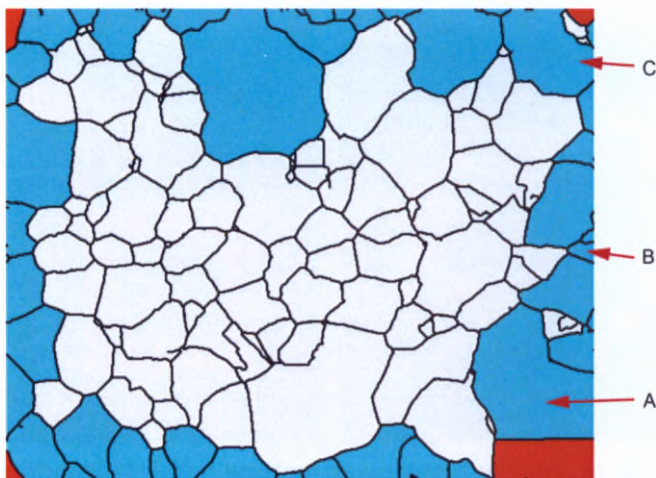
N_w – liczba obiektów leżących wewnątrz obrazu; są to białe obiekty na il. 12.6,

N_B – liczba obiektów przeciętych przez brzeg obrazu; są to niebieskie obiekty na il. 12.6,

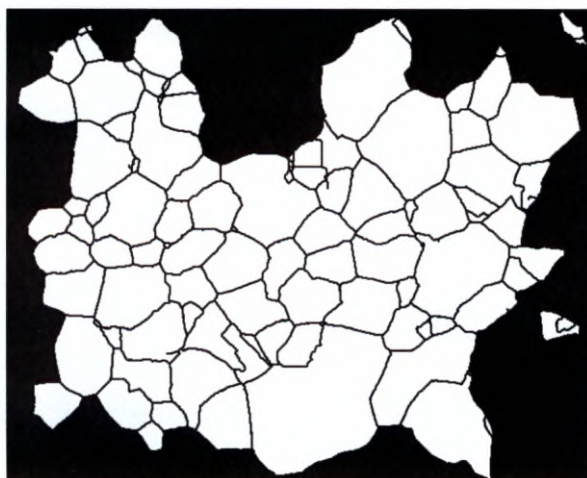
N_N – liczba obiektów leżących w narożnikach obrazu.

Współczynniki przy poszczególnych liczebnościach wynikają z założenia, że obiekty przecięte przez brzeg leżą w dwóch stykających się obrazach. Co prawda, jakiś obiekt może leżeć w większości (a więc będzie to powyżej połowy) w analizowanym obrazie (obiekt A na il. 12.6), ale za to inny będzie wewnątrz analizowanego obrazu jedynie w niewielkim fragmencie (obiekt B na il. 12.6). Średnio po połowie obiektu będzie znajdowało się wewnątrz obrazu i poza nim. Pewnym problemem, który powoduje powstanie systematycznego błędu, są obiekty przecięte przez dwa brzegi (obiekt C na il. 12.6), ale są to sytuacje rzadkie, które nie mają istotnego wpływu na wyniki zliczania obiektów. Ponadto nie występują w każdym obrazie. Metodę Jeffriesa można stosować niezależnie od tego, czy obiekty stykają się i wypełniają całą powierzchnię, jak na il. 12.5 i 12.6, czy też są od siebie odizolowane.

Możemy teraz wyznaczyć średnią powierzchnię obiektów, a konkretnie ziaren w analizowanym obrazie. Ma on rozmiar 600x481 pikseli, zatem cały obraz składa się z 288 600 pikseli. Obraz na il. 12.6 zawiera 108 ziaren wewnętrznych, 67 przeciętych przez brzeg i 4 narożne, co zgodnie z metodą Jeffriesa daje nam 142,5 ziarna. Zatem średnia powierzchnia ziarna wynosi 2025 pikseli.



Il. 12.6. Metoda Jeffriesa zliczania obiektów w obrazie. Opis w tekście



Il. 12.7. Wygląd obrazu po usunięciu obiektów przeciętych przez brzeg

Czasami wymagana jest informacja o całych obiektach, a ściślej rzecz biorąc o ich rozkładzie pól powierzchni, i wtedy nie możemy brać pod uwagę obiektów przeciętych przez brzeg. Systemy analizy obrazu ułatwiają taką operację i dzięki operacji *border kill*, omówionej w ramach operacji morfologicznych, zostają tylko obiekty zaznaczone na il. 12.6 kolorem białym. Otrzymujemy wtedy obraz z il. 12.7.

Jeżeli przeanalizujemy obraz na il. 12.7, to sumaryczna powierzchnia ziaren wynosi 165 316, co przy 108 ziarnach wewnętrznych daje średnią 1531 pikseli. Obliczając tutaj średnią, oczywiście zaniżyliśmy wynik, bo przecież część pikseli poświęciliśmy na granice pomiędzy obiektami. Możemy zatem wynik skorygować, dzieląc pole powierzchni całego obszaru zajętego przez obiekty (wystarczy operacja zamknięcia o wielkości 1, bo granice pomiędzy ziarnami mają taką właśnie grubość) przez ich liczbę, czyli 108. Otrzymamy nieco większą średnią powierzchnię obiektu, wynoszącą 1653 piksele. Zestawienie omawianych wyników zawiera tabela 12.1.

Tabela 12.1. Średnia powierzchnia obiektu z il. 12.6 w funkcji metody wyznaczenia

Metoda Jeffriesa	Ziarna wewnętrzne	Ziarna wewnętrzne z korektą
2025	1531	1653

Różnice w wynikach (tab. 12.1) są bardzo duże. W tym konkretnym przypadku duża średnia powierzchnia, wyznaczona za pomocą metody Jeffriesa, wynika z faktu, że brzeg obrazu przeciął znaczną liczbę ziaren o dużej powierzchni. A te duże powierzchnie wliczyliśmy jako przypisane tylko połowie liczby ziaren przeciętych przez brzeg. Konsekwentnie, obliczenia pomijające ziarna przecięte przez brzeg obrazu

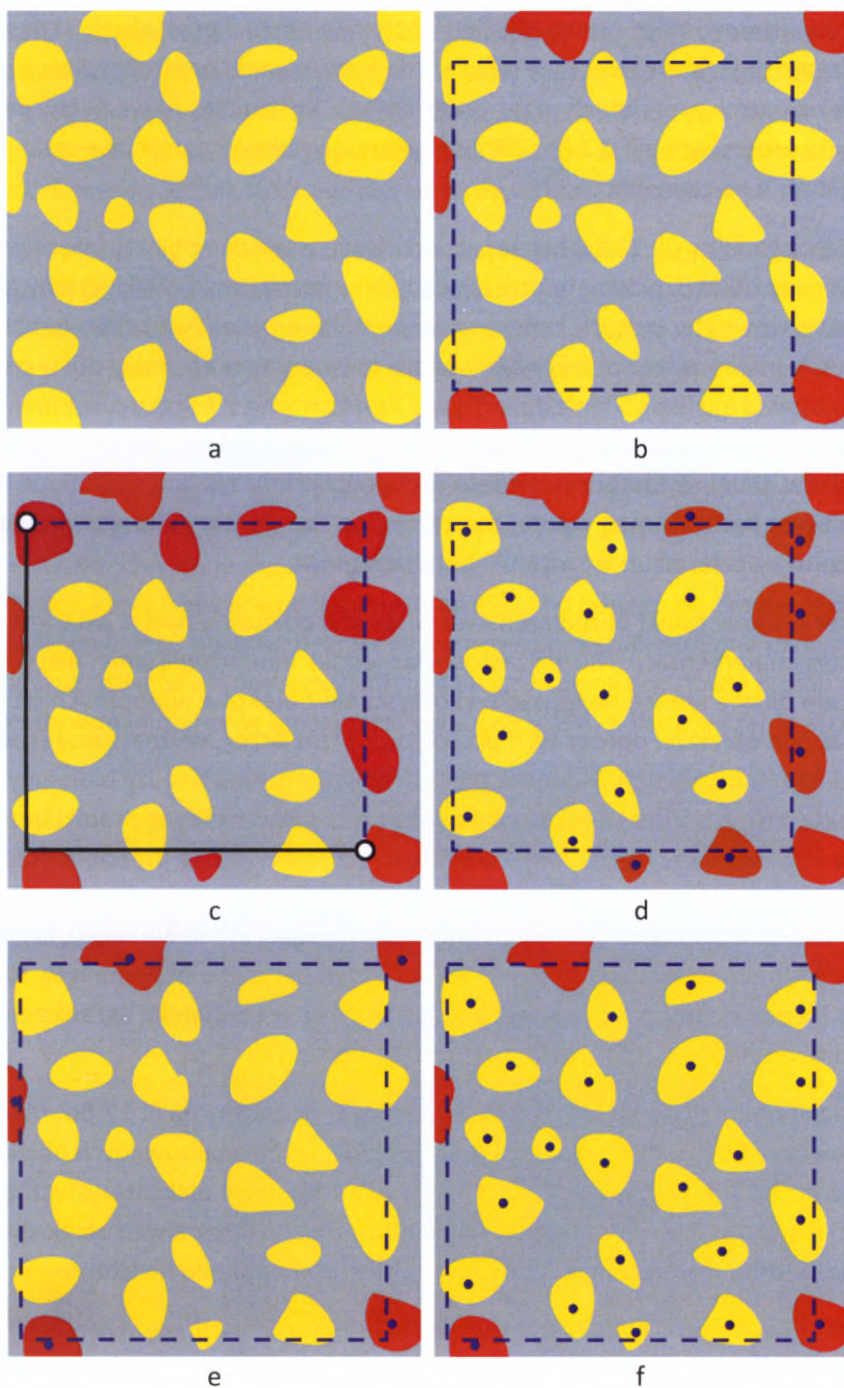
prowadzą do zaniżenia ich średniego pola powierzchni. Taki wynik jest w znacznym stopniu konsekwencją systematycznego błędu wprowadzonego przez usunięcie ziaren przeciętych przez brzeg obrazu. W tym konkretnym przypadku nakładają się na niego błędy wynikające z niejednorodności badanej struktury. Dlatego w praktyce zaleca się, aby przy tego rodzaju pomiarach brać pod uwagę obrazy, w których mamy przynajmniej ok. 500 ziaren. Wtedy obserwowane tutaj fluktuacje będą najprawdopodobniej nieznaczne.

Warto też zwrócić uwagę, że przyjęcie poprawki, związanej ze stratą części pikseli na odtworzenie granic ziaren, spowodowało wzrost średniej powierzchni o ok. 8% (patrz tab. 12.1). Gdybyśmy analizowali obraz zawierający ok. 500 ziaren, czyli trzykrotnie więcej niż w przypadku il. 12.5, to różnica w średnim polu powierzchni, wyznaczonym z uwzględnieniem pikseli zużytych na rysowanie granic ziaren w stosunku do pomiaru bez tej poprawki, wynosiłaby ok. 14%. To naprawdę znacząca różnica, która pokazuje, jak rozbieżne mogą być wyniki pomiarów tylko dlatego, że przyjęliśmy nieco inne zasady ich prowadzenia. Różnice te wynikają tylko z samych pomiarów. Nie ma tu jakiegokolwiek analizy znaczenia wcześniejszych przekształceń obrazu, choć one również mogą istotnie wpłynąć na wyniki.

W ten sposób wyjaśniliśmy, dlaczego nie można bezkrytycznie przyjmować wyników komputerowych pomiarów. Zawsze potrzebna jest dokładna informacja, jak zostały one przeprowadzone. W dodatku, chcąc zachować porównywalność wyników, cały czas powinniśmy stosować tę samą metodę. Jeżeli badania będą prowadzone przez dłuższy czas, to zwykle lepiej jest używać konsekwentnie starej metody, nawet jeśli opracowaliśmy lepszą. Tę nowszą i lepszą należy zastosować dopiero do kolejnej serii badań, której wyniki nie będą bezpośrednio porównywane z wynikami badań archiwalnych.

Usuwanie obiektów przeciętych przez brzeg obrazu wprowadza do analizy błąd systematyczny. Polega on na tym, że przecięcie dużych obiektów jest bardziej prawdopodobne niż przecięcie małych. W ten sposób wśród obiektów pozostałych po usunięciu przeciętych przez brzeg mamy nadmiar mniejszych, co powoduje zaniżenie wyniku. Potwierdza to analiza wyników widocznych w tabeli 12.1, gdyż średnia powierzchnia ziaren, nawet po korekcie, jest mniejsza od tej, którą wyznaczaliśmy metodą Jeffriesa.

Okazuje się, że błędy wynikające z prostego usunięcia obiektów przeciętych przez brzeg, można dość łatwo wyeliminować dzięki zastosowaniu tzw. ramki bezpieczeństwa (il. 12.8). Załóżmy, że mamy zbadać rozkład wielkości obiektów widocznych w postaci żółtych plam na il. 12.8a. Oczywiście nie chcemy analizować obiektów przeciętych przez brzeg, bo nie wiemy, jaka jest ich całkowita powierzchnia.



Il. 12.8. Ramka bezpieczeństwa. Opis w tekście

Wprowadzamy zatem ramkę bezpieczeństwa w postaci prostokąta, skonstruowanego tak, że jest to największy możliwy do narysowania prostokąt, który nie obejmuje obiektów przeciętych przez brzeg obrazu. Ilustruje to narysowana przerywaną linią prostokąt na il. 12.8b. Z kolei wykluczone obiekty przecięte przez brzeg są zaznaczone na czerwono.

Do dalszej analizy (il. 12.8c) bierzemy pod uwagę wszystkie obiekty leżące wewnątrz ramki oraz obiekty przecięte przez lewy i dolny brzeg ramki. Wykluczamy jednak te obiekty, wewnątrz których kończą się wspomniane wyżej dwa brzegi ramki. Inaczej mówiąc, usuwamy też obiekty obejmujące lewy górny oraz prawy dolny róg ramki. Jest tu pełna analogia z metodą Jeffriesa. Tam braliśmy $1/4$ liczby obiektów leżących w narożnikach, tutaj bierzemy cały obiekt, ale tylko leżący w jednym z czterech narożników. Dalej, w metodzie Jeffriesa uwzględnialiśmy $1/2$ liczby obiektów przeciętych przez brzeg obrazu, a tutaj bierzemy pod uwagę wszystkie obiekty przecięte przez dwa z czterech boków ramki bezpieczeństwa.

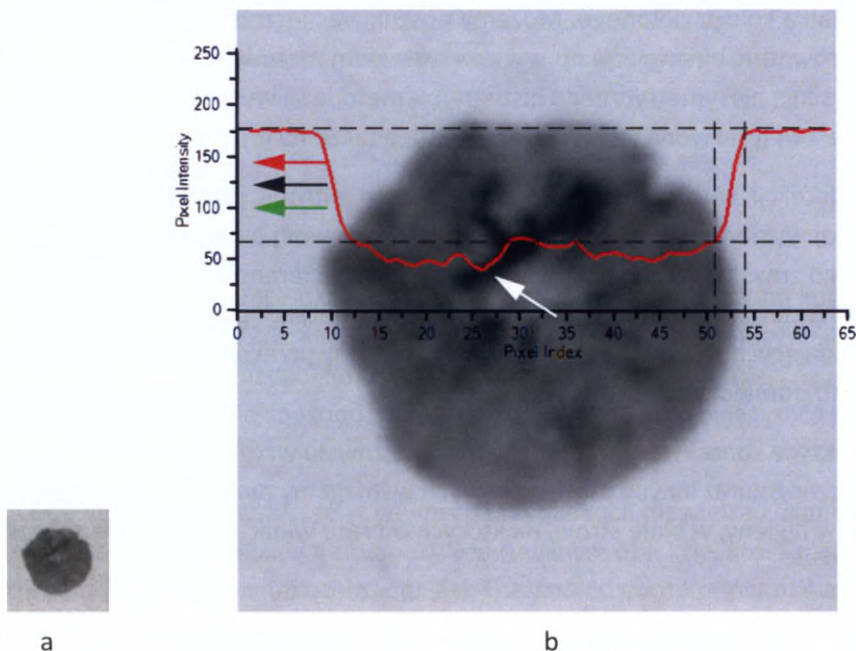
Wykorzystanie ramki bezpieczeństwa w sposób pokazany na il. 12.8.c jest dość łatwe do implementacji. Można zastosować jednak inne rozwiązanie, nieco trudniejsze, ale dające szansę na lepsze wykorzystanie obiektów widocznych na obrazie. W każdym obiekcie, oprócz tych przeciętych przez brzeg, wyznaczamy środek ciężkości (ciemne kółeczka widoczne na il. 12.8d). Do dalszej analizy bierzemy tylko te obiekty, których środki ciężkości leżą we wnętrzu albo na brzegu ramki bezpieczeństwa (żółte obiekty na il. 12.8d). Obiekty wykluczone (np. te przecięte przez prawy brzeg ramki) będą wzięte pod uwagę w sąsiedniej ramce. W ten sposób, zakładając pokrycie całej dostępnej przestrzeni przez przylegające do siebie ramki, każdy obiekt będzie wzięty pod uwagę tylko jeden raz i żaden nie zostanie pominięty. Nie mamy więc błędu systematycznego, wprowadzanego np. w przypadku ograniczenia się do usunięcia obiektów przeciętych przez brzeg obrazu.

Wariant wykorzystania ramki bezpieczeństwa, pokazany na il. 12.8d, również nie wprowadza błędu systematycznego. Co prawda, liczba analizowanych obiektów jest różna na il. 12.8c i 12.8d, ale nie jest to błąd metody, tylko efekt statystycznej zmienności wyników. Przy dostatecznie dużej liczbie przeanalizowanych obiektów różnice staną się albo niemierzalne, albo przynajmniej statystycznie nieistotne.

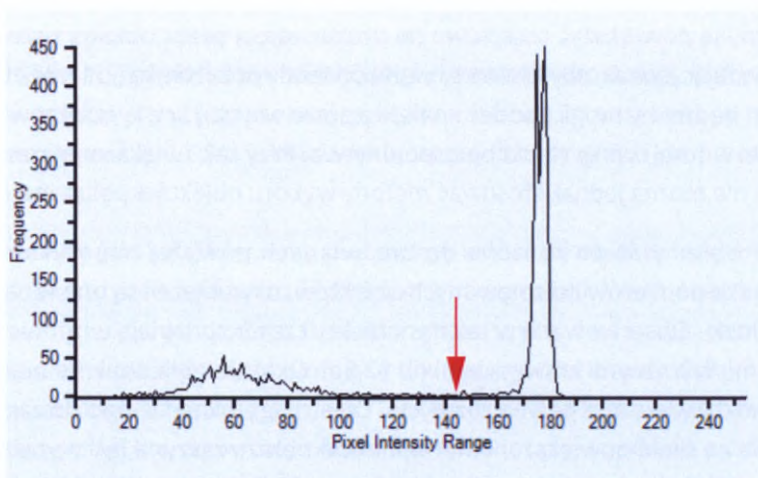
Metoda pokazana na il. 12.8d jest trudniejsza w implementacji, gdyż wymaga wyznaczenia współrzędnych środków ciężkości i następnie ich filtrowania. Metoda ta ma jednak pewien potencjał, pozwalający zwiększyć liczbę analizowanych w danym obrazie obiektów. Skoro nie bierzemy pod uwagę obiektów, których środki ciężkości leżą poza ramką,

to możemy ją powiększyć częściowo na obszar zajęty przez obiekty przecięte przez brzeg, uważając jednak, aby ich środki ciężkości leżały poza ramką (il. 12.8e). Dzięki temu zabiegowi będziemy mogli poddać analizie jeszcze większą liczbę obiektów niż w przypadku pierwotnej wersji ramki bezpieczeństwa. Przy tak zwiększonej ramce bezpieczeństwa nie można jednak stosować metody wyboru obiektów pokazanej na il. 12.8c.

Często problemy nieco zbliżone do omawianych powyżej napotykamy również w przypadku pomiarów odizolowanych obiektów, nawet jeżeli są one w całości obecne na obrazie. Obserwowane przez nas obiekty często sprawiają wrażenie ograniczenia ostrymi, wyraźnymi krawędziami (il. 12.9a). Okazuje się jednak, że niemal zawsze brzeg obiektu wykazuje pewne rozmycie. Często ogranicza się ono do zaledwie kilku pikseli, ale na silnie powiększonym fragmencie obrazu zaczyna być wyraźnie widoczne (il. 12.9b). Pokazany na tym rysunku brzeg ciemnego wydzielenia na jasnym tle jest rozmyty, a granice tego rozmycia wskazują pionowe przerywane linie z prawej strony il. 12.9b. Przyjmuje się, że najlepszy (czyli dający najmniejsze błędy) jest taki próg detekcji, który trafia w środek zakresu przejścia od jasnego do ciemnego obszaru w okolicach brzegu obiektu. Granice tego zakresu pokazują poziome przerywane linie, a wyznaczony na ich podstawie (czyli tak naprawdę na podstawie profilu pokazanego czerwoną linią) próg binaryzacji, wynoszący 124, pokazuje czarna strzałka.



Il. 12.9. Wpływ progu binaryzacji na wyniki pomiarów. Opis w tekście



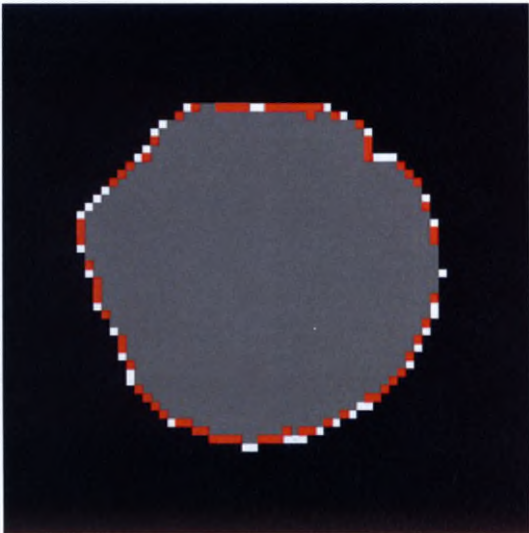
Il. 12.10 Histogram poziomów szarości i próg detekcji metodą maksimum entropii dla obrazu pokazanego na il. 12.9

Jeżeli opisany proces binaryzacji będziemy chcieli zautomatyzować, to jako dolny kres przejścia od jasnego do ciemnego obszaru najprawdopodobniej przyjmiemy globalne minimum poziomów szarości w obiekcie. Odpowiednią wartość pokazuje biała strzałka na il. 12.9, a wynikający stąd próg czułości, wynoszący 100, wskazuje strzałka koloru zielonego. Możemy poszukiwać jeszcze innego rozwiązania, czyli przeprowadzić binaryzację np. z wykorzystaniem metody maksimum entropii. Z powodu silnie niesymetrycznego histogramu metoda ta wyznaczy próg detekcji na poziomie 144 (czerwona strzałka na il. 12.10, a także na il. 12.9).

Próg detekcji, wyznaczony metodą maksimum entropii, jest najbliższy temu, co najczęściej wybieramy podczas klasycznej interaktywnej binaryzacji. Wynika to z cech naszego zmysłu wzroku, który lepiej rozpoznaje ciemne obszary na jasnym tle niż jasne na ciemnym tle. Dlatego wybieramy próg detekcji położony bliżej wartości odpowiadających jasnym elementom, choć jesteśmy przekonani, że wybieramy środek zakresu pomiędzy światłami i cieniami.

W praktyce konsekwencją tej cechy naszego zmysłu wzroku jest ubieranie personelu medycznego oraz innych osób, od których wymagamy zachowania najwyższych standardów higieny, w białe stroje, na których od razu widać nawet lekkie zabrudzenia.

Opisywane trzy metody binaryzacji dają trzy nieco inne wynikowe obrazy binarne, których różnice pokazuje il. 12.11. Wyniki te są bardzo zbliżone, zwłaszcza jeśli nie oglądalibyśmy wykrytych obiektów w co najmniej 10-krotnym powiększeniu, jak to się dzieje na il. 12.11.



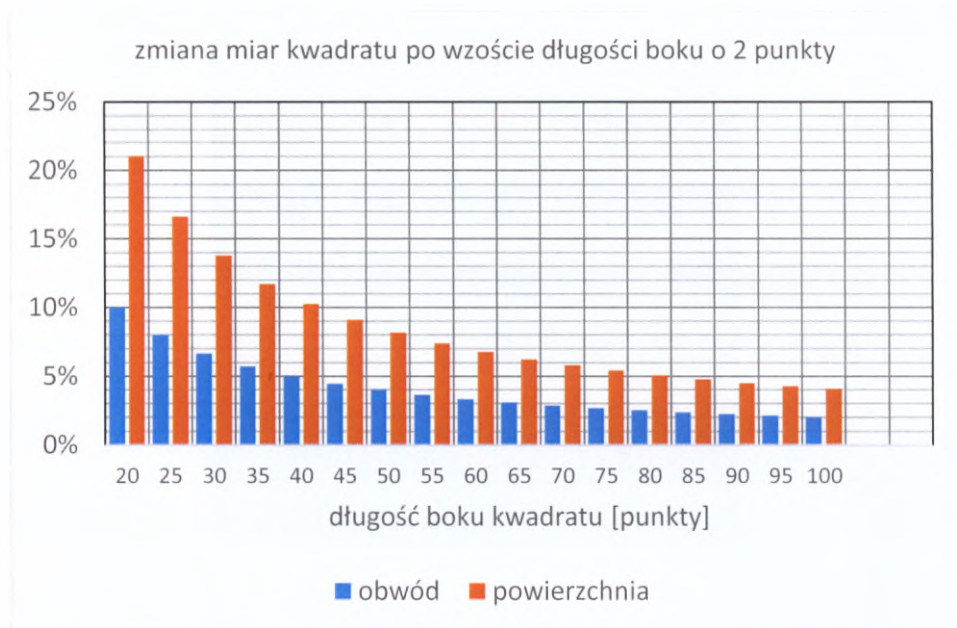
Il. 12.11. Wyniki detekcji obiektu z il. 12.9 za pomocą różnych progów detekcji. Próg dolny zawsze wynosił 0. Wynik detekcji z progiem 100 pokazuje szary obszar rysunku. Zwiększenie progu do wartości 124 powoduje dodanie pikseli zaznaczonych kolorem czerwonym, a zwiększenie progu do wartości 144 dodatkowo powoduje dodanie pikseli zaznaczonych na biało

Tabela 12.2. Wpływ progu detekcji na pole powierzchni obiektu z il. 12.9

Zakres progów detekcji	Pole powierzchni	Względny przyrost
0-100	1305	-
0-124	1375	5,36%
0-144	1418	8,66%

Wspomniane drobne różnice okazują się większymi, jeśli spojrzymy na konkretne liczby, co ilustruje tabela 12.2. Każdy z omawianych progów detekcji można uznać za poprawny. Po raz kolejny uzyskujemy wynik, wskazujący na to, że konieczne jest bardzo precyzyjne podawanie sposobu przetwarzania obrazu, jego binaryzacji i wreszcie procedury pomiarowej.

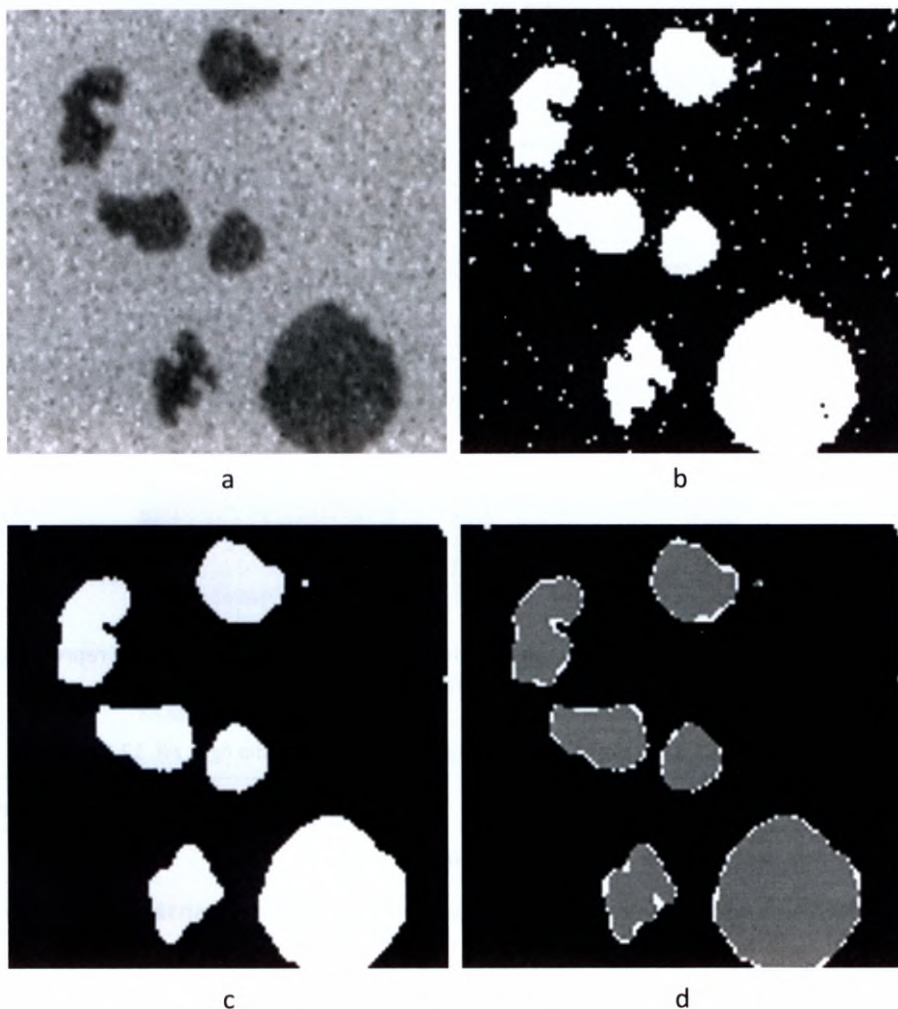
Na il. 12.11 możemy zauważyć, że w pewnych przypadkach niewielka zmiana progu detekcji może wywołać skokową zmianę w wykrywanym obiekcie, która najczęściej sprowadza się do dodania lub odjęcia pojedynczej warstwy pikseli wokół obiektu. Można łatwo wyliczyć, jak taka zmiana wpłynie na powierzchnię i obwód kwadratu. Wyniki odpowiedniej analizy przedstawia il. 12.12. Jak widać, różnice w wynikach są znaczące. Można z nich wyciągnąć wniosek, że jakakolwiek ilościowa charakterystyka



Il. 12.12. Zmiana obwodu i powierzchni kwadratu (boki zorientowane poziomo i pionowo) spowodowana dodaniem na zewnątrz pojedynczej warstwy pikseli

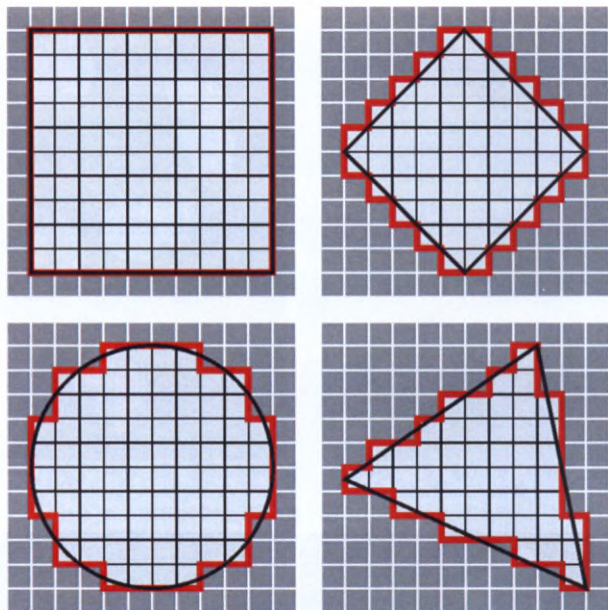
drobnych obiektów, o rozmiarach poniżej kilkunastu pikseli, jest obarczona ryzykiem bardzo dużego błędu, który może być porównywalny z mierzoną wielkością. Małe obiekty, zwłaszcza o rozmiarach pojedynczych pikseli, powinno się co najwyżej zliczać, bez podawania ich pola powierzchni, obwodu itp.

Różnice w polach powierzchni mogą być konsekwencją nie tylko różnic w przyjętych sposobach binaryzacji i wynikających stąd różnic w progach detekcji. Ilustruje to dobrze il. 12.13. Pokazany na il. 12.13a obraz z dość wysokim poziomem szumu został poddany binaryzacji metodą maksimum entropii, co dało wynik pokazany na il. 12.13b. Widoczne drobne białe punkty to wynik binaryzacji szumu i można go łatwo usunąć za pomocą filtra medianowego, praktycznie bez wpływu na kształt wykrytych obiektów. Można też zmienić kolejność i ograniczyć szum w obrazie wyjściowym, a dopiero później przeprowadzić binaryzację. Chcąc skutecznie ograniczyć szum w obrazie z il. 12.13a, konieczne jest przeprowadzenie filtracji medianowej na obszarze istotnie większym niż w przypadku obrazu z il. 12.13b. Wynikiem tego postępowania jest obraz binarny z il. 12.13c, w którym krawędzie obiektów są bardziej wygładzone niż na il. 12.13b. Różnice pomiędzy wynikami obu metod pokazano na il. 12.13d. Podobnie jak w przypadku różnych progów czułości, obie metody detekcji obiektów są poprawne, ale różnice w wielkości pól powierzchni zbliżają się do 10%.



Il. 12.13. Wpływ sposobu detekcji na końcowy wynik: a) wyjściowy obraz z wysokim poziomem szumu, b) łatwy do skorygowania wynik binaryzacji metodą maksimum entropii, c) wynik binaryzacji metodą maksimum entropii, przeprowadzonej po redukcji szumu, d) porównanie wyników detekcji

Do tej pory analizowaliśmy wpływ sposobu binaryzacji na pole powierzchni obiektu po binaryzacji. Obecnie zajmiemy się innym problemem, którym są błędy oceny pola powierzchni, spowodowane przybliżaniem prostej figury geometrycznej przez zbiór pikseli obrazu. Cztery przykładowe figury, o łatwym do ustalenia polu powierzchni, przedstawia il. 12.14. Analizowane figury zostały zaznaczone czarną linią, a cyfrowe przybliżenia ich brzegu linią czerwoną. Względne błędy oceny pola powierzchni tych figur przedstawia tabela 12.3.



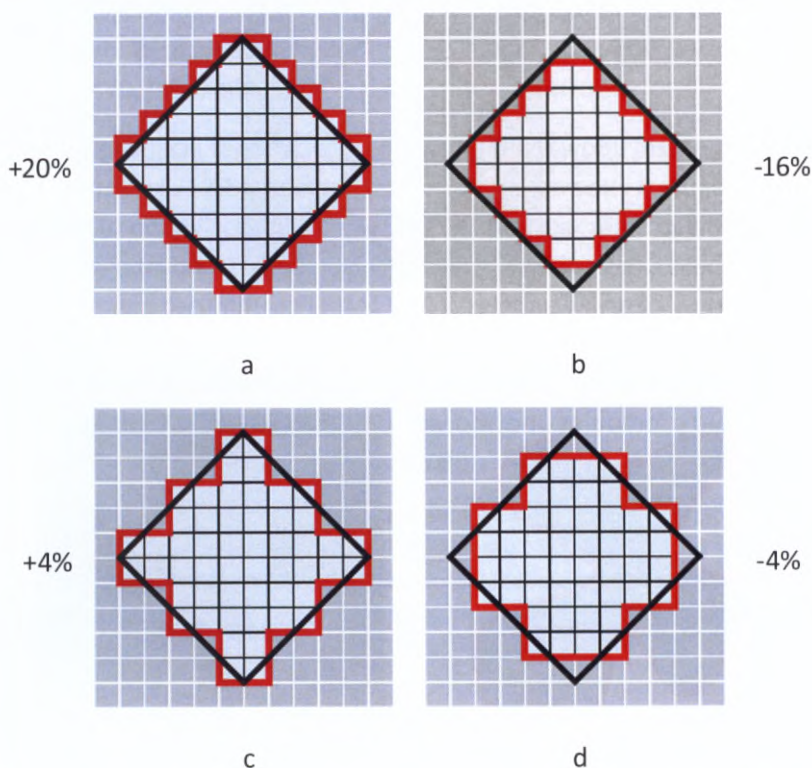
Il. 12.14. Proste figury użyte do badania błędów wyznaczania obwodu oraz ich reprezentacja w cyfrowym obrazie. Opis w tekście

Tabela 12.3. Wybrane błędy oceny pola powierzchni dla figur z il. 12.14

Figura	Błąd oceny pola powierzchni
Kwadrat o boku zorientowanym poziomo	0
Kwadrat o boku zorientowanym pod kątem 45°	+20%
Koło	+1,9%
Trójkąt	+4,3%

Wartości błędów przedstawione w tabeli 12.3 nie stanowią podstawy do ogólnej oceny błędów szacowania pola powierzchni przez systemy analizy obrazu. Ich celem jest jedynie zwrócenie uwagi na istniejący problem. Wśród przedstawionych na il. 12.14 figur warto zwrócić uwagę na kwadrat o boku zorientowanym pod kątem 45°, gdyż w jego przypadku otrzymaliśmy wyjątkowo duży błąd oceny pola powierzchni, wynoszący aż 20%. Wynika to faktu, że idealny brzeg tego kwadratu przebiega wzdłuż przekątnej pikseli traktowanych jako małe kwadraty. Z tego powodu cyfrowa reprezentacja kwadratu, pokazana na il. 12.14, daje znaczne zawyżenie pola powierzchni.

Oczywiście, można zrobić tak jak na il. 12.15b, czyli do budowy cyfrowej wersji kwadratu zaliczyć tylko te piksele, które w całości zawarte są w modelowym kwadracie

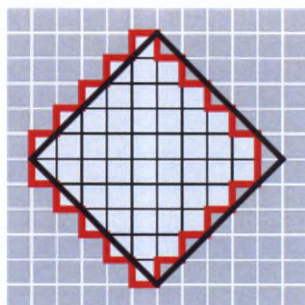


Il. 12.15. Różne sposoby wykonania cyfrowej reprezentacji kwadratu oraz wynikające z nich błędy oceny pola powierzchni

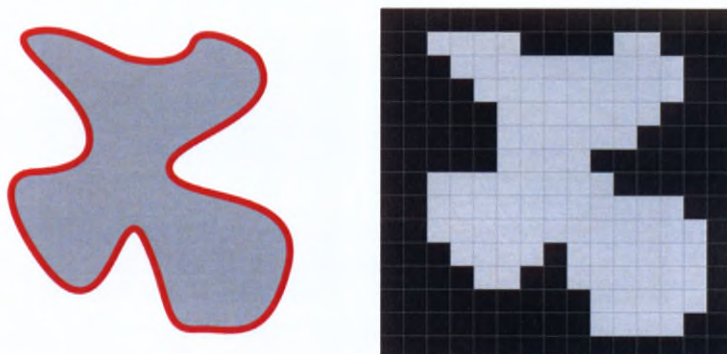
zaznaczonym czarną linią. Wtedy pole powierzchni będzie zaniżone o 16%. Błąd jest mniejszy, ale to nie oznacza, że taki sposób wyboru pikseli jest lepszy. Jeżeli mielibyśmy do czynienia z kwadratem o boku długości $\sqrt{2}$, to oba sposoby dałyby błędy na poziomie 100%. W pierwszym wariacie pole powierzchni wynosiłoby 4, w drugim byłoby to 0, a rzeczywiste pole powierzchni wynosiłoby 2.

Gdybyśmy do figury zaliczali co drugi piksel przecięty przez brzeg, to mielibyśmy dwa warianty, pokazane na il. 12.15 c i 12.15d. W obu przypadkach bezwzględna wartość błędu oceny pola powierzchni wynosiłaby 4%, przy czym raz wynik byłby zawyżony, a raz zaniżony.

Można też przyjąć wariant przybliżenia figury w sposób pokazany na il. 12.16. Przecięte przez modelową figurę piksele z jej lewej strony dodajemy do cyfrowej reprezentacji, a z prawej strony je odrzucamy. To, czy taki sposób postępowania jest najlepszy, jest co najmniej dyskusyjne, ale w tym przypadku nie wystąpi błąd oceny pola powierzchni. Omawiany przypadek kwadratu o boku zorientowanym pod kątem 45° jest wyjątkowo



Il. 12.16. Cyfrowa reprezentacja kwadratu, która daje bezbłędny wynik pomiaru pola powierzchni



Il. 12.17. Obiekt i jego reprezentacja w cyfrowym obrazie

niekorzystny do oceny pola powierzchni. W przypadku innych figur kryterium wyboru pikseli jest proste: modelowany kwadratem piksel zaliczamy do cyfrowej wersji figury, jeżeli jego wspólna część z przybliżoną figurą ma pole powierzchni równe co najmniej 50% pola powierzchni kwadratu modelującego dany piksel. Taki sposób działania jest w przybliżeniu realizowany podczas binaryzacji, gdy kryterium zaliczania do obiektu jest ustalone na poziomie średniej wartości ciemnych i jasnych punktów w okolicy brzegu obiektu. Problem ten był analizowany podczas omawiania il. 12.9.

Do podstawowych i często wykorzystywanych parametrów charakteryzujących geometrię obiektów, pokazanych schematycznie na il. 12.2, należy obwód. Niestety, jest to parametr trudny do zmierzenia. Proces jego wyznaczania będziemy analizowali na podstawie szacowania obwodu obiektu pokazanego na il. 12.17, gdzie z prawej strony znajduje się cyfrowy zapis tego obiektu.

Najprostszą formą wyznaczania obwodu, zwykle występującą w systemach analizy obrazu pod nazwą PERIMETER, jest po prostu traktowanie punktów tworzących brzeg obrazu jako kwadraty o boku jednostkowej długości, a następnie ich zliczanie.

Taka metoda daje dokładny wynik dla kwadratu, którego boki są zorientowane poziomo i pionowo. W przypadku figury pokazanej na il. 12.14 tak wyznaczany obwód będzie wynosił 70, jeżeli jako jednostkę przyjmujemy piksel.

Konsekwencją przyjęcia tego sposobu wyznaczania obwodu jest to, że cztery różne figury, pokazane na il. 12.14, będą miały identyczne obwody o długości wynoszącej 40 pikseli. Jest to oczywiście błędne. Analityczne wyznaczenie obwodu dla figur z il. 12.14 nie sprawia trudności i można łatwo wyznaczyć wartości błędów szacowania obwodu za pomocą parametru PERIMETER. Względne wartości tych błędów przedstawia tabela 12.4. Parametr PERIMETER zawyża rzeczywisty obwód nawet o ponad 40%. Warto przy tym zwrócić uwagę, że ten błąd nie zależy od wielkości badanego obiektu. Tabela 12.4 zawiera też błędy oceny obwodu za pomocą metody Croftona, która jest omówiona w dalszej części rozdziału.

Tabela 12.4. Wybrane błędy oceny obwodu za pomocą parametru PERIMETER oraz CROFTON PERIMETER dla figur pokazanych na il. 12.14

Figura	Błąd oceny obwodu	
	PERIMETER	CROFTON
Kwadrat o boku zorientowanym poziomo	0	-8%
Kwadrat o boku zorientowanym pod kątem 45°	+41,4%	-1,3% (max. 5,2%)
Koło	+27,3%	+3%
Trójkąt	+22,8%	-5,8%

Ocena obwodu nieregularnych figur nie jest łatwa. Rozwiązanie tego problemu zawdzięczamy pracom A.L. Cauchy'ego i M. Croftona, którzy znaleźli je jeszcze w XIX wieku, a więc na długo przed wynalezieniem komputerów. Jeżeli analizujemy dowolną, wypukłą figurę (il. 12.18a), to możemy na podstawie długości rzutów wykonywanych pod różnymi kątami, zmieniającymi równomiernie dla wszystkich możliwych orientacji, wyliczyć średnią długość tego rzutu:

$$\bar{d} = \sum_{i=1}^n d_i \quad (12.2)$$

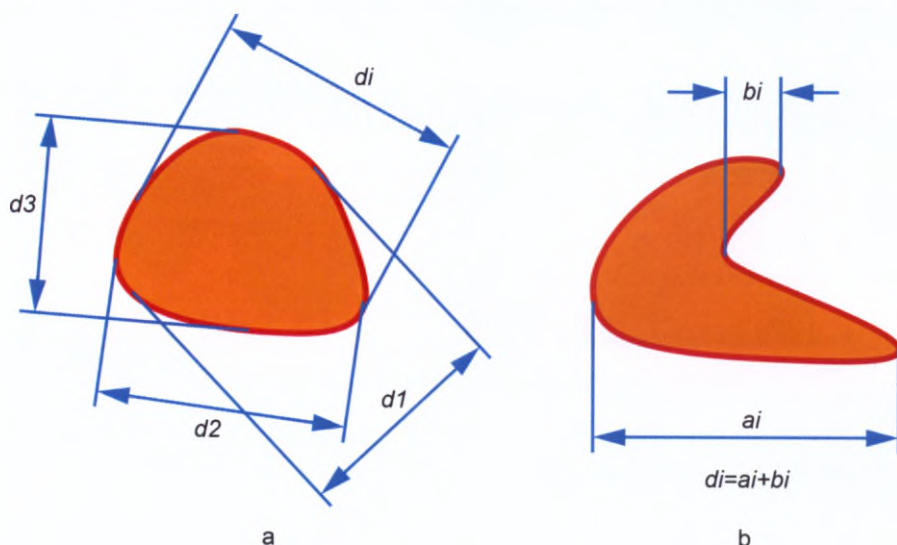
a następnie obwód L tej figury:

$$L = \pi \bar{d} \quad (12.3)$$

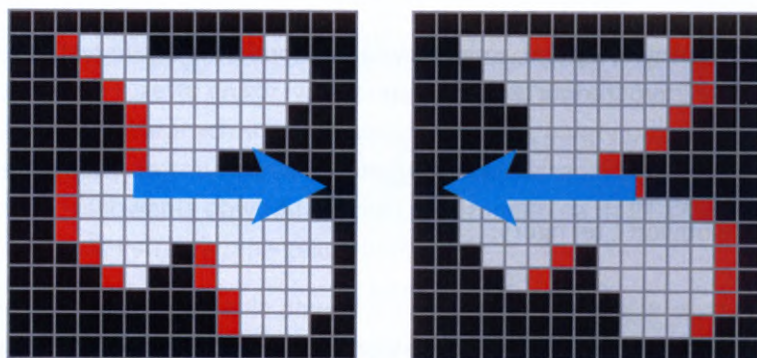
W przypadku figury wklęsłej wykorzystujemy tzw. rzut całkowity, którego ideę pokazuje il. 12.18b.

W przypadku cyfrowego obrazu wyznaczenie długości rzutu jest wbrew pozorom bardzo proste. Omówimy to na podstawie rzutu pionowego (il. 12.19). Całkowity rzut pionowy w przypadku patrzenia na figurę z lewej strony (lewa część il. 12.19) będzie po prostu liczbą punktów, które na obrazie figury zaznaczono czerwonym kolorem. Wystarczy w tym celu policzyć, ile punktów obrazu wykazuje konfigurację 01, czyli taką, w której z dwóch sąsiednich punktów ten z lewej ma wartość 0, a ten z prawej ma wartość 1. Jak łatwo zauważyć na il. 12.19, w przypadku próby wyznaczania rzutu od prawej strony figury otrzymamy identyczny wynik.

Oczywiście w przypadku obrazu z kwadratową matrycą pikseli nie damy rady badać długości rzutów pod wszystkimi kątami. Ograniczamy się do czterech kierunków:



Il. 12.18. Wyznaczanie obwodu nieregularnej figury. Opis w tekście



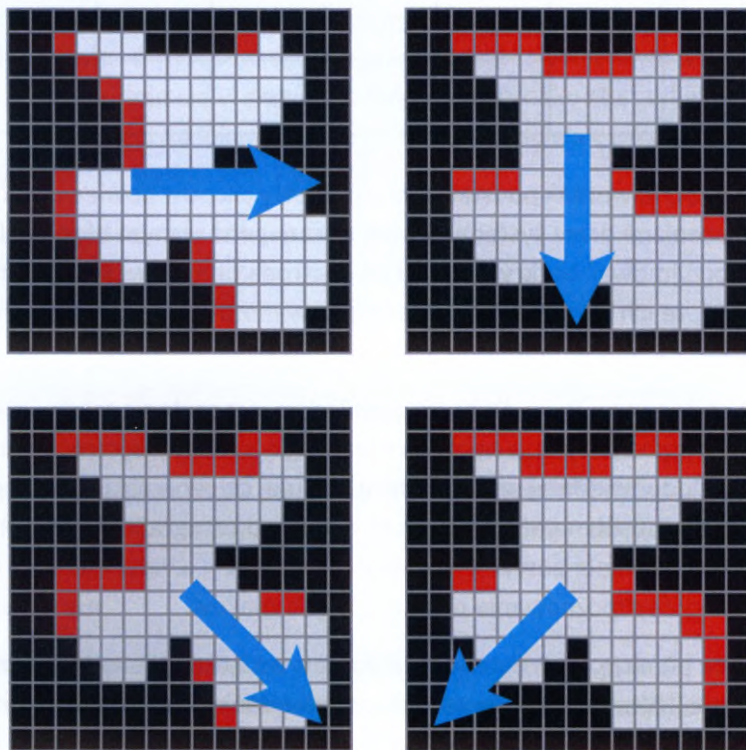
Il. 12.19. Wyznaczanie długości rzutu pionowego. Opis w tekście

0° , 45° , 90° i 135° (il. 12.20). Rzuty w kierunku poziomym i pionowym są bardzo łatwe do wyznaczenia (patrz omówienie il. 12.19). Jeżeli chcemy wyznaczyć długość rzutu pod kątem 45° lub 135° , to musimy wziąć pod uwagę, że poszczególne linie rzutowania są zagęszczone, bo są odległe nie o jeden piksel, tylko o $\sqrt{2}/2$. Ponieważ mamy cztery kierunki rzutowania, to do obliczania obwodu wykorzystamy średnią z czterech pomiarów. Jeżeli długość rzutu dla kierunku x° oznaczmy przez L_{x° , to obwód L badanej figury możemy wyznaczyć zgodnie ze wzorem:

$$L = \frac{\pi}{4} \left[L_0 + L_{90} + \frac{1}{\sqrt{2}} (L_{45} + L_{135}) \right] \quad (12.4)$$

Zależność tę nazywamy wzorem lub regułą Croftona, a obwód wyznaczany z jej wykorzystaniem jest najczęściej opisywany jako CROFTON PERIMETER.

Wyznaczenie obwodu zgodnie z regułą Croftona też nie jest pozbawione błędów, ale błędy te są mniejsze niż pokazane w tabeli 12.4. Dla kwadratu obróconego o kąt 45° metoda Croftona daje różne błędy, ale nigdy nie przekraczają one 5,2%, co zaznaczono w tabeli 12.4.



Il. 12.20. Wyznaczanie rzutów figury pod różnymi kątami. Opis w tekście

Najistotniejszym wnioskiem płynącym z dotychczasowych analiz jest to, że metoda Croftona daje zazwyczaj nawet kilkukrotnie mniejsze błędy oceny obwodu obiektu niż parametr PERIMETER. Wyjątkiem jest kwadrat o boku zorientowanym poziomo, gdzie PERIMETER daje dokładną wartość obwodu. Obwód wyznaczony metodą Croftona może być w tym przypadku zaniżony nawet o blisko 20% (tab. 12.5). Ten duży błąd odpowiada jednak ocenie bardzo małego obiektu, dla którego nie powinno się szacować parametrów geometrycznych, gdyż cyfrowa reprezentacja koła i kwadratu jest w tym przypadku identyczna. Warto zwrócić uwagę, że wraz ze zwiększaniem się obiektu błędy metody Croftona maleją (tab. 12.5), podczas gdy błędy oceny za pomocą parametru PERIMETER pozostają stałe. Błąd oceny obwodu dla bardzo dużego kwadratu o boku długości 1000 pikseli zmaleje w przypadku metody Croftona do -2,1%. Oczywiście mówimy o zmniejszaniu wartości bezwzględnej błędu.

Tabela 12.5. Względne błędy wyznaczania obwodu metodą Croftona dla kwadratu o boku zorientowanym poziomo i dla koła

	Długość średnicy koła lub boku kwadratu							
	2	3	4	5	10	20	50	100
	-19,1%	-14,5%	-12,1%	-10,7%	-8,0%	-6,6%	-5,7%	-5,5%
	3,0%	8,9%	-5,8%	-0,5%	3,0%	-2,3%	0,2%	0,03%

Podsumowując, nawet tak podstawowe parametry jak pole powierzchni oraz obwód mogą być oceniane przez systemy analizy obrazu ze znacznym błędem. Dlatego nie należy ich przyjmować bezkrytycznie. Dobrą stroną pomiarów komputerowych jest natomiast to, że ich wyniki są powtarzalne.

Zajmiemy się teraz rozpoznawaniem i oceną kształtu. Jest ona niezwykle ważna w naszym codziennym życiu. Dobrym przykładem jest czytanie tekstu, również tego. Do prawidłowego odczytania tekstu konieczne są co najmniej trzy elementy:

- przyjęcie odpowiedniej umowy, która opisuje, co oznaczają poszczególne znaki oraz w jaki sposób należy ich używać, np. zasad ortografii lub interpunkcji,
- znajomość opisanej powyżej umowy,
- umiejętność rozpoznawania poszczególnych znaków.

Jeżeli wydaje nam się, że drugi z wymienionych punktów jest mało istotny, to jesteśmy w błędzie. Nie można umowy dotyczącej znaczenia poszczególnych znaków zapisać tylko za jej pomocą, bo nie odczytamy tego zapisu bez jej znajomości! Tłumaczy to dobrze problemy z odczytywaniem przez nas chińskiego pisma. Mamy wprawdzie



II. 12.21. Proste kształty mogą zawierać znaczną ilość informacji. Opis w tekście

dostęp do pewnej informacji, ale pozostaje ona bezużyteczna, jeśli nie znamy sposobu jej kodowania.

Jeżeli znamy odpowiednią umowę, to nawet proste kształty mogą nieść ze sobą znaczną ilość informacji. Przykładem może być znak drogowy pokazany na il. 12.21. Mamy tu dwa kształty, które kodują dwie porcje informacji. Tymi kształtami są: trójkąt i zawarty w środku czarny zygzak, przypominający leżącą literę S.

Pierwszy z tych kształtów przekazuje nam dwie informacje:

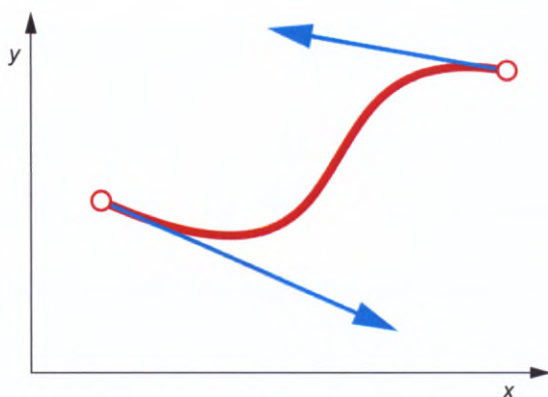
- W pobliżu jest potencjalnie niebezpieczne miejsce (wynika to z faktu, że znak ma kształt trójkąta równobocznego z wierzchołkiem skierowanym do góry).
- Miejsce to znajduje się w pewnej odległości od znaku. Ta odległość może być różna w zależności od lokalizacji znaku. Jeżeli znajduje się on poza obszarem zabudowanym, to wynosi ona 150-300 m.

Drugi kształt przynosi nam kolejne informacje:

- Tym niebezpiecznym miejscem są dwa zakręty.
- Pierwszy z nich jest zakrętem w lewo.
- Drugi jest oddalony od pierwszego nie więcej niż 300 m.
- Drugi zakręt może być w prawo lub w lewo – zatem trzeba uważać.

Tak więc bardzo prosty element graficzny niesie ze sobą zaskakująco wiele informacji. Ich przekazanie odbywa się bardzo szybko, o wiele szybciej niż przeczytanie tych wiadomości zakodowanych za pomocą pisma. Nic zatem dziwnego, że rozpoznawanie i ocena kształtu jest problemem, z którym musi się zmierzyć komputerowa analiza obrazu.

Kształt jest trudnym do scharakteryzowania parametrem. Zaczyna się od problemów z jego zdefiniowaniem. Niektóre poważne źródła, np. encyklopedia.pwn.pl, w ogóle nie zawierają hasła „kształt”. Poszczególne źródła różnie definiują kształt i zawsze można mieć do tych definicji pewne zastrzeżenia, co tylko dowodzi trudności w ich



Il. 12.22. Najprostszy odcinek krzywej Béziera. Opis w tekście

formułowaniu. Wydaje się, że najlepszą, dość rozbudowaną definicję podaje wikipedia.org, choć początkowa część definicji opisuje kształt w sposób pośredni. Według tego źródła:

„W geometrii przyjmuje się, że dwa podzbiory przestrzeni euklidesowej posiadają taki sam kształt, jeżeli jeden może być przemieniony w drugi za pomocą kombinacji przesunięć, obrotów i równomiernego skalowania”. Innymi słowy, kształt zbioru punktów jest całością dotyczącą tego zbioru geometrycznej informacji, która jest niezmienna względem przesunięć, obrotów oraz zmian wielkości. Proste? Niezupełnie.

Zdefiniowanie kształtu, niezależnie od jego trudności, nie przybliży rozwiązania problemów dotyczących jego oceny. Wynika to z faktu, że nawet najprostszy krzywoliniowy element – np. narysowany ołówkiem – może być wykonany na nieskończenie wiele sposobów i z tego powodu nie da się go jednoznacznie opisać pojedynczą liczbą.

Il. 12.22 przedstawia najprostszy odcinek krzywej Béziera, która jest popularnym narzędziem do rysowania w programach graficznych. Omawiany fragment krzywej jest zdefiniowany za pomocą dwóch punktów, nazywanych węzłami, które zaznaczono na il. 12.22 w postaci białych kółek z czerwoną obwódką. Przebieg krzywej pomiędzy tymi węzłami zależy od kierunku stycznych oraz promieni krzywizny w obu węzłach. Reprezentują to widoczne na il. 12.22 strzałki. Jeżeli ten odcinek krzywej umieścimy w układzie współrzędnych x - y , to do jego jednoznacznego opisania potrzebujemy sześciu liczb: odległości między węzłami w kierunku x oraz y , kątów pomiędzy styczną i osią x w obu węzłach oraz promieni krzywizny w obu węzłach. Jeżeli obrócimy ten kształt tak, aby oba węzły znajdowały się w tej samej odległości od osi x (obróć figurę nie zmieniając jej kształtu), to wystarczy podać odległość między węzłami, dwa kąty i dwa promienie. Opis odcinka krzywej możemy zatem zredukować do pięciu liczb. I na tym koniec.

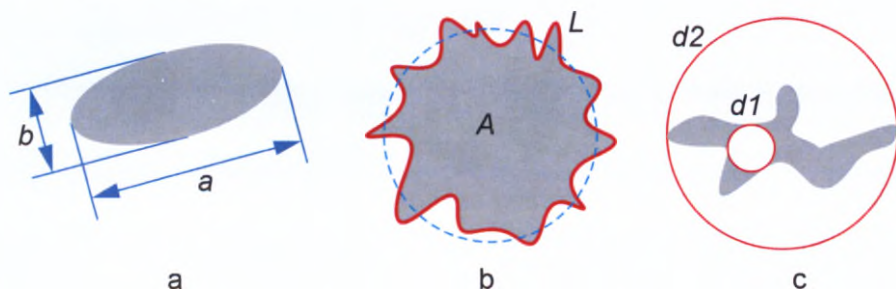
Wniosek: nawet najprostsze kształty wymagają co najmniej kilku liczb do jednoznacznego opisu. Zatem nie jest możliwe takie scharakteryzowanie kształtu, jak to występuje w przypadku wielkości, gdzie często wystarczy podanie jednego parametru, np. pola powierzchni.

Wymyślono dwa podstawowe rozwiązania tego problemu:

- wykorzystanie sztucznej inteligencji (zwłaszcza sieci neuronowych) do rozpoznawania i klasyfikowania kształtów,
- wyznaczanie wartości współczynników kształtu.

Pierwsze rozwiązanie jest całkiem skuteczne i wykorzystują je np. systemy rozpoznawania pisma (OCR – *Optical Character Recognition*) lub odczytywania tablic rejestracyjnych. Możemy przypisać obiekt do pewnej klasy, ale nie ocenimy, czy zmienił się znacznie, czy może tylko niewiele? Inaczej mówiąc, nie zmierzmy w ten sposób zmian kształtu, a może to być potrzebne, np. w diagnostyce medycznej, gdzie zmiany kształtu komórek lub ich kolonii mogą być sygnałem pogłębiania się procesu chorobowego. Problem braku miary zmian kształtu można przynajmniej częściowo rozwiązać przez wprowadzenie odpowiednio dużej liczby wzorców, ale może to czasami utrudnić prawidłową klasyfikację.

Drugie rozwiązanie to pewna proteza, półśrodek, którego używamy, aby pokonać ograniczenia wynikające ze złożoności kształtu i braku możliwości jego zmierzenia. Współczynniki kształtu są bezwymiarowymi (bo muszą być niezależne od wielkości) parametrami, które opisują stopień zgodności z idealnym lub modelowym kształtem, najczęściej kołem.



Il. 12.23. Wielkości używane do wyznaczania współczynników kształtu. Opis w tekście

Najprostszym przypadkiem odstępstwa od koła jest spłaszczone koło, czyli elipsa (il. 12.23a). Ten (lub zbliżony do niego) kształt można opisać za pomocą prostego współczynnika charakteryzującego wydłużenie, który jest stosunkiem najdłuższego rzutu do wymiaru w kierunku do niego prostopadłym:

$$k_1 = \frac{a}{b} \quad (12.5)$$

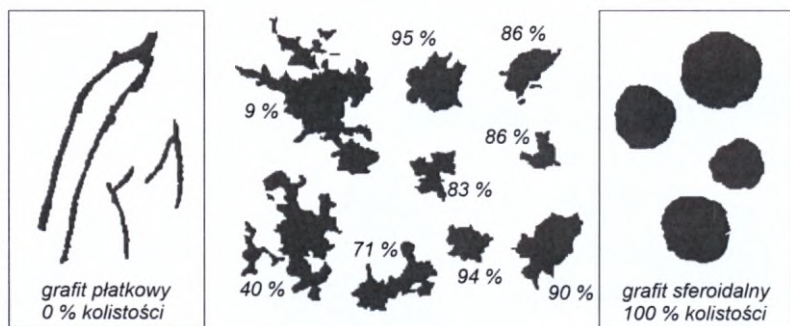
Ten współczynnik będzie jednak zupełnie bezużyteczny, jeśli obiekt nie wykazuje wydłużenia, a jedynie poszarpany brzeg (il. 12.23b). Wtedy można wykorzystać pole powierzchni A oraz obwód L do budowy współczynnika kształtu:

$$k_2 = \frac{4\pi A}{L^2} \quad (12.6)$$

Współczynnik ten przyjmuje wartość 1 dla koła i wartości mniejsze od 1 dla pozostałych kształtów. Im bardziej kształt jest nieregularny, tym mniejszą wartość wykazuje współczynnik k_2 . Długość obwodu jest podnoszona do kwadratu po to, aby zachować niezmiennosc wartości współczynnika przy zmianie wielkości obiektu.

Może się zdarzyć tak, że obiekt będzie wydłużony i będzie posiadał postrzępione brzegi, czyli będzie łączył te cechy, na które były wrażliwe współczynniki zdefiniowane do tej pory (il. 12.23c). W takiej sytuacji można próbować charakteryzować kształt za pomocą stosunku średnic koła opisanego na obiekcie i wpisanego do niego:

$$k_3 = \frac{d_2}{d_1} \quad (12.7)$$



Il. 12.24. Ocena kształtu wydzieliń grafitu. Opis w tekście

Il. 12.24 pokazuje klasyfikację wydzieliń grafitu w żeliwie. Jak widać, pomimo braku możliwości ilościowej oceny kształtu, wykorzystanie współczynnika k_3 , pozwoliło wydzielić trzy grupy obiektów, a nawet ilościowo scharakteryzować stopień przynależności do skrajnych klas.

W systemach komputerowej analizy obrazu często mamy możliwość wyznaczenia wartości kilku predefiniowanych współczynników kształtu:

- **BRFillRatio** (*Bounding Rectangle Fill Ratio*) – stosunek pola powierzchni obiektu do pola powierzchni opisanego na nim prostokąta, którego boki są równoległe do krawędzi obrazu. Można łatwo wyznaczyć jego wartość dzieląc pole powierzchni przez długości poziomej i pionowej średnicy Fereta. Osiąga on wartość 1 dla prostokąta o jednym z boków zorientowanym poziomo, a dla pozostałych figur wartości te są mniejsze od 1. Najniższe wartości dają figury wklęsłe. Wadą tego współczynnika są zmiany jego wartości dla różnej orientacji tego samego kształtu.
- **BoundingBoxToPerimeter** – stosunek obwodów prostokąta opisanego powyżej i obwodu obiektu mierzonego parametrem *Perimeter*. Dla wszystkich figur wypukłych (kwadrat, koło, trójkąt itp.) jego wartość wynosi 1. Dla obiektów wklęsłych jego wartość jest mniejsza od 1. Im bardziej rozbudowany jest brzeg obiektu, tym mniejsza jest wartość tego współczynnika.
- **LogOfHeightToWidth** – współczynnik nieco zbliżony do współczynnika k_1 , przy czym bierze się pod uwagę stosunek szerokości (mierzonej w kierunku poziomym) do wysokości (mierzonej w kierunku pionowym) obiektu. Wartość współczynnika jest równa logarytmowi dziesiętnemu stosunku wymiarów.
- **Compactness** (ściskość lub zwartość):

$$\text{Compactness} = \frac{16 \cdot (\text{pole powierzchni})}{(\text{perimeter})^2} \quad (12.8)$$

Dla kwadratu o poziomej orientacji ten współczynnik przyjmie wartość 1, a dla koła 0,785.

- **Circularity** (kołowość lub kolistość) – jest to współczynnik odpowiadający współczynnikowi kształtu k_2 . Obwód dla tego współczynnika wyznacza się metodą Croftona.
- **Convexity** (wypukłość) – jest to stosunek pola powierzchni obiektu do pola powierzchni jego wypukłego otoczenia. Osiąga maksymalną wartość 1 dla każdego wypukłego obiektu.
- **Elongation** (wydłużenie) wykorzystuje te same parametry, czyli długość najdłuższego rzutu a oraz długość rzutu w kierunku prostopadłym b , co współczynnik k_1 (il. 12.23a), przy czym jest liczone nieco inaczej:

$$\text{Elongation} = \frac{a - b}{a + b} \quad (12.9)$$

i osiąga wartość 0 dla obiektu niewykazującego wydłużenia. Dla obiektu o stosunku długości do szerokości 2:1 osiąga wartość 0,6, dla stosunku 3:1 – wartość 0,8, zaś dla stosunku 4:1 – wartość 0,88.



Il. 12.25. Szkielety wybranych figur (powtórzenie ilustracji z rozdziału poświęconego operacjom morfologicznym)

Można zdefiniować praktycznie dowolną liczbę współczynników kształtu, przy czym zawsze będą one spełniały swoją rolę tylko w odniesieniu do ograniczonej klasy obiektów. Bywają również budowane półautomatyczne klasyfikatory obiektów, które wykorzystują logikę rozmytą, a kryterium oceny kształtu jest kombinacja liniowa wartości wielu współczynników kształtu, branych pod uwagę z różnymi wagami.

Do oceny kształtu można również wykorzystać szkielety obiektów (il. 12.25). Jak wiemy z rozdziału poświęconego operacjom morfologicznym, szkielety matematyczne są zbiorami środków wszystkich kół wpisanych w figurę, opierających się co najmniej dwoma punktami o brzeg figury. W ten sposób (il. 12.25) szkieletem koła jest punkt, szkieletem kwadratu jego przekątne, a szkieletem trójkąta dwusieczne jego kątów na odcinku od wierzchołka trójkąta do punktu przecięcia dwusiecznych. Szkieletem pierścienia jest okrąg. Metody analizy obrazu pozwalają na uzyskanie dobrych przybliżeń szkieletów matematycznych. Wyjątkiem jest koło, dla którego zamiast pojedynczego punktu otrzymujemy szkielet przypominający literę X.

Przypomnijmy niektóre własności szkieletów:

- Szkielet figury leży zawsze w jej wnętrzu.
- Szkielet składa się z linii i nie ma pola powierzchni.
- Jeżeli w figurze występują otwory, to szkielet utworzy tyle zamkniętych pętli, ile jest otworów.
- Im bardziej postrzępiony jest brzeg figury, tym więcej gałęzi znajduje się w szkielecie.

W szczególności dwie ostatnie cechy mogą być przydatne do oceny kształtu.

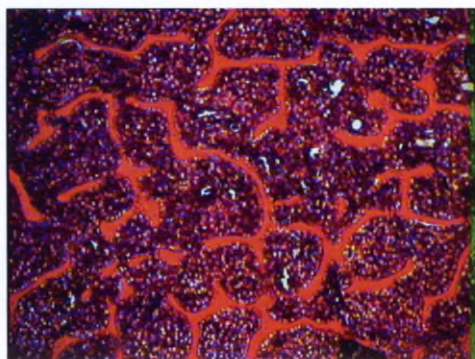
Ponieważ linie jest łatwiej analizować niż całe figury, możemy dokonywać pewnych pomiarów na szkieletach. Można wyznaczyć m.in. takie charakterystyki: liczba węzłów, liczba punktów końcowych, liczba gałęzi, liczba zamkniętych pętli czy średnia długość gałęzi. Niektóre parametry są ze sobą związane, np.

$$(\text{liczba gałęzi}) = \frac{3 \cdot (\text{1. węzłów}) + (\text{1. punktów końcowych})}{2} \quad (12.10)$$

gdzie słowo „liczba” skrócono do jednej litery z kropką: l .

Przykładem wykorzystania szkieletów może być ocena beleczek kostnych w kręgosłupie (il. 12.26a). Po wczytaniu obrazu (il. 12.26a) dokonano detekcji beleczek kostnych (il. 12.26b), a następnie na podstawie analizy szkieletów (il. 12.26c) oceniono stopień rozgałęzienia struktury beleczkowej. Otrzymane wyniki pozwoliły na wykazanie, że bardziej rozgałęziona struktura beleczek kostnych wpływa na zwiększenie wytrzymałości kręgosłupa na ściskanie.

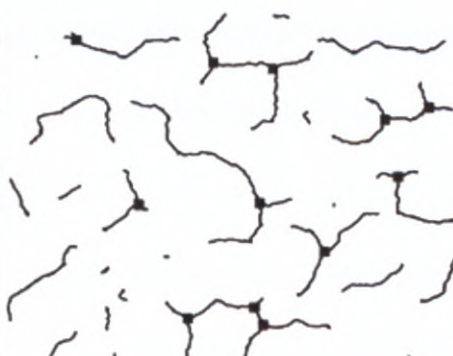
Wspomnienie szkieletów jako narzędzia do oceny kształtu jest dobrą okazją do omówienia pewnej metody wyznaczania liczby obiektów na podstawie obrazu granic między nimi, który możemy uznać za szkielet. Przykład takiego szkieletu pokazuje il. 12.27. Założmy, że chcemy wyznaczyć liczbę ziaren N na podstawie znajomości



a



b



c

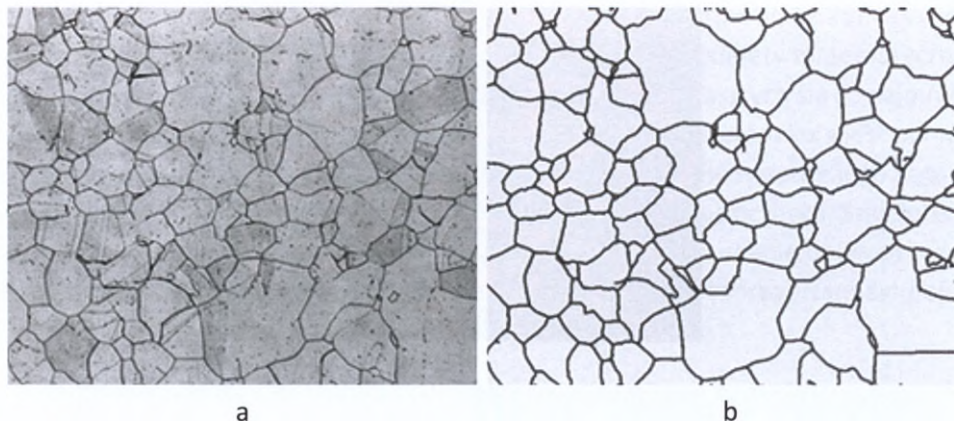
Il. 12.26. Wykorzystanie szkieletów do charakteryzowania kształtu beleczek kostnych.
Opis w tekście

liczby węzłów (czyli miejsc, gdzie spotyka się kilka krawędzi) szkieletu W . Do przeprowadzenia tej analizy będą potrzebne dwa założenia:

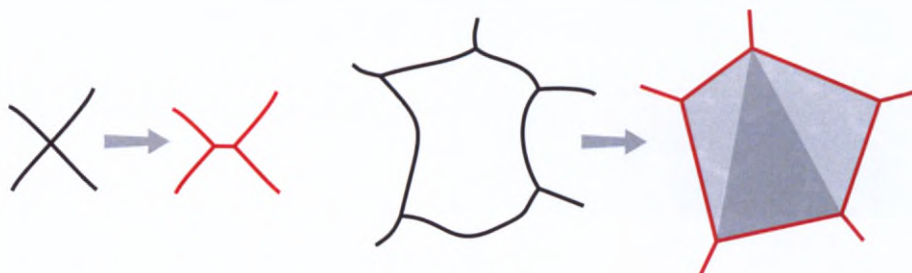
- Wszystkie figury w szkielecie są wielokątami wypukłymi.
- W jednym węźle spotykają się trzy linie.

Powyższe założenia nie mają żadnego wpływu na liczbę obiektów, natomiast mają znaczenie dla wyprowadzenia interesującej nas zależności.

W praktyce nigdy na przekroju struktury komórkowej nie spotkają się cztery obiekty. Jeżeli jednak z powodu niskiej rozdzielczości obrazu mamy wrażenie, że w jednym węźle spotykają się cztery obiekty, to należy przyjąć, że są to dwa węzły położone blisko siebie (lewa strona il. 12.28). Z kolei dowolny kształt ziarna może być przetworzony bez zmiany liczby węzłów i ziaren w wielokąt wypukły (prawa strona il. 12.28). Ta wypukłość wielokątów jest potrzebna do tego, aby móc je dzielić na trójkąty. I tak każdy czworokąt można podzielić na dwa trójkąty, pięciokąt na trzy, sześciokąt na cztery itd.



Il. 12.27. Obraz struktury ziarnistej (a) oraz szkielet obrazu granic ziaren (b)



Il. 12.28. Założenia do metody punktów węzłowych. Opis w tekście

Zakładamy, że nasz zbiór obiektów składa się z n_3 trójkątów, n_4 czworokątów, n_5 pięciokątów itd. Jeśli tak, to sumaryczna liczba obiektów wynosi:

$$N = n_3 + n_4 + n_5 + \dots \quad (12.11)$$

Każdy wielokąt, posiadający k kątów, można podzielić na $k-2$ trójkątów. Ponieważ suma kątów trójkąta wynosi π , to suma wszystkich kątów S w wielokątach wynosi:

$$S = n_3(3-2)\pi + n_4(4-2)\pi + n_5(5-2)\pi + n_6(6-2)\pi + \dots \quad (12.12)$$

Z drugiej strony wszystkie kąty spotykają się w węzłach i tworzą w każdym węźle kąt pełny, czyli 2π . Zatem można sumę wszystkich kątów wyrazić jako:

$$S = 2W\pi \quad (12.13)$$

Porównując równania (12.12) i (12.13), otrzymujemy:

$$2W\pi = n_3(3-2)\pi + n_4(4-2)\pi + n_5(5-2)\pi + n_6(6-2)\pi + \dots \quad (12.14)$$

Równanie to można zapisać nieco inaczej:

$$2W\pi = (3 \cdot n_3 + 4 \cdot n_4 + 5 \cdot n_5 + 6 \cdot n_6 + \dots)\pi - 2(n_3 + n_4 + n_5 + n_6 + \dots)\pi \quad (12.15)$$

Możemy to równanie podzielić obustronnie przez π . Dodatkowo, suma $(3 \cdot n_3 + 4 \cdot n_4 + 5 \cdot n_5 + \dots)$ to liczba wszystkich kątów. Ponieważ w jednym węźle spotykają się trzy kąty, to możemy ich sumę wyrazić również jako $3W$. Z kolei druga suma zapisana w nawiasie to po prostu liczba wszystkich komórek N . Nasze równanie przyjmuje zatem postać:

$$2W = 3W - 2N \quad (12.16)$$

Po elementarnych przekształceniach otrzymujemy końcową postać poszukiwanej zależności:

$$N = 0,5 \cdot W \quad (12.17)$$

Czyli liczba komórek jest równa połowie liczby węzłów. Poprawność tej zależności można łatwo sprawdzić, rysując sobie dowolną strukturę komórkową i porównując liczby obiektów wyznaczonych metodą Jeffriesa oraz przedstawioną powyżej metodą punktów węzłowych.

Obecnie, w epoce narzędzi komputerowych, metoda ta praktycznie nie jest używana. Została jednak opisana przy okazji pomiarów, gdyż pokazuje, jak dzięki prostym, ale sprytnym rozwiązaniom można rozwiązać niektóre problemy. Granice struktur komórkowych (il. 12.27a) powstają w wyniku trawienia wypolerowanej powierzchni.

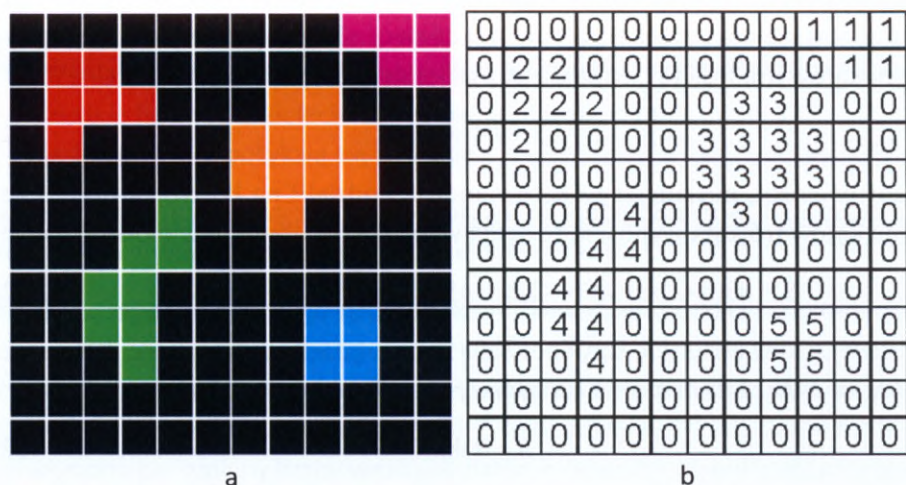


Fig. 12.29. Obraz etykietowany: a) sposób wyświetlania, b) wartości w pamięci komputera

Czasami z powodu różnych trudności nie wszystkie granice są dobrze wytrawione. Ale na ogół dobrze widoczne są punkty węzłowe, gdyż w tych właśnie miejscach odporność chemiczna materiału jest słaba. Dlatego, zamiast męczyć się z wyrafinowanymi algorytmami uzupełniającymi brakujące granice pomiędzy obiektami, można w takiej sytuacji wykorzystać metodę punktów węzłowych.

Wróćmy na chwilę do obrazów etykietowanych, które były omawiane w ramach binaryzacji. Taki obraz jest wyświetlany jako mozaika kolorowych plam (il. 12.29a), ale każda z nich odpowiada jednej liczbie i są one pamiętane tak, jak pokazuje to il. 12.29b. Pierwszy punkt obrazu, w lewym górnym rogu, otrzymuje etykietę 0 i jest ona przydzielana każdemu następnemu punktowi, gdy poruszamy się w prawo, a po zakończeniu wiersza przechodzimy o jeden piksel w dół i znowu przesuwamy się od lewej do prawej. Gdy natrafiamy na pierwszy punkt pierwszego obiektu, dajemy mu etykietę 1. Taką samą etykietę otrzymują wszystkie piksele należące do tego obiektu, zgodnie z zasadami spójności (obiekty mogą być czterospójne lub ośmiospójne). Gdy napotykamy nowy obiekt, przypisujemy mu kolejne wartości etykiet: 2, 3, 4 itd. Istotna informacja: wszystkie piksele należące do tego samego obiektu mają identyczną etykietę. Prowadzimy ten proces aż do osiągnięcia ostatniego piksela.

Jakie ma to znaczenie dla pomiarów? Przede wszystkim ułatwia identyfikację obiektów, gdy potrzebujemy znać parametry każdego obiektu, a jednocześnie przyspiesza analizę. Wyobraźmy sobie, że potrzebujemy wyznaczyć rozkład powierzchni obiektów w obrazie z il. 12.29. Patrzymy, jaki jest zakres wartości, i widzimy, że największą wartością jest 5. Od razu wiemy, że mamy do czynienia z pięcioma obiektami.



Il. 12.30. Pomiary na obrazie kolorowym. Opis w tekście



Il. 12.31. Obraz zakłócony małymi obiektami, które są przede wszystkim szumem

Tworzymy zatem odpowiednią tabelę, zawierającą pięć wierszy i dwie kolumny. W pierwszej kolumnie będziemy mieli etykiety (od 1 do 5), a w drugiej pola powierzchni liczone w pikselach (początkowo ta kolumna będzie wypełniona zerami). Następnie przeglądamy po kolei w wierszach od góry do dołu wszystkie wartości pikseli. Jeżeli natrafimy na piksel z etykietą różną od zera, to w wierszu o numerze równym etykietce zwiększamy wartość w kolumnie pola powierzchni o 1. W ten sposób, po zaledwie jednokrotnym przejściu przez wszystkie wartości w obrazie, otrzymamy gotowy rozkład pól powierzchni wszystkich obiektów. Zazwyczaj wykorzystanie etykietyzacji podczas pomiarów nie jest widoczne dla użytkownika systemu analizy obrazu, ale obrazy tego typu są analizowane w tle.

Zwykle mówi się, że do pomiarów nadają się tylko obrazy binarne. To prawda, zwłaszcza w przypadku pomiarów wielkości geometrycznych. Czasami można jednak wykonać pewne pomiary na obrazach szarych lub nawet kolorowych (il. 12.30).

Jeżeli zarejestrowaliśmy obraz znaku drogowego (il. 12.30a), to możemy go przetworzyć tak, aby uzyskać binarny obraz tła znaku (il. 12.30b). Jeżeli wykorzystamy go jako maskę do obrazu wyjściowego (il. 12.30c), to możemy np. zmierzyć, jaki kolor został użyty w tym znaku jako tło, i tym samym sprawdzić, czy zostało to zrobione poprawnie. Niniejszy przykład stanowi jedynie ilustrację pewnego sposobu analizy, który możemy zastosować, chcąc wykonać pomiary na obrazach szarych lub kolorowych. Użycie obrazów binarnych jako pomocniczego narzędzia będzie jednak niezbędne.

Na zakończenie kilka uwag o prezentacji wyników badań. Jeżeli analizujemy obiekty w obrazie (il. 12.31), to najczęściej rozkład wielkości przedstawiamy za pomocą histogramu ważonego liczbą obiektów. Inaczej mówiąc, grupujemy obiekty w pewne klasy wielkości i podajemy liczebność obiektów w każdej z tych klas. Dane do odpowiedniego histogramu na potrzeby obrazu z il. 12.31 zawiera tabela 12.6.

Tabela 12.6. Histogram pola powierzchni obiektów w obrazie z il. 12.31

pozioma średnica Fereta – górny kres klasy	15	30	45	60
liczba obiektów	1248	40	20	5

Nie ma szans na rysowanie jakiegokolwiek wykresu, gdyż pierwszy przedział klasowy zawiera ponad 30 razy więcej obiektów niż jakikolwiek inny przedział. Oczywiście, można odfiltrować te małe wydzielenia i rozkład miałby zupełnie inny charakter. Okazuje się jednak, że nie zawsze jest to potrzebne, tym bardziej że może to być odczytane jako nieuprawniona ingerencja w dane pomiarowe.



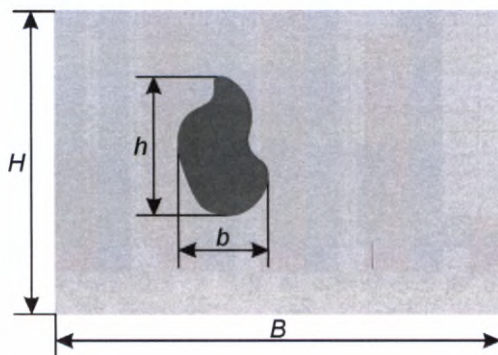
II. 12.32. Ważony powierzchnią rozkład wielkości (pola powierzchni) obiektów z il. 12.31

Wyobraźmy sobie, że prowadzimy zbiórkę pieniędzy. Kilkadziesiąt osób dało nam po przysłowiowej złotówce, ale jedna osoba przekazała aż 1000 zł. Jeśli przeanalizujemy dane, to wyjdzie nam, że tylko jedna osoba dała cokolwiek więcej i będzie to dla tej osoby oczywiście krzywdzące. Można przedstawić te same dane w postaci ważonej wielkością wpłaty. Wtedy te kilkadziesiąt osób da nam w sumie kilkadziesiąt złotych, co jest wielokrotnie mniejsze od tysiąca. Takie zestawienie zdecydowanie lepiej pokaże, którzy darczyńcy mieli większy wpływ na zebraną kwotę.

W przypadku badań materiałów zwykle okazuje się, że o jego cechach decydują obiekty największe, choć liczbowo jest ich dość mało. Dlatego często lepszym rozwiązaniem od histogramów ważonych liczebnością są histogramy ważne powierzchnią. Dane, które posłużyły do wypełnienia tabeli 12.6, można teraz przeliczyć. Wyniki przedstawia histogram na il. 12.32. Poszczególne słupki mówią nam, jaki ułamek powierzchni wszystkich obiektów stanowiły obiekty o średnicy Fereta z przedziałów: 0-15, 15-30, 30-45 oraz 45-60.

Podczas omawiania błędów, spowodowanych przecięciem części obiektów przez brzeg obrazu, wprowadziliśmy pojęcie ramki bezpieczeństwa, której wykorzystanie pozwala otrzymać rozkład obiektów pozbawiony systematycznego błędu. Pomimo skuteczności i poprawności działania, wykorzystanie ramki bezpieczeństwa może być trudne, zwłaszcza dla początkującego użytkownika, gdyż wymaga dodatkowych operacji związanych z filtrowaniem obiektów.

Możliwe jest jednak uzyskanie poprawnego rozkładu wielkości obiektów bez używania ramki bezpieczeństwa. Wystarczą nam następujące dane (il. 12.33): rozmiar obrazu w pikselach, czyli jego wysokość H oraz szerokość B , średnice Fereta dla każdego obiektu, wyrażone w pikselach h oraz b , a także wartość parametru, który ma być analizowany – może to być obwód, średnica, pole powierzchni itp.



Il. 12.33. Wielkości potrzebne do korekty Miles–Lantuéjoul. Opis w tekście

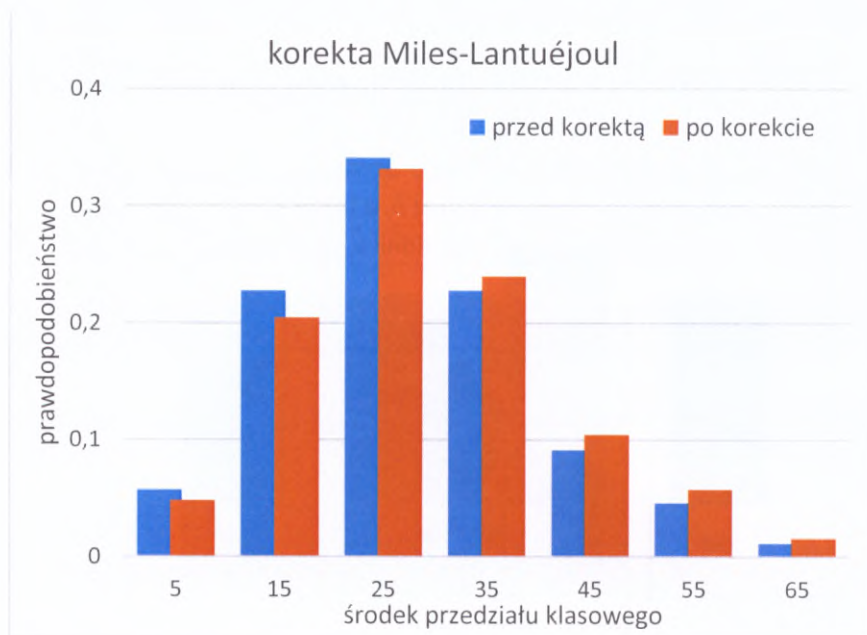
Odpowiednia procedura jest znana jako korekta Miles–Lantuéjoul, co upamiętnia nazwiska jej twórców. Cała procedura przebiega następująco:

- Przygotowujemy obraz binarny obiektów do analizy i za pomocą operacji *border kill* usuwamy wszystkie obiekty przecięte przez brzeg obrazu.
- Mierzymy parametry dla każdego obiektu, oznaczonego indeksem i : rozmiar obrazu w pikselach, czyli jego wysokość H oraz szerokość B , średnice Fereta dla każdego obiektu, wyrażone w pikselach h_i oraz b_i , a także wartość parametru, który ma być analizowany. Musimy również znać całkowitą liczbę obiektów N .
- Im większy jest obiekt, tym mniejsze prawdopodobieństwo $P(i)$, że nie zostanie on przecięty przez brzeg obrazu:

$$P(i) = \frac{(B - b_i) \cdot (H - h_i)}{B \cdot H} \quad (12.18)$$

W związku z tym dla każdego obiektu wyznacza się współczynnik korekcyjny $W(i)$, który kompensuje zwiększone ryzyko przecięcia dużych obiektów przez brzeg obrazu:

$$W(i) = \frac{B \cdot H}{(B - b_i) \cdot (H - h_i)} \cdot \frac{N}{\sum_{i=0}^{N-1} \frac{1}{P(i)}} \quad (12.19)$$



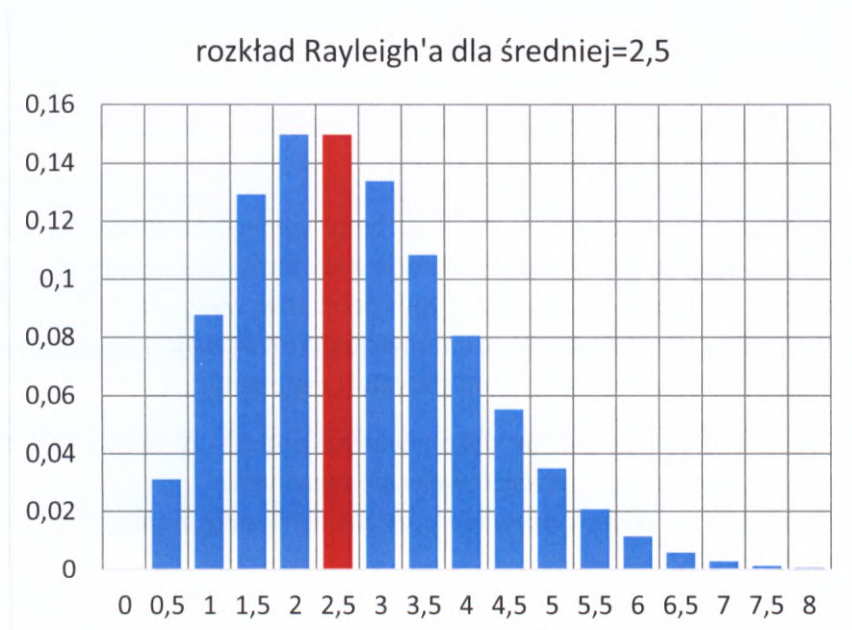
Il. 12.34. Korekta Miles–Lantuéjoul. Opis w tekście

- Wyznaczamy histogram względnej częstotliwości obiektów (liczba obiektów w klasie podzielona przez całkowitą liczbę obiektów), czyli prawdopodobieństwa zaliczenia do danej klasy zgodnie z mierzoną wielkością, np. polem powierzchni.
- Korygujemy histogram w poszczególnych klasach, zastępując liczbę obiektów w klasie sumą wag tych obiektów, wyznaczanych zgodnie ze wzorem (12.19).

Il. 12.34 pokazuje wyraźnie, że prosta operacja *border kill* zawyża udział drobnych obiektów, a jednocześnie zaniża udział dużych obiektów.

Rozkład pokazany na il. 12.34 dotyczy danych modelowych. Założono, że obiekty są kołami o średnicy równej środkom klas. Jest ich kolejno: 5, 20, 30, 20, 8, 4 i 1. Znajdują się w prostokącie o wymiarach 500x200 pikseli.

I wreszcie pewna interesująca informacja dotycząca opisu rozkładów wielkości. Jeżeli badamy zestaw obiektów, którymi mogą być odłamki skał, ziarna w próbce metalu, ziarenka kawy itp., to nie mają one identycznej wielkości, tylko wykazują pewien rozkład. Jego kresem dolnym jest oczywiście zero, dla pewnej wartości występuje maksimum (najwięcej obiektów ma rozmiary w pobliżu tej wartości), a potem następuje systematyczny spadek. Obiektów znacznie większych od tych odpowiadających



Il. 12.35. Rozkład Rayleigha. Czerwonym kolorem zaznaczono wartość odpowiadającą średniej dla całej populacji. Opis w tekście

maksimum częstości występowania jest już bardzo mało. Statystycy opracowali wiele rozkładów, ale jeden z nich wydaje się wyjątkowo ciekawy. Jest to rozkład Rayleigha (il. 12.35). Wykazuje on pewne interesujące cechy:

- Wykazano, że dobrze pasuje m.in. do opisu rozkładu wielkości wielu cząstek w materiałach.
- Ma bardzo podobny przebieg do powszechnie używanego rozkładu logarytmowo-normalnego (na ogół testy statystyczne nie wykazują istotnych różnic w dopasowaniu obu tych rozkładów do danych doświadczalnych).
- Wyznaczenie parametrów tego rozkładu jest bardzo proste. Wystarczy znajomość średniej arytmetycznej analizowanych wielkości.

Poniżej przedstawiamy konkretne wzory. Funkcja gęstości prawdopodobieństwa dla rozkładu Rayleigha ma postać:

$$f(d) = a \cdot d \cdot e^{-\frac{a}{2}d^2} \quad (12.20)$$

gdzie:

$$a = \frac{\pi}{2(\bar{d})^2} \quad (12.21)$$

Symbol $\bar{d}$ oznacza średnią arytmetyczną analizowanej wielkości. Warto spróbować użyć tego niezbyt popularnego rozkładu podczas analizy danych doświadczalnych, zwłaszcza jeśli dotyczą rozkładu wielkości różnych obiektów badanych za pomocą metod komputerowej analizy obrazu.

Zagadnienia weryfikacji zgodności rozkładu empirycznego z założonym rozkładem wielkości, podobnie jak wyznaczanie oraz ocena istotności różnic w wartościach statystycznych parametrów opisujących wyniki pomiarów, wykraczają poza zakres niniejszego podręcznika. Odpowiednie informacje Czytelnik znajdzie w bogatej literaturze, dotyczącej zastosowania metod statystyki matematycznej.

W ramach omawiania pomiarów wykonywanych podczas komputerowej analizy obrazu przedstawiono dość szczegółowo wiele zagadnień:

- podstawowe parametry używane podczas cyfrowych pomiarów w analizie obrazu,
- elementy algorytmów wybranych pomiarów,
- błędy wynikające z cyfrowej, dyskretnej budowy obrazów,
- błędy wynikające z takich czynników jak niewłaściwa binaryzacja czy zły dobór mierzonych parametrów,

- elementy interpretacji i prezentacji wyników,
- pewne wskazówki praktyczne.

Wszystko to razem powinno dać solidne podstawy do prób samodzielnego prowadzenia pomiarów w analizie obrazu oraz ich interpretacji.

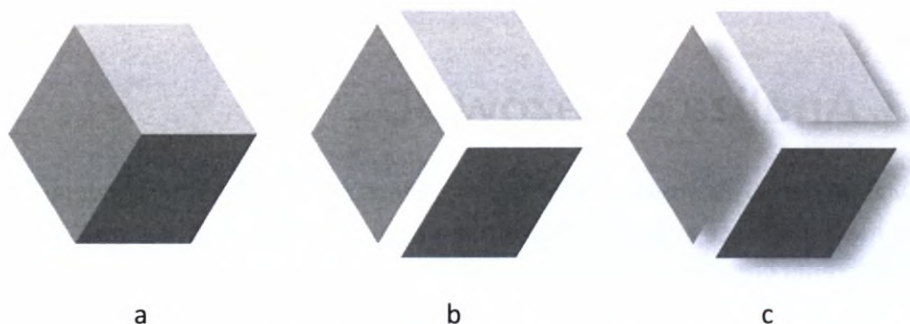
13. Analiza obrazów 3D

Świat jest trójwymiarowy, a informacja o przestrzennej organizacji naszego otoczenia ma decydujące znaczenie dla poruszania się czy używania narzędzi. Dzięki doświadczeniu często do orientacji w przestrzeni wystarczy nam płaski obraz, jak np. pokazanego na il. 13.1 pałacu Bellestremów w Pławniowicach. Bez trudu rozpoznajemy, że kształt budynku przypomina literę C, i oceniamy, że grupa kilku osób znajduje się w przybliżeniu w połowie odległości między miejscem, z którego wykonano zdjęcie, a wejściem do pałacu. Obraz na il. 13.1 wykazuje silnie zniekształconą geometrię obiektów położonych w okolicach brzegu obrazu, np. baszta w lewej strony jest zakrzywiona. Podobnie zniekształcony obraz otrzymuje nasze oko, a korekta geometrii dokonuje się w naszym mózgu.

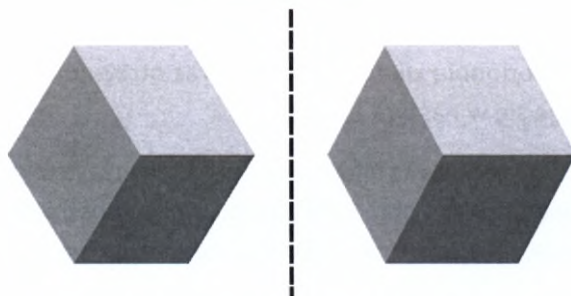
Potrąfimy zatem odczytać pewną informację dotyczącą przestrzeni 3D, czyli trójwymiarowej, na podstawie płaskiego obrazu. Tego typu umiejętności są bardzo rozpowszechnione w świecie zwierząt. Ptaki, ryby i ssaki roślinożerne, jak zające, sarny, jelenie, słonie lub konie, mają oczy umieszczone po dwóch stronach głowy, co pozwala na obserwację jedynie płaskich obrazów. Takie rozmieszczenie oczu ma jednak bardzo istotną zaletę. Umożliwia bowiem kontrolę bardzo dużego zakresu przestrzeni, obejmującego również obszar za zwierzęciem, co ułatwia dostrzeżenie drapieżnika i uratowanie życia.



Il. 13.1. Pałac Bellestremów w Pławniowicach, zarejestrowany obiektywem o bardzo krótkiej ogniskowej i nie poddany korekcie geometrii



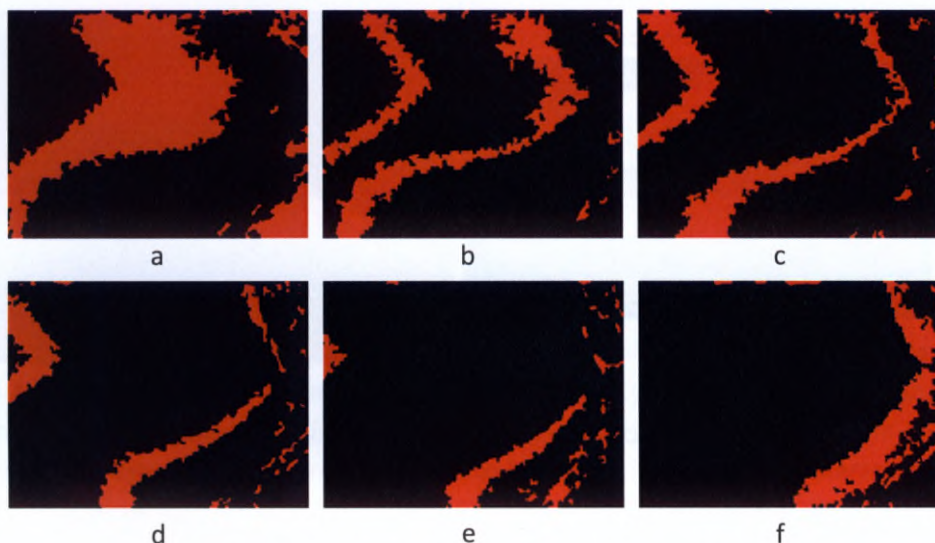
Il. 13.2. Niewielkie przesunięcia trzech rombów powodują istotne zmiany w odbiorze treści obrazu. Opis w tekście



Il. 13.3. Jeżeli rozdzielimy te obrazy kartką umieszczoną wzdłuż pionowej linii i będziemy patrzyli lewym okiem na lewy rysunek, a prawym okiem na prawy rysunek, to zobaczymy trójwymiarowy, przestrzenny obraz sześcienu

Il. 13.2a przedstawia trzy stykające się romby. Wypełnienie różnymi odcieniami szarości ich powierzchni sprawia, że sprawiają wrażenie sześciennnej kostki. Jeżeli jednak trochę rozsuniemy te romby (il. 13.2b), zaczynamy mieć wątpliwości, czy są to odrębne figury, czy może jednak ściany sześcienu o stępionych krawędziach. Do obiektów na il. 13.2b wystarczy jednak dodać symulowany cień (il. 13.2c), abyśmy mieli wrażenie, że to jednak trzy płytki, a nie ściany jakiejś bryły. Zatem nasze doświadczenie i wyobrażnia odgrywają istotną, jeśli nie decydującą rolę w interpretacji zawartości obrazu.

Patrząc z kolei na il. 13.3 widzimy dwa lekko różniące się od siebie obrazy sześciennnej kostki. Wprawdzie podczas analizy il. 13.2 wykazaliśmy, że nie ma żadnej pewności, że jest to kostka, ale teraz na chwilę o tym zapominamy. Jeżeli za pomocą odpowiednio zorientowanej kartki papieru zmusimy nasze oczy do oddzielnej obserwacji tych obrazów, to po chwili będziemy mieli wrażenie oglądania zdecydowanie trójwymiarowej bryły, unoszącej się nad kartką. Jest to efekt widzenia stereoskopowego, polegającego na tym, że mózg, składając dwa lekko różniące się od siebie obrazy,



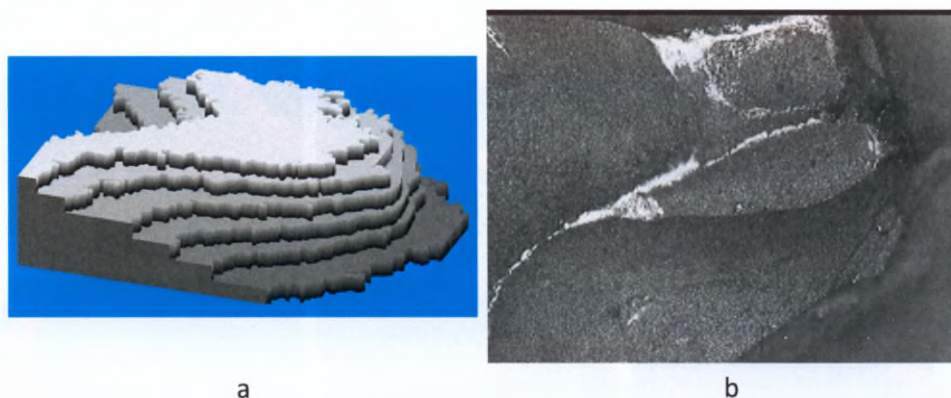
Il. 13.4. Kolejne obszary maksymalnej ostrości obrazów z il. 14.1

buduje trójwymiarowy model z precyzyjną oceną odległości poszczególnych obszarów, a nawet punktów tego obiektu.

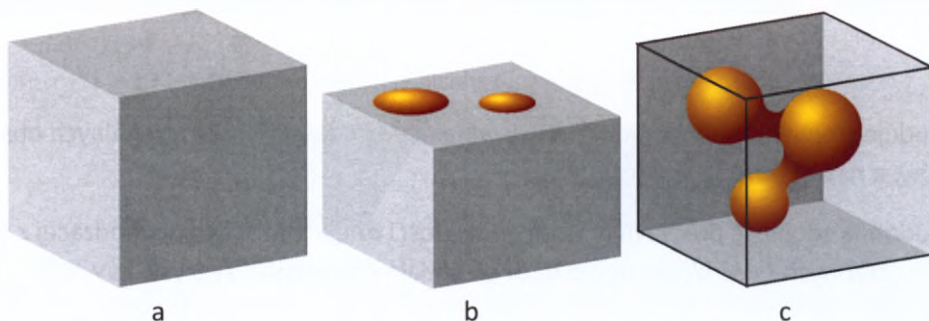
Podobną zdolność precyzyjnej oceny odległości dzięki informacji pochodzącej z rozmieszczonych dość blisko siebie oczu posiadają drapieżniki: tygrysy, lwy, wilki, lisy, a także zaprzyjaźnione z nami koty i psy. W odróżnieniu od roślinożerców nie jest im potrzebna obserwacja całego otoczenia, tylko precyzyjna lokalizacja ofiary, bez której nie będzie sukcesu podczas polowania. Opisane powyżej zagadnienia są przedmiotem zainteresowania fotogrametrii, dzięki której powstają precyzyjne plany i rekonstrukcje różnych budowli, a także mapy. Fotogrametria wykracza jednak poza ramy niniejszej książki i nie będziemy poświęcali jej czasu i miejsca w dalszej części.

Obiektywną informację 3D można też uzyskać na podstawie analizy serii płaskich obrazów, jeżeli są zogniskowane w różnych płaszczyznach (por. il. 14.1). Kolejne obszary maksymalnej ostrości, używane jako maski do składania ostrego obrazu (il. 13.4), mogą być wykorzystane do modelowania obserwowanej powierzchni (il. 13.5a). Tego typu analizę wykonuje się za pomocą mikroskopów konfokalnych. Używa się w nich monochromatycznego światła laserowego, które umożliwia utworzenie ostrego obrazu w niewielkim zakresie odległości i dzięki temu zbudowanie dość wiernego modelu obserwowanych obiektów.

Przedstawione metody oceny struktury 3D mają wspólną jedną podstawową wadę: pozwalają na ocenę jedynie zewnętrznej warstwy całej analizowanej struktury (il. 13.6a). Nie mamy żadnej informacji o jej wewnętrznej budowie.



Il. 13.5. a) uproszczony model 3D bryły, b) ostry obraz jej powierzchni



Il. 13.6. Analiza 3D. Opis w tekście

Można opisanym powyżej ograniczeniom częściowo zaradzić, wykonując fizyczne przekroje (il. 13.6b). Pozwala to uzyskać wiele cennych informacji, a metodami pozyskiwania tych danych zajmuje się stereologia, do której wstęp zawiera jeden z poprzednich rozdziałów. Niestety, analiza stereologiczna nie da nam informacji na temat ewentualnego połączenia obiektów widocznych na przekroju na il. 13.6b. Taką informację otrzymamy dopiero dzięki obrazowi 3D (il. 13.6c).

Jak wynika z tego rozumowania, obrazem 3D będziemy nazywali trójwymiarowy zbiór danych, opisujący lokalizację oraz kolor, stopień szarości lub inne atrybuty wszystkich punktów (nazywanych woksłami) należących do tego zbioru.

Obrazy 3D są w istocie jedynym dostępnym źródłem informacji, pozwalającym na pełną ocenę kształtu, wzajemnego położenia oraz spójności obiektów. Używa się ich m.in. w badaniach medycznych, biologicznych, materiałoznawczych oraz geologicznych. Na przykład badanie charakteru porowatości próbek skał pozyskanych z odwiertów decyduje o przydatności do ewentualnego pozyskiwania ropy lub gazu.

Jeżeli pory są zamknięte, czyli brakuje połączeń między nimi, to nie nadają się do eksploatacji złoża, bo ropa lub gaz nie będą mogły dostać się do szybu.

Pozyskiwanie obrazów 3D na drodze wykonywania dużej liczby fizycznych przekrojów jest często trudne technicznie (np. wymaga zamrażania preparatów biologicznych lub mozolnego szlifowania i polerowania twardych materiałów), zajmuje dużo czasu i jest badaniem niszczącym. Z tego ostatniego względu nie nadają się do badania żywych organizmów.

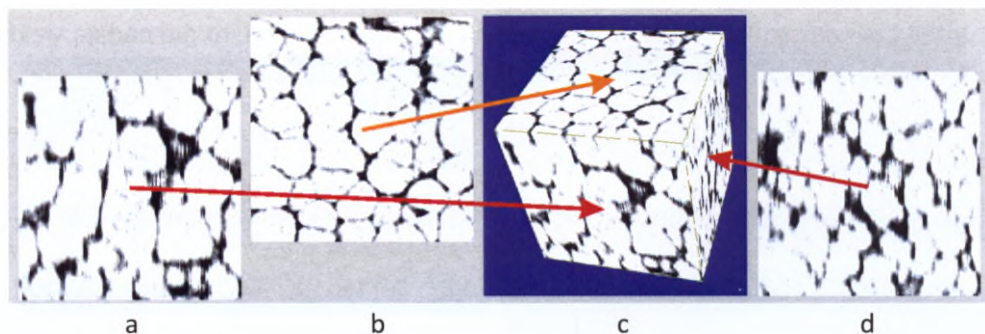
Najczęściej obrazy 3D pozyskuje się dzięki zastosowaniu rentgenowskiej tomografii komputerowej. W przypadku badań medycznych używane są również: rezonans magnetyczny, ultrasonografia 3D oraz badania wykorzystujące izotopy promieniotwórcze. Używane przez te metody techniki badawcze i wstępne przetwarzanie danych są skomplikowane, a ich omówienie wykracza poza ramy niniejszej książki, dlatego ograniczymy się do omówienia podstaw analizy obrazów 3D, bez wnikania w problemy ich pozyskiwania.

Główne problemy związane z analizą obrazów 3D skupiają się wokół trzech zagadnień:

- akwizycji obrazów 3D, a konkretnie silnych ograniczeń technicznych,
- problemów z wizualizacją obrazów 3D oraz wyników analizy,
- wysokiej zasobochłonności, wynikającej z bardzo dużej objętości plików i skomplikowanej analizy oraz problemów z wyznaczaniem wartości parametrów opisujących badane struktury 3D.

Podczas tomografii badany element jest prześwietlany pod różnymi kątami w jednej płaszczyźnie. W badaniach medycznych cały układ pomiarowy obraca się wokół nieruchomego pacjenta, a w pozostałych badaniach zwykle obraca się próbka, co bardzo upraszcza konstrukcję tomografu. Zawsze, nawet w przypadku tzw. spiralnej tomografii komputerowej, powstają wirtualne przekroje, z których składa się końcowy obraz 3D. Taki sposób powstawania obrazu 3D powoduje, że w przypadku np. bardzo cienkich ścianek komórek mogą one zostać wykryte tylko wtedy, gdy przecinają płaszczyznę skanowania pod dużym kątem (il. 13.7b). Przekroje, będące efektem przecinania obrazu 3D przez płaszczyzny prostopadłe do płaszczyzny skanowania (il. 13.7a i d), są pozbawione tej części ścian komórkowych, które były w przybliżeniu równoległe do płaszczyzny skanowania.

W przypadku obrazów medycznych bardzo często odległości między płaszczyznami są większe niż odległości między punktami obrazu w płaszczyźnie skanowania. Utrudnia to późniejsze analizy, gdyż uzyskanie wokseli równooddalonych we



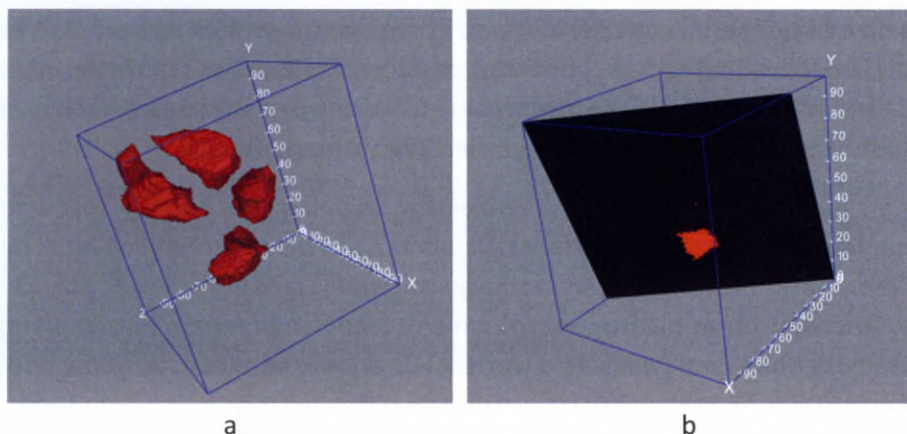
Il. 13.7. Problemy akwizycji obrazów 3D. Płaszczyzna skanowania odpowiadała górnej powierzchni testowego sześcianu (obraz b), natomiast obrazy (a) oraz (d) powstały w wyniku dodawania kolejnych płaszczyzn skanowania

wszystkich kierunkach wymaga przeliczania obrazu i wykorzystania metod interpolacyjnych, co grozi niekiedy powstaniem artefaktów, czyli elementów niewystępujących w rzeczywistości.

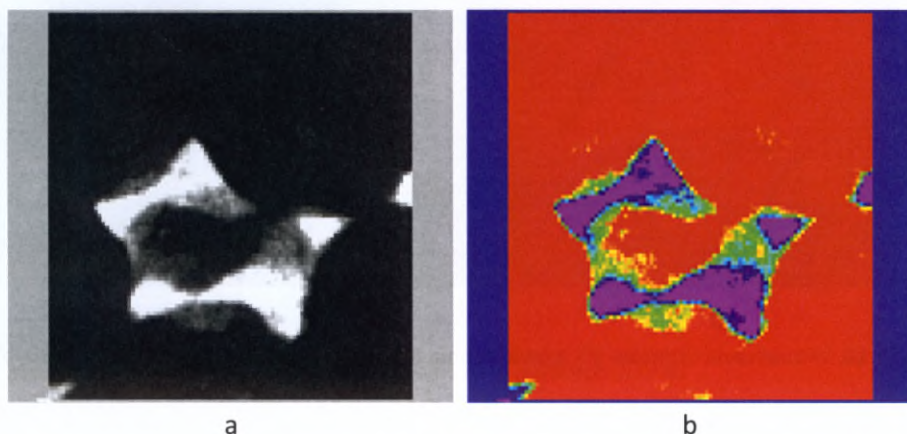
Zwykle obrazy 3D mają istotnie gorszą zdolność rozdzielczą niż obrazy 2D. Na przykład w badaniach medycznych rozmiar pojedynczego woksela jest rzędu milimetrów, co nie pozwala na precyzyjną analizę bardzo drobnych struktur. Dodatkowo, tomografia rentgenowska rozpoznaje jako odrębne tylko obszary różniące się istotnie masą atomową. Powoduje to, że bardzo dobrze wykrywa porowatość, natomiast nie rozpoznaje oddzielnych ziaren w jednofazowych materiałach metalicznych. Taki materiał jest dla tomografii po prostu jednorodny. Nie zmienia to faktu, że obrazy 3D są bardzo cennym materiałem badawczym.

Kolejnym problemem jest wizualizacja obrazów 3D, gdyż nie jest możliwe pokazanie przestrzennego obrazu na płaskim ekranie lub wydruku. Jeżeli mamy pojedyncze obiekty, w przypadku których ważny jest ich kształt, a nie wewnętrzna budowa, to ich pokazanie nie stanowi większego problemu (il. 13.8). Można wyświetlić obiekty przy różnie zorientowanym układzie współrzędnych (il. 13.8a), jak również ich przekroje, przy czym nie muszą one być równoległe do żadnej płaszczyzny wyznaczonej osiami współrzędnych (il. 13.8b).

Obrazy 3D mogą wykazywać zróżnicowane poziomy szarości w obrębie obiektu, co jest widoczne na il. 13.9a, a jeszcze wyraźniej pokazuje il. 13.9b, w którym poziomy szarości zostały przetworzone na kolory. Na il. 13.9b widać, że wnętrza beleczek siatki są jednorodne, ale okolice powierzchni wykazują budowę warstwową. Można to wykorzystać do wizualizacji obiektów przez wyświetlanie wszystkich wokseli o wybranym poziomie szarości. Tworzy się wtedy obraz powłok odwzorowujących



Il. 13.8. Wizualizacja obrazów 3D odizolowanych obiektów: a) obiekty w przestrzeni, b) przykład przekroju



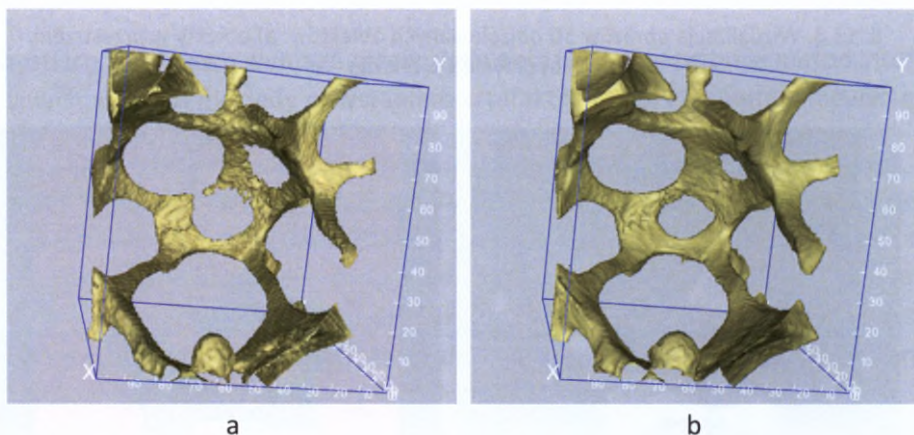
Il. 13.9. Przekrój przez obraz 3D siatki pianki metalicznej: a) obraz w skali szarości, b) obraz w wersji pseudokolorowej

kształty obiektu. Zmiana poziomu szarości powoduje zmianę w kształcie wyświetlanych obiektów (il. 13.10).

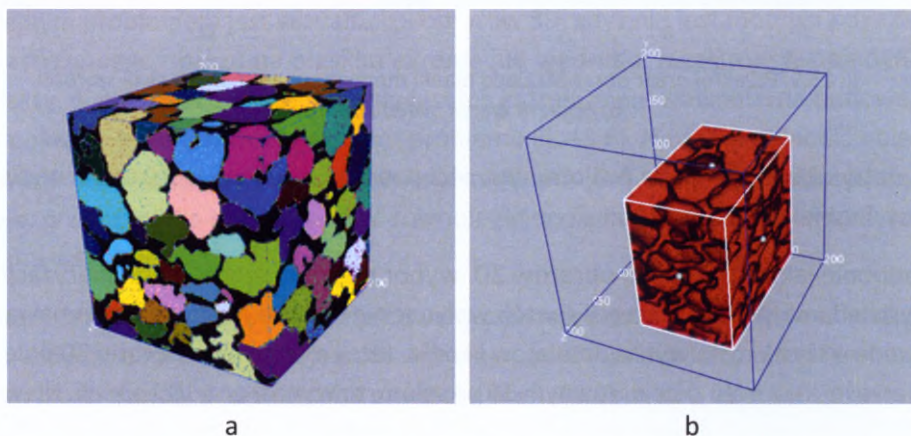
Podobnie jak w przypadku obrazów 2D, wybór poziomu szarości do binaryzacji lub wyświetlania może w znaczący sposób wpłynąć nie tylko na naszą ocenę wizualną, ale przede wszystkim na wyniki pomiarów i analiz. Jeżeli mierzymy w obrazie 2D pole powierzchni kwadratu o boku równym 10 pikselom, to wynosi ono 100 pikseli. Niewielki błąd w binaryzacji może nam dodać wokół kwadratu warstwę o grubości 1 piksela. Mniej pomylić się, np. o pół piksela, nie można. Taki błąd spowoduje przyrost pola powierzchni o 44%. Analogiczny błąd przy analizie objętości sześcianu o boku 10 wokseli

da nam zawyżenie objętości już o blisko 73%. Mamy zatem niebezpieczeństwo bardzo poważnych błędów podczas pomiarów obrazów 3D. Co prawda może się wydawać, że 10 wokseli to niewiele, ale trzeba pamiętać o mniejszej rozdzielczości obrazów 3D w porównaniu z obrazami 2D dotyczącymi tego samego obiektu.

W przypadku gęsto upakowanych obiektów, wypełniających całą przestrzeń, wyświetlanie powierzchni isosurface daje obraz całkowicie nieczytelny. Jeżeli dysponujemy obrazem binarnym, to można go wyświetlić jako etykietowany (il. 13.11a). Wyświetlając obraz binarny, można zmieniać stopień przezroczystości przestrzeni między obiektami (woksele o wartości 0), a także wybierać dowolny podzbiór obrazu 3D (il. 13.11b).



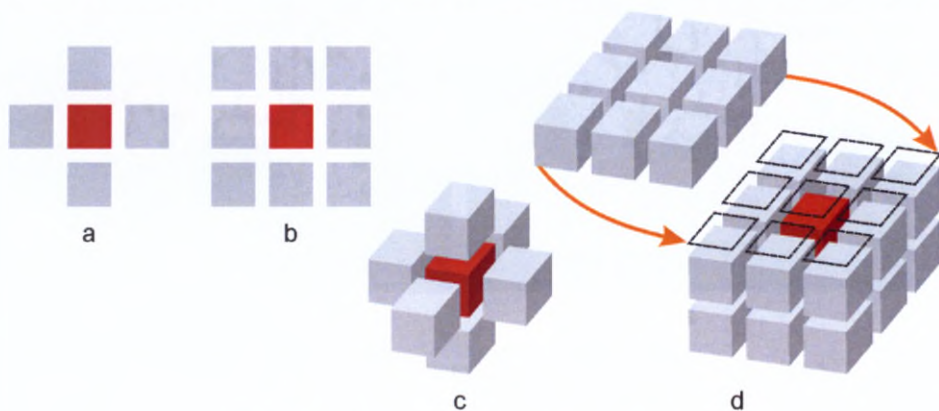
Il. 13.10. Wizualizacja obrazów 3D; powierzchnia isosurface dla poziomu szarości: a) 230, b) 128



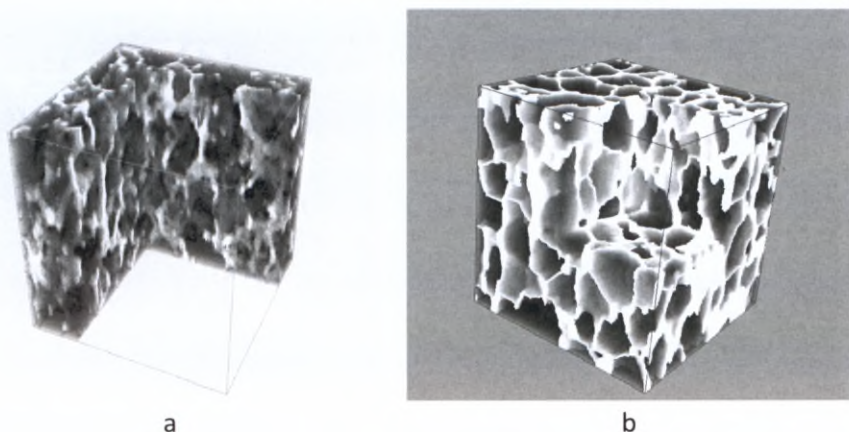
Il. 13.11. Wizualizacja obrazów 3D: a) obraz etykietowany dużej liczby obiektów, b) obserwacja podzbioru analizowanego obrazu 3D

Filtry i przekształcenia morfologiczne w obrazach 3D wykonuje się podobnie jak w przypadku obrazów 2D. Zmianie ulegają jedynie zasady spójności (il. 13.12). Odpowiednikiem 4 najbliższych sąsiadów w obrazach 2D (il. 13.12a) jest 6 sąsiadów w obrazach 3D (il. 13.12c), a odpowiednikiem sąsiedztwa 8-spójnego (il. 13.12b) jest 26 najbliższych sąsiadów w obrazie 3D (il. 13.12d). Konsekwencją tych zmian jest przede wszystkim wydłużenie czasu poszczególnych operacji, co może być istotnym utrudnieniem w przypadku dużych rozmiarów obrazu lub mniejszej mocy obliczeniowej komputera. Należy też spodziewać się błędów lub zaskakujących wyników w przypadku elementów 3D o niewielkich rozmiarach.

Prześledzimy teraz proces detekcji pustek w polimerowej piance (il. 13.13a). Istotnym problemem był w tym przypadku brak rejestracji w wyjściowym obrazie części



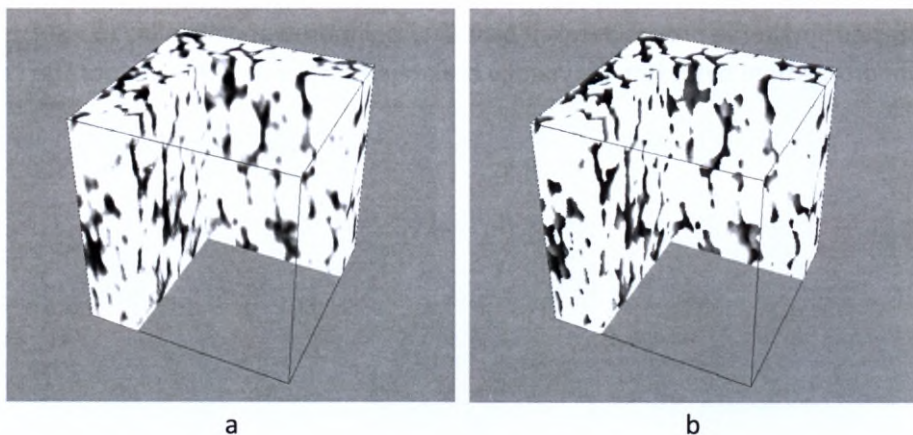
Il. 13.12. Zasady spójności w obrazach 2D i 3D



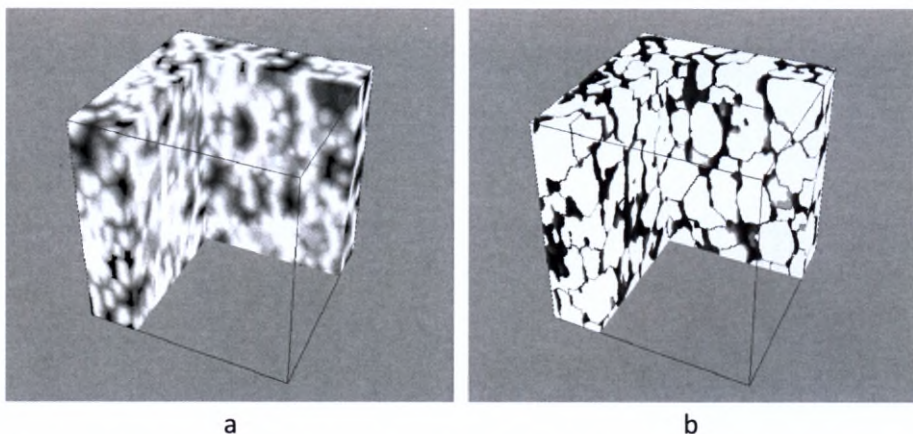
Il. 13.13. Obraz 3D pianki polimerowej: a) oryginalny obraz z tomografu, b) przetworzony obraz z odtworzonymi komórkami

granic porów. Widać to na il. 13.7, który dotyczy tej samej pianki. Ponieważ początkowy obraz zawiera dość dużo szumów, pierwszym krokiem była filtracja medianowa (il. 13.14a). Następny krok stanowiła binaryzacja, co objawia się ostrzejszymi krawędziami poszczególnych obszarów (il. 13.14b).

Kolejnym krokiem było wygenerowanie funkcji odległości 3D, która jest przestrzennym odpowiednikiem przekształcenia omawianego w przypadku przekształceń morfologicznych. Lokalne minima negatywu tej funkcji, widoczne jako czarne plamy na il. 13.15a, pełniły rolę znaczników do detekcji działów wodnych. Ta ostanía operacja



Il. 13.14. Analiza obrazu z il. 13.13a: a) mediana, b) binaryzacja. W celu poprawy czytelności obrazu pustki są wyświetlane jako bryły i dlatego pustki są białe i nieprzezroczyste, a granice między nimi czarne



Il. 13.15. Analiza obrazu w il. 13.13a: a) funkcja odległości i lokalne minima, b) końcowy wynik detekcji; pory pokazane jako bryły

wygenerowała siatkę o grubości pojedynczych wokseli. Dodana do zbinaryzowanego obrazu (il. 13.14b) doprowadziła do końcowego wyniku, pokazanego na il. 13.15b oraz, w innej formie graficznej, na il. 13.13b. Detekcja działów wodnych pozwoliła na uzupełnienie brakujących granic między pustkami, zwłaszcza w przypadku granic o orientacji zbliżonej do płaszczyzny skanowania.

Przedstawione w tym rozdziale elementy analizy obrazów 3D stanowią jedynie wstęp do bardzo rozbudowanej gałęzi badań, zwłaszcza w zakresie medycyny. Analizy 3D mają swoje ograniczenia, ale podstawy ich analizy nie odbiegają od tego, o czym mówimy w przypadku obrazów 2D.

14. Budowa algorytmów

Komputerowa analiza obrazu może być skutecznym narzędziem, wspomagającym badania i analizy w wielu obszarach naszej aktywności. Aby pokazać bogactwo potencjalnych i realnych zastosowań, wymienimy niektóre z nich. Poniższy spis nie jest ani kompletny, ani uporządkowany, pomijając przedstawienie zestawienia w kolejności alfabetycznej:

- automatyczna weryfikacja dokumentów,
- biologia,
- biometria (rozpoznawanie twarzy, linii papilarnych, tęczówki oka itp.),
- diagnostyka medyczna,
- geografia i geologia,
- kontrola jakości,
- kryminalistyka,
- logistyka,
- muzealnictwo (analiza autentyczności zbiorów),
- nauka o materiałach,
- patomorfologia,
- rozpoznawanie kodów paskowych i QR kodów,
- rozpoznawanie numerów rejestracyjnych,
- rozpoznawanie znaków drogowych oraz inne systemy autonomicznych pojazdów,
- sortowanie poczty,
- systemy bezpieczeństwa,
- zastosowania militarne (rozpoznawanie celów, sterowanie, identyfikacja itp.).

Każda z tych dziedzin ma swoją metodykę badań, aparat pojęciowy, podstawy teoretyczne itp. Niekiedy dziedziny te są na tyle odległe od siebie, że specjalistom trudno się porozumieć, tym bardziej że przyswojenie sobie chociażby podstaw tej drugiej dziedziny wymaga znacznej ilości czasu i niekiedy ogromnego wysiłku. Pomimo przydatności metod analizy obrazu w każdej z wymienionych powyżej form aktywności nie stanowią one istotnego elementu żadnej z nich.

Z powyższego wstępu płyną bardzo ważne dla praktyki wnioski. Jeżeli wprowadzenie metod analizy obrazu jest konieczne, to:

- Najczęściej specjalista z danej dziedziny co najwyżej słyszał o komputerowej analizie obrazu.

- Jego zainteresowanie wykorzystaniem tego narzędzia jest albo dziełem przypadku (np. dowiedział się od kolegi, że takie coś istnieje), albo wymuszone przez akty prawne (niektóre normy kontroli jakości wymagają stosowania metod analizy obrazu).
- Będzie nas prosił o pomoc, bo sam nie ma ani czasu, ani ochoty, żeby uczyć się podstaw analizy obrazu.
- Może się też zdarzyć, że dysponuje sprzętem do analizy obrazu i oczekuje, że zrobimy z tym sprzętem coś takiego, że zacznie on wykonywać pracę, do której jest przeznaczony. Najlepiej, gdyby obsługa sprowadzała się do naciśnięcia jednego przycisku.

Nakreślone powyżej nastawienie specjalistów jest, wbrew pozorom, całkiem racjonalne. Jeżeli któryś z nich z dowolnych powodów rzeczywiście chce lub potrzebuje prowadzić analizę obrazów, to współpraca może okazać się całkiem interesująca. Niewiele osób zna się na analizie obrazu, co daje również pewną szansę na godziwe wynagrodzenie. Widać jednak wyraźnie, że jeśli chcemy, aby nasza współpraca przyniosła rzeczywiste efekty, to musimy choć trochę poznać dziedzinę, w której ma być użyta analiza obrazu.

Jeżeli jeden materiałoznawca zażąda od nas oceny wtrąceń niemetalicznych, a drugi porowatości, to musimy wiedzieć, jak je rozpoznać, bo na obrazach do badań mogą wyglądać bardzo podobnie. Tak samo z lekarzami. Jeśli usłyszymy od kardiologa, że mamy archiwalne obrazy z rezonansu tylko w płaszczyźnie poziomej, więc nie będziemy mogli bezpośrednio porównać wyników z projekcją czterojamową, to musimy wiedzieć, o co w ogóle chodzi, żeby zaproponować jakikolwiek tryb prac nad wdrożeniem metod analizy obrazu.

Ogólna zasada jest następująca: musimy poznać problem na tyle dobrze, abyśmy mogli porozumieć się ze specjalistą z dziedziny, a ponadto powinniśmy potrafić sami, w sposób manualny, wyznaczyć poszukiwane obiekty lub sklasyfikować analizowany obraz. Nawet jeżeli chcemy wykorzystać sieci neuronowe, co pozornie zwalnia nas z głębszej wiedzy na temat analizy obrazu, powinniśmy potrafić ocenić samodzielnie analizowane obrazy po to, żeby móc wstępnie zweryfikować prawidłowość pracy systemu.

Kolejna praktyczna uwaga dotyczy weryfikacji skuteczności opracowywanego systemu analizy. Zakładając dla uproszczenia, że badany obiekt jest albo dobry, albo zły, możemy wyróżnić cztery odmiany wyników:

- prawdziwie pozytywny (ang. *true positive*, *TP*) – dobry oceniamy jako dobry,
- prawdziwie negatywny (ang. *true negative*, *TN*) – zły oceniamy jako zły,
- fałszywie pozytywny (ang. *false positive*, *FP*) – zły oceniamy jako dobry i ten stan nazywamy błędem pierwszego rodzaju,
- fałszywie negatywny (ang. *false negative*, *FN*) – dobry oceniamy jako zły i ten stan nazywamy błędem drugiego rodzaju.

Jeżeli wyniki naszych analiz zostały zweryfikowane i znamy liczby wyników ze wszystkich czterech wymienionych grup, to możemy określić różne miary jakości naszej metody. Najczęściej podawanymi parametrami są:

- **Czułość** (ang. *sensitivity*) lub udział prawdziwie pozytywnych (ang. *true positive rate*, *TPR*) oznacza po prostu udział wyników poprawnie ocenionych jako dobre wśród wszystkich dobrych:

$$TPR = \frac{\sum TP}{\sum TP + \sum FN} \quad (14.1)$$

- **Swoistość** (ang. *specificity*, *SPC*) lub udział prawdziwie negatywnych (ang. *true negative rate*, *TNR*) oznacza po prostu udział wyników poprawnie ocenionych jako złe wśród wszystkich złych:

$$TNR = \frac{\sum TN}{\sum TN + \sum FP} \quad (14.2)$$

- **Precyzja** (ang. *precision*) lub wartość predykcyjna dodatnia (ang. *positive predictive value*, *PPV*) oznacza po prostu udział wyników poprawnie ocenionych jako dobre wśród wszystkich ocenionych jako dobre:

$$PPV = \frac{\sum TP}{\sum TP + \sum FP} \quad (14.3)$$

- **Dokładność** (ang. *accuracy*, *ACC*) to po prostu udział poprawnych ocen (nieważne, dobrych czy złych) wśród wszystkich ocen:

$$ACC = \frac{\sum TP + \sum TN}{\sum TP + \sum TN + \sum FN + \sum FP} \quad (14.4)$$

Podsumowując, można powiedzieć w sposób nie do końca ścisły, ale oddający intencje tych miar, że:

- metoda jest **czuła**, jeżeli nie odrzuca niepotrzebnie dobrych wyników,
- metoda jest **swoista**, jeżeli poprawnie rozpoznaje złe wyniki,
- metoda jest **precyzyjna**, jeżeli nie dopuszcza złych wyników do dobrych,
- metoda jest **dokładna**, jeżeli wyniki są poprawne.

Byłoby wspaniale, gdyby nasza metoda była jednocześnie bardzo czuła, swoista, precyzyjna oraz dokładna. Na to są raczej małe szanse. W praktyce powinna być przede wszystkim maksymalnie precyzyjna. Inaczej mówiąc, nasza metoda będzie w praktyce bezpieczna, jeżeli nie dopuści do oceny złych wyników jako dobre. *Nota bene*, ten wniosek jest po prostu zdroworozsądkowy.

Zatem naszą metodę możemy zareklamować następująco: jeżeli metoda ocenia coś jako dobre, to jest to z pewnością dobre. A co zrobić, jeżeli metoda oceni to coś jako złe? Wtedy nie odrzucamy od razu, tylko sprawdzamy jeszcze raz, być może innymi metodami. Ktoś może powiedzieć, że taka metoda, nawet idealnie swoista i precyzyjna, będzie do niczego, jeżeli okaże się mało czuła i mało dokładna. To nie do końca prawda! Wyobraźmy sobie, że opracowaliśmy bardzo tanią metodę, o idealnej precyzji i swoistości, ale o niskiej czułości i dokładności. Sprawia to, że po badaniu naszą metodą i tak musimy ponownie zbadać większość przypadków, bo tych, których wyników jesteśmy pewni, jest niewiele. Jeżeli jednak te kolejne badania będą bardzo drogie, to nasza bardzo tania metoda pozwoli zaoszczędzić wiele pieniędzy nawet przy niewielkiej czułości i dokładności.

**PODSUMOWANIE: NASZA METODA NIE MUSI WSZYSTKIEGO OCENIĆ.
ALE TO, CO OCENIA, POWINNA OCENIAĆ DOBRZE.**

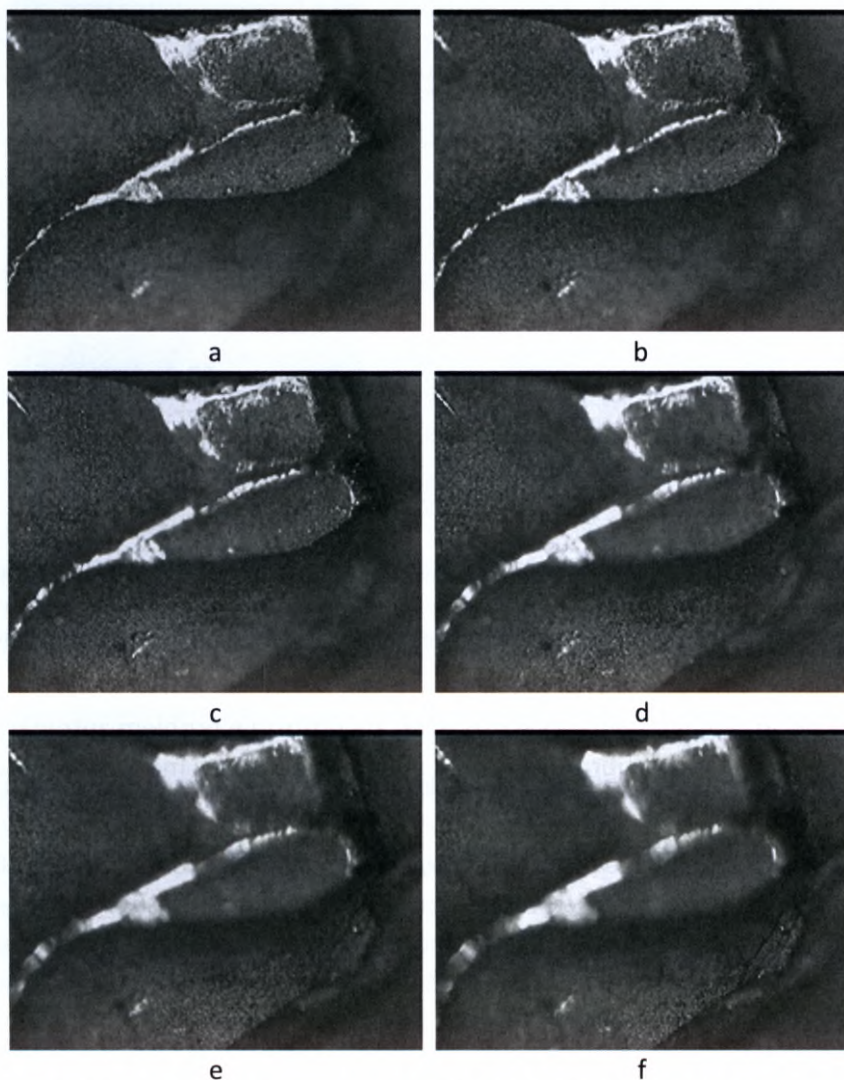
Wiemy już, jakie są oczekiwania wobec naszej metody (w tym przypadku będzie nią odpowiednio opracowany algorytm). Musimy teraz zastanowić się, jak go stworzyć w możliwie prosty i szybki sposób. W tym miejscu pojawia się podstawowy problem: metody analizy obrazu leżą na pograniczu ścisłego formalizmu matematycznego (zrezygnowano z niego na przykład przy omawianiu przekształceń morfologicznych, aby książka była zrozumiała dla większego grona Czytelników) oraz sztuki, np. poezji.

To, że poszczególne przekształcenia są sformalizowane i ściśle zdefiniowane, nie budzi wątpliwości. Skąd jednak porównanie do poezji? Analogia jest, wbrew pozorom, daleko posunięta. Znajomość słownictwa, a także zasad gramatyki, ortografii oraz interpunkcji nie wystarczy do napisania wiersza. Potrzebny jest jeszcze pomysł, co i w jaki sposób chcemy w tym wierszu zawrzeć. W dodatku nie mamy ścisłych reguł, jak wiersz napisać. W przypadku analizy obrazu jest podobnie. Mamy wiele operacji, których działanie zostało opisane w poprzednich rozdziałach, znamy ich sposób działania i możliwe do uzyskania wyniki. Ale to wciąż nie wystarczy do zbudowania skutecznie działającego algorytmu. Aby go stworzyć, musimy użyć wyobraźni i często wypróbować kilka różnych rozwiązań. Jest to niewątpliwie proces twórczy, stąd pewna analogia do sztuki.

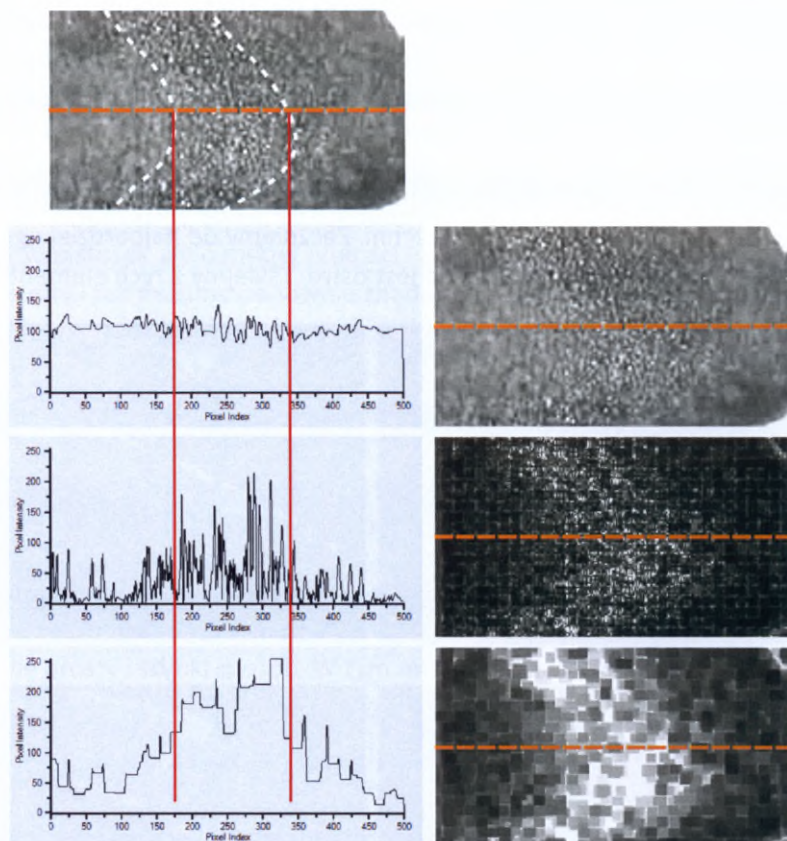
Prześledzimy proces budowania algorytmu na przykładzie budowy ostrego obrazu z serii sześciu częściowo ostrych obrazów, przedstawiających fragment powierzchni pęknięcia złamanego wiertła (il. 14.1). Ponieważ powierzchnia pęknięcia nie jest płaska, tylko rozbudowana przestrzennie, to niemożliwe jest wykonanie ostrego zdjęcia za pomocą mikroskopu świetlnego, który ma niewielką głębię ostrości (głębia ostrości to zakres odległości od obserwowanego przedmiotu, w którym jego elementy są

odzworowywane jako ostre; poza głębią ostrości obraz jest rozmyty). Ustawiając płaszczyznę ostrości na różnych wysokościach, otrzymano serię obrazów tego samego elementu, w których obszar ostrości przesunął się od najwyższej części centralnej (il. 14.1a) do części zewnętrznej (il. 14.1f).

Naszym zadaniem jest zbudowanie na podstawie tej sekwencji zdjęć jednego obrazu, który będzie ostry na całej powierzchni. Zaczniemy od najbardziej ogólnej koncepcji: weźmy z każdego obrazu to, co jest ostre, i sklejmy z tych elementów nowy,



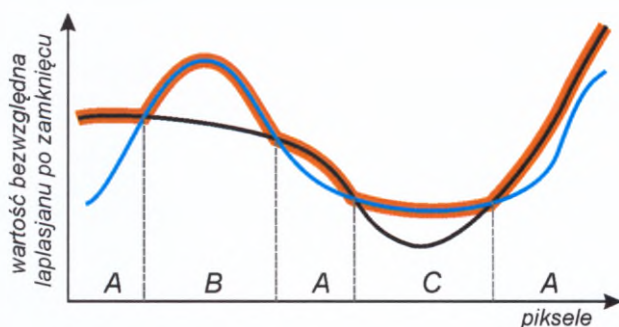
Il. 14.1. Sekwencja częściowo ostrych zdjęć tego samego elementu, wykonanych przy różnych ustawieniach ostrości. Opis w tekście



Il. 14.2. Detekcja ostrych obszarów obrazu. Od góry: fragment obrazu z zaznaczonym manualnie obszarem ostrości, ten sam obraz i profil jego środkowej części, obraz bezwzględnej wartości laplasjanu i jego profil, obraz bezwzględnej wartości laplasjanu po operacji zamknięcia i jego profil

ostry obraz. Pomysł wydaje się rozsądny, ale pozostaje problem automatycznego rozpoznania ostrości obrazu, gdyż nie mamy obiektywnej miary tego parametru.

Aby znaleźć odpowiednie rozwiązanie, trzeba zastanowić się, dlaczego pewne elementy wydają się być ostrzejsze od innych. Przyglądnijmy się fragmentowi il. 14.1c, pokazanemu w górnej części il. 14.2. Środkowa część tego obrazu, ograniczona przerywanymi białymi liniami, wydaje się być ostrzejsza od pozostałej części obrazu. Spowodowane to jest lokalnie lepszym kontrastem, czyli większą różnicą poziomów szarości sąsiadujących pikseli, niż w części sprawiającej wrażenie rozmytej. Potwierdza to profil poziomów szarości wyznaczony dla poziomej linii, zaznaczonej kolorem pomarańczowym. Zgodnie z naszymi oczekiwaniami, zęby piły, którą przypomina profil, są większe i węższe w obszarze uznanym przez nas jako ostry. Niestety, ten profil



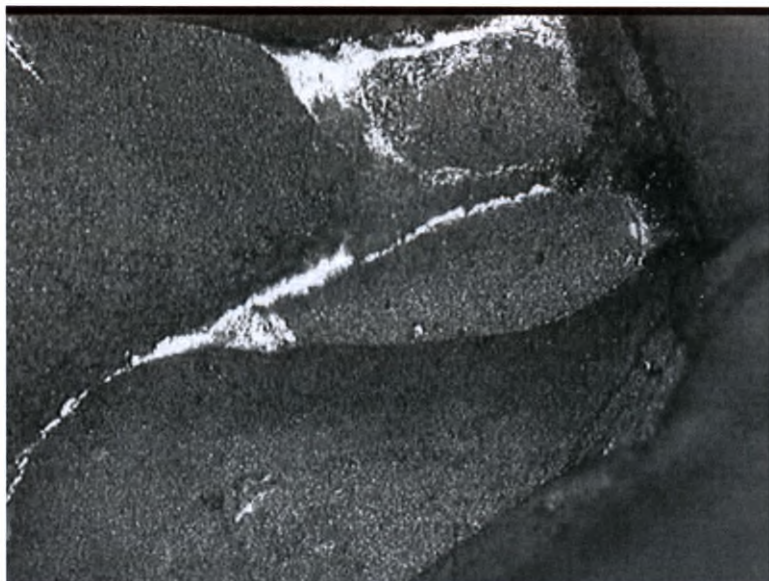
Il. 14.3. Wybór fragmentów najostrzejszych fragmentów obrazu do budowy złożonego końcowego obrazu. Opis w tekście

nie nadaje się do oceny ostrości, bo jego średnia wartość zależy od tego, czy trafia w jaśniejszy, czy też ciemniejszy obszar obrazu.

Mamy jednak odpowiednie narzędzie, które rozwiązuje ten problem. Jest nim laplasjan, który wzmacnia lokalne różnice poziomów szarości i jest niewrażliwy na zmiany średniej wartości poziomów szarości. Dla obszarów, w których poziom szarości jest stały, laplasjan przyjmuje wartość 0. Dla naszych potrzeb musimy pozbyć się ujemnych wartości laplasjanu, stąd też użyjemy jego bezwzględnej wartości (środkowy profil i obraz na il. 14.2). Widać, że bezwzględne wartości laplasjanu są wyższe w obszarze uznanym przez nas za ostry.

Ponieważ wysokie wartości są oddzielone głębokimi dolinami, warto ten obraz nieco wygładzić, np. przez operację zamknięcia. Efekt widoczny jest w dolnej części il. 14.2. Rejon ostrości pierwotnego obrazu jest teraz wyraźnie widoczny jako jasny pas w centrum obrazu. Pozostaje teraz tylko wybrać odpowiednie fragmenty każdego obrazu i złożyć je w całość.

Przeanalizujemy ten problem na przykładzie dwóch modelowych profili obrazów po wygładzeniu wartości bezwzględnej laplasjanu (il. 14.3). Mamy dwa obrazy – nazwijmy je zgodnie z kolorami linii na il. 14.3 czarnym i niebieskim (w przykładzie z il. 14.1 takich obrazów mamy sześć, ale cała metoda jest identyczna z przedstawioną na przykładzie dwóch obrazów). W obszarach oznaczonych literą A na il. 14.3 ostrzejszy jest obraz czarny i on powinien być brany pod uwagę przy składaniu wynikowego obrazu. W obszarze oznaczonym literą B ostrzejszy jest obraz niebieski. Ciekawy jest obszar C, w którym obraz niebieski jest najmniej ostry, ale mimo to ostrzejszy niż obraz czarny i dlatego w tym obszarze powinno się wykorzystać do złożenia obrazów właśnie fragment obrazu niebieskiego. Podsumowując, wybieramy obszary względnie najostrzejsze, ale jeżeli w składowych obrazach nie ma obszarów zdecydowanie



Il. 14.4. Końcowy wynik tworzenia ostrego obrazu

ostrzych, to metoda składania kilku obrazów nie wygeneruje obszaru bardziej ostrego, niż występujące w składowych obrazach.

W celu wybrania odpowiedniego fragmentu tworzymy dodatkowy obraz pomocniczy, będący maksimum wartości wszystkich składowych obrazów. Reprezentuje to pomarańczowa linia na il. 14.3. Jeżeli od obrazu maksimum odejmiemy się którąkolwiek składową, to wynik da wartość 0 w tych obszarach, w których obraz powiązany z daną składową jest najostriejszy. Zatem przez binaryzację z progiem 0 można utworzyć maskę do wyboru odpowiedniego fragmentu składowego obrazu. Zestaw takich masek pokazano na il. 13.4.

W ten sposób stworzyliśmy dość krótki, ale skuteczny i w istocie zaawansowany algorytm. Można go w uproszczony, syntetyczny sposób (bez definiowania wielkości matryc, elementów strukturalnych itp.) zapisać w następującej formie:

- Odczytaj n obrazów.
- Dla każdego z obrazów wykonaj sekwencję operacji: laplasjan, wartość bezwzględna, zamknięcie.
- Wykonaj obraz odniesienia: maksimum ze wszystkich obrazów po zamknięciu.
- Dla każdego z obrazów: odejmując od obrazu odniesienia wynik zamknięcia, ustal obszar, w którym dany obraz jest najostriejszy.
- Tak ustalony obszar wykorzystaj jako maskę dla obrazu, tworząc cegiełkę do budowy ostrego obrazu.

- Zsumuj poszczególne cegiełki, a ewentualne wolne obszary wypełnij zawartością ostatniego obrazu.
- Pochwal się wynikiem.

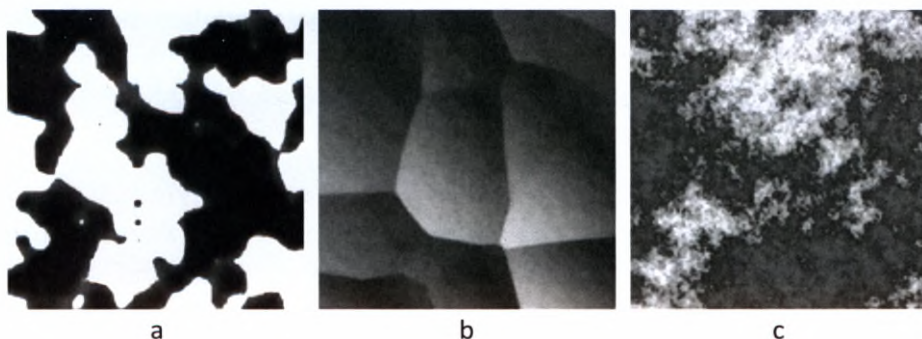
Omawiany algorytm jest dość krótki, choć powstaje sporo obrazów pośrednich, gdyż większość operacji jest sześciokrotnie powtarzana. Końcowy wynik jest poprawny i efektywny (il. 14.4). W omawianym algorytmie mamy dwa kluczowe problemy:

- Znalezienie sposobu wykrywania lokalnie ostrzejszych miejsc,
- Znalezienie sposobu wyboru fragmentów obrazu stanowiących wkład do budowy finalnego, ostrego obrazu.

Doświadczenia z rozwiązywania problemów analizy obrazu wskazują, że bardzo często o końcowym sukcesie decyduje znalezienie skutecznego rozwiązania jednego lub dwóch takich kluczowych elementów. Pozostała część algorytmu zwykle nie stanowi dużego problemu.

W obrazach możemy wyróżnić trzy podstawowe typy spotykane w praktyce:

- Typ A: granice poszczególnych obiektów są widoczne dzięki różnicy poziomów szarości (il. 14.5a). Przykładami takich obrazów mogą być: pory i wtrącenia niemetaliczne, zarysy ciemnych obiektów na jasnym tle (uszczelki, ziarna ściernie lub roślin itp.) lub tablice rejestracyjne. Skrajnym przypadkiem takiej sytuacji jest obraz binarny.
- Typ B: granice poszczególnych obiektów są widoczne dzięki temu, że tworzą je linie różniące się poziomem szarości od otoczenia lub widzimy je dzięki lokalnie dużym gradientom poziomów szarości (il. 14.5b). Przykładami takich obrazów mogą być: łączenia płytek na ścianie lub podłodze, granice między kostkami w bruku, wytrawione granice ziaren, rysy, żyłki w liściach itp.



Il. 14.5. Modelowe obrazy ilustrujące trzy typy problemów, z którymi często spotykamy się w analizie obrazu. Opis w tekście

- Typ C: granice pomiędzy obiektami są rozmyte i trudne do jednoznacznej oceny. W przypadku oceny przez kilka osób, każdy wskazuje na nieco inne położenie granic obiektów (il. 14.5c). Przykładami takich obrazów mogą być: chmury, drzewa (zwłaszcza o stykających się koronach), ciemne obiekty z cieniem, ślady na śniegu lub piasku, osoby w tłumie itp.

Powyższy podział został wprowadzony wyłącznie na potrzeby tego rozdziału i nie jest używany poza nim. Nie obejmuje też wszystkich możliwych rodzajów obrazów, ale ilustruje podstawowe cechy najczęściej spotykanych problemów. Trzeba też wziąć pod uwagę, że duże, skomplikowane obrazy mogą w różnych obszarach wykazywać zgodność ze wszystkimi pokazanymi na il. 14.5 typami obrazów.

Obrazy odpowiadające typowi A to przede wszystkim obrazy binarne i w ich przypadku nie ma problemów z detekcją obiektów. Trzeba tylko pamiętać o wpływie rodzaju spójności na zaliczanie poszczególnych pikseli do jednego obiektu. Można te obrazy modyfikować przede wszystkim za pomocą przekształceń morfologicznych oraz logicznych. Pewna doza wyobraźni oraz metoda prób i błędów zwykle dość szybko doprowadzają do pozytywnego końcowego wyniku analizy.

Trudniejsza jest analiza obrazów szarych. Zazwyczaj zaczyna się od najprostszego rozwiązania, którym jest wypróbowanie kilku sposobów binaryzacji. Czasami zdarza się (ale dość rzadko), że dobry wynik daje od razu któraś z metod automatycznej binaryzacji. Jeżeli tak nie jest, to należy wypróbować binaryzację interaktywną, podczas której obserwujemy, jak zmiany progu detekcji wpływają na wynik.

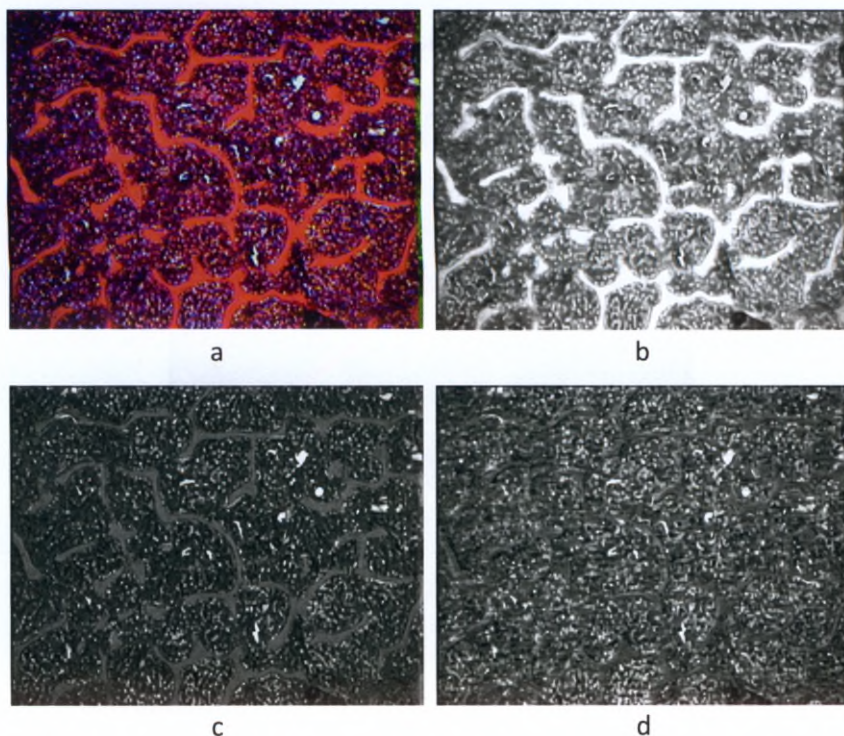
Próby manualnej binaryzacji pokazują od razu, czy damy radę wykryć obiekty bez daleko idącej wstępnej modyfikacji obrazu. Jeżeli tak, a jedynym problemem jest zły dobór progów detekcji przez metody automatyczne, to mamy połowę pracy za sobą. Wystarczy bowiem popracować nad odpowiednią metodą doboru progów detekcji. Czasami, zwłaszcza w sytuacjach, gdy nie przewidujemy masowych analiz, warto jest zatrzymać się na tym etapie i pozostawić metodę w formie interaktywnej, bez silenia się na pełną automatyzację.

Jeżeli próba interaktywnej binaryzacji wykazuje brak możliwości otrzymania satysfakcjonujących wyników, to trzeba poszukać przyczyny. Czasami można ją dostrzec bez dodatkowych analiz, ale zwykle znalezienie skutecznego rozwiązania wcale nie jest oczywiste. Bardzo dobrym narzędziem jest wtedy analiza linii profilu obrazu. Wiadać będzie na nim, czy obraz wykazuje obecność cienia, czy poziom szumu nie jest porównywalny z zakresem zmian poziomów szarości pomiędzy tłem i obiektami, czy rozpoznane przez nas krawędzie znajdują odbicie w pikach profilu itp. Analiza profilu

pozwała zdiagnozować problem i wskazuje, gdzie szukać rozwiązania. Zdarza się też, że analiza profilu nie pomoże wprost w znalezieniu rozwiązania, ale wykaże błędność przyjętej przez nas koncepcji. Wtedy należy szukać nowego sposobu. Interesującym przykładem w tym zakresie jest przedstawiony w rozdziale poświęconym analizie obrazów kolorowych wpływ przyjętego modelu barw na wyniki detekcji.

Problem budowy algorytmu detekcji obiektów w obrazach typu A przeanalizujemy dodatkowo na podstawie obrazu beleczek kostnych (il. 14.6a). Beleczki są jasnoczerwonym elementem i nasz wzrok dostrzega je z łatwością. Możemy spróbować więc rozbić początkowy obraz na składowe RGB, aby w składowej czerwonej (il. 14.6b) wykryć beleczki, tym bardziej że składowa zielona (il. 14.6c) i niebieska (il. 14.6d) zdecydowanie nie nadają się do tego celu.

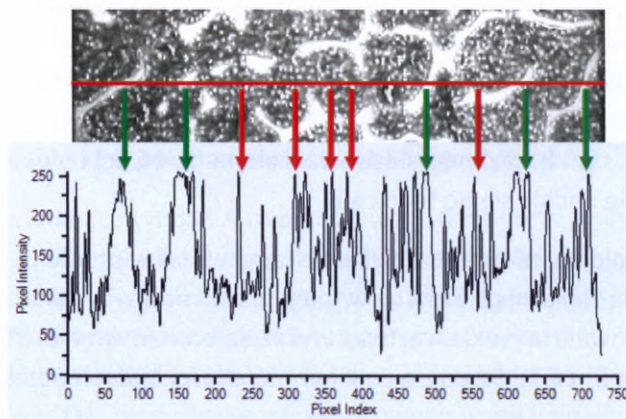
Zgodnie z wcześniejszymi sugestiami, korzystamy z analizy profilu (il. 14.7). Okazuje się, że beleczki są wyraźnie widoczne w miejscu maksimów profilu (zielone strzałki na il. 14.7), ale równie wysokie wartości wykazują odizolowane białe punkty (czerwone strzałki na il. 14.7). Są one wynikiem obecności białych obiektów (komórki



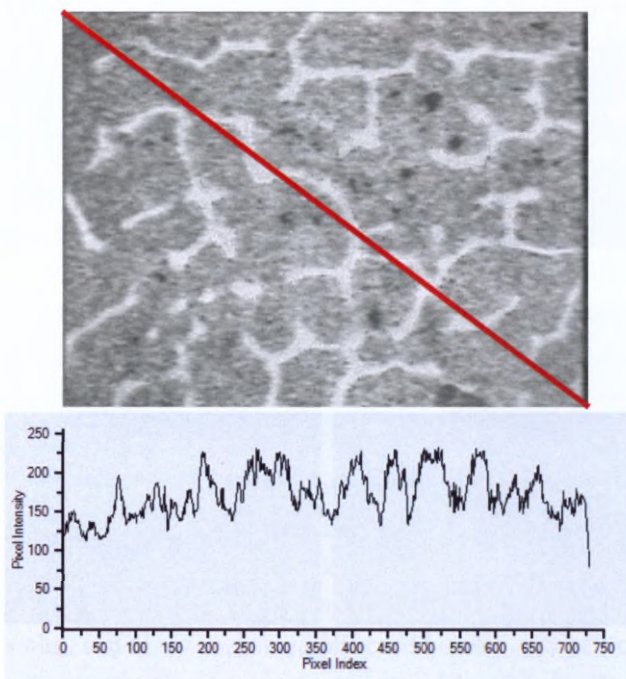
Il. 14.6. Obraz beleczek kostnych w kręgu: a) zarejestrowany obraz kolorowy, b) składowa czerwona, c) składowa zielona, d) składowa niebieska

tłuszczowe) w obrazie wyjściowym, a biały kolor zawiera wysokie wartości wszystkich składowych RGB, w tym składowej czerwonej.

Jak łatwo zauważyć (il. 14.6c), składowa zielona wykazuje również jasne plamy w miejscu białych elementów na obrazie wyjściowym, ale miejsca czerwonych beleczek kostnych pozostają ciemne. Jeżeli odejmiemy składową zieloną od czerwonej,



Il. 14.7. Binaryzacja składowej czerwonej nie daje dobrej detekcji beleczek. Opis w tekście



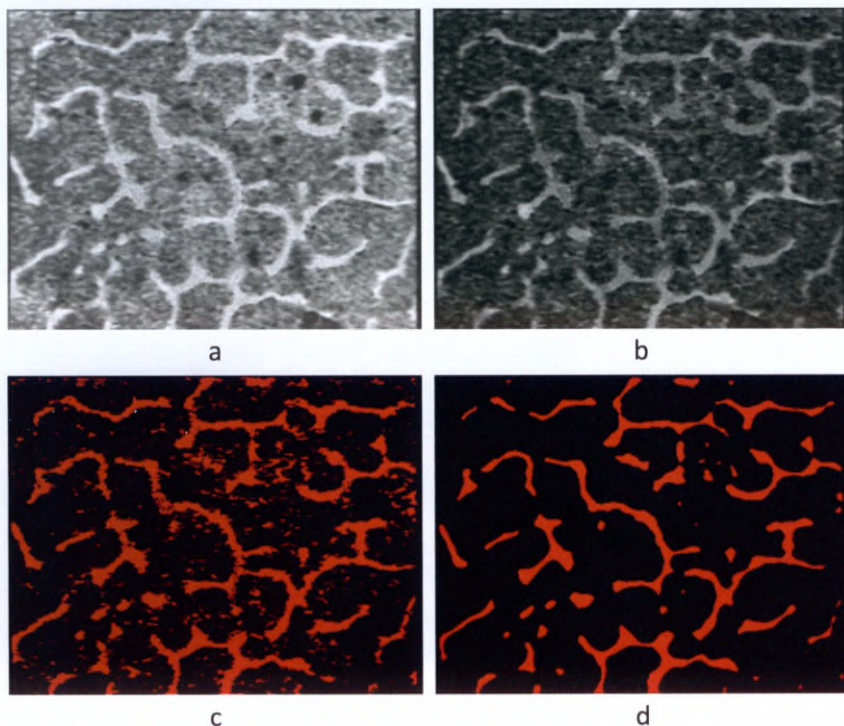
Il. 14.8. Profil wyznaczony wzdłuż czerwonej linii wskazuje na potrzebę korekcji cienia

to można spodziewać się łatwiejszego do detekcji obrazu. I tak jest w istocie (il. 14.8), ale obraz wciąż nie da dobrego wyniku detekcji. Jeżeli przeanalizujemy profil, który potwierdza swoją przydatność jako narzędzie, to stwierdzimy, że potrzebna jest korekcja cienia (wykres na dole il. 14.8).

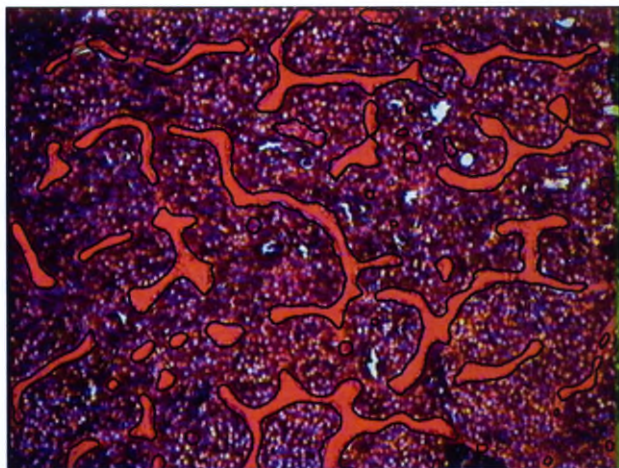
Po wykonaniu korekcji cienia jesteśmy blisko sukcesu (il. 14.9a), ale binaryzacja dalej okaże się daleka od doskonałości. Można wspomóc się detekcją lokalnych maksimów (*white top hat*, il. 14.9b). Binaryzacja tego obrazu da już akceptowalny wynik, który można skorygować np. przez filtrację medianową (il. 14.9d). Podobny wynik da złożona sekwencja operacji morfologicznych. Zatem wybieramy prostsze rozwiązanie. Użyty skany końcowy wynik (il. 14.9d) daje bardzo dobre odwzorowanie belczek (il. 14.10).

Można także wykonać binaryzację składowej czerwonej (il. 14.11a) i filtrację medianową. Wynik (il. 14.11b) będzie jednak wyraźnie gorszy od otrzymanego za pomocą przedstawionego wcześniej bardziej rozbudowanego algorytmu.

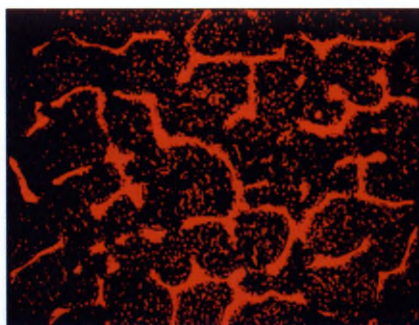
Można jeszcze inaczej... Jeżeli filtracji medianowej poddamy różnicę składowych czerwonej i zielonej (il. 14.12a), to można ją zbinaryzować. Konieczne będzie



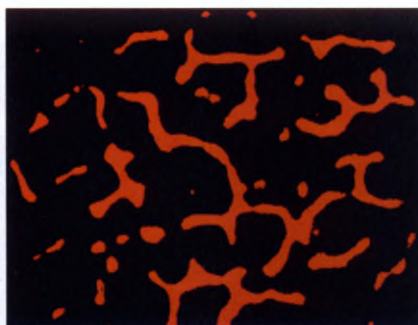
Il. 14.9. Detekcja beleczek kostnych: a) wynik korekcji cienia, b) *white top hat* obrazu (a), c) binaryzacja obrazu *white top hat*, d) mediana obrazu binarnego



Il. 14.10. Końcowy wynik detekcji nałożony na obraz początkowy

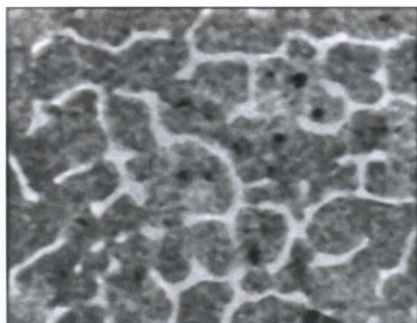


a

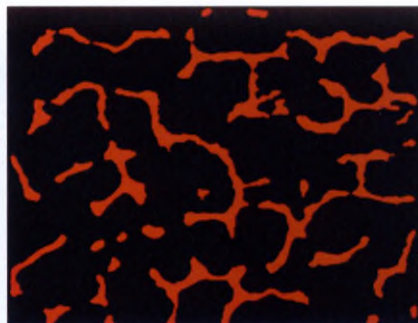


b

Il. 14.11. Detekcja beleczek kostnych: a) binaryzacja składowej czerwonej (il. 14.9b),
b) filtracja medianowa obrazu binarnego



a



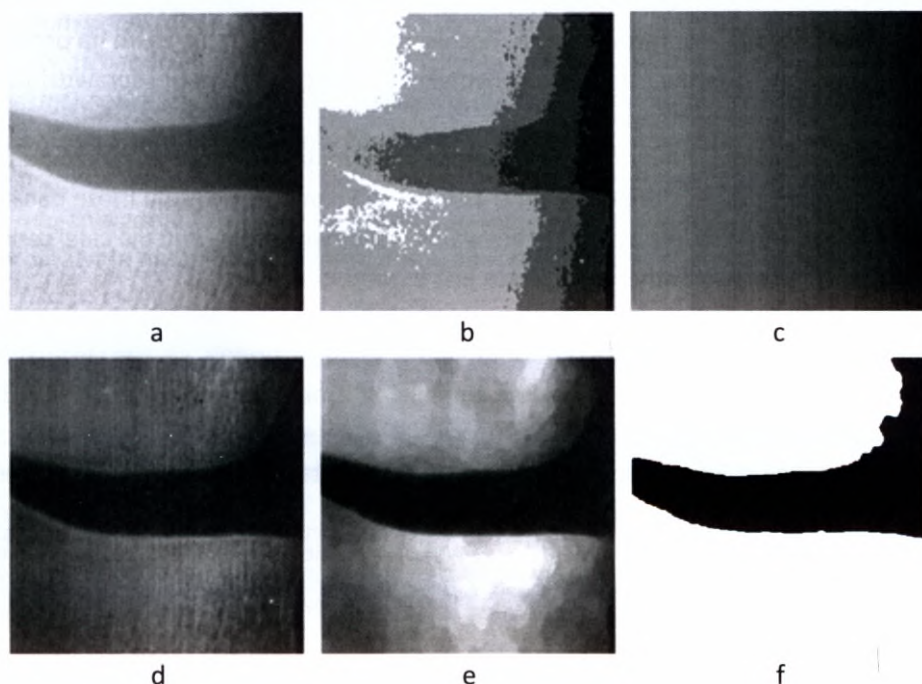
b

Il. 14.12. Różnica składowych: czerwonej i zielonej po filtracji medianowej (a)
oraz końcowy wynik detekcji. Opis w tekście

usunięcie kilku małych, odizolowanych wydzielen, co zrobimy za pomocą rekonstrukcji. Końcowy wynik ma pewne niedoskonałości (il. 14.12b), ale jest lepszy niż wynik binaryzacji składowej czerwonej (il. 14.11b).

Przedstawiona powyżej możliwość uzyskania podobnych wyników za pomocą istotnie różniących się od siebie rozwiązań jest typowa dla metod analizy obrazu. Mnogość różnych rozwiązań stanowi pewną trudność, ale jednocześnie daje możliwość ulepszania istniejących algorytmów.

Inny, ciekawy, choć w gruncie rzeczy prosty przykład konstrukcji algorytmu do detekcji obrazu klasy A pokazuje il. 14.13. Naszym celem jest automatyczna detekcja szpary stawowej w stawie kolanowym, pokazanej na zdjęciu rentgenowskim (il. 14.13a). Zadanie pozornie proste okazuje się być trudne ze względu na zmienny poziom szarości zdjęcia, wynikający ze zmiennej grubości tkanek prześwietlanych promiowaniem rentgenowskim (il. 14.13b). Wygląda to na problem wymagający korekcji cienia. Standardowe podejście w tym przypadku nie da jednak dobrych wyników. Trzeba poszukać jakiegoś nowego, nietypowego rozwiązania.



Il. 14.13. Detekcja szpary stawowej w stawie kolanowym: a) początkowy obraz rentgenowski, b) wyniki binaryzacji z różnym progiem, c) erozja pionowym elementem liniowym o rozmiarze równym wysokości obrazu, d) różnica obrazu początkowego i zerodowanego, e) zamknięcie obrazu po erozji, f) końcowy wynik

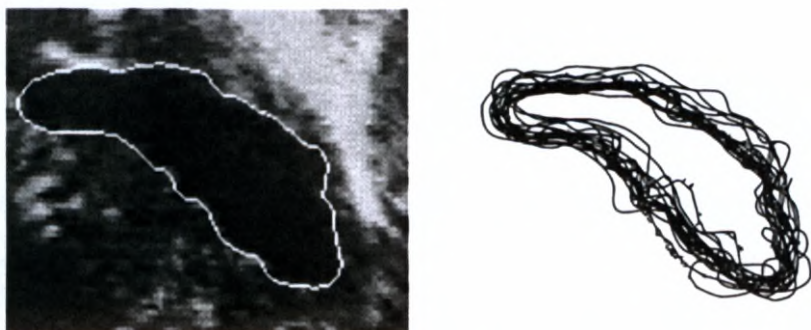
Szpara stawowa jest najciemniejszym elementem obrazu na il. 14.13a, a poziom szarości zmniejsza się systematycznie w kierunku poziomym. Możemy uzyskać obraz tego ciemnego tła po prostu przez erozję pionowym liniowym elementem strukturalnym o wielkości odpowiadającej wysokości obrazu (il. 14.13c). Po odjęciu tego tła od obrazu wyjściowego jesteśmy już blisko końcowego sukcesu (il. 14.13d). Można próbować od razu przeprowadzić binaryzację. Jeżeli chcemy zabezpieczyć się przed ewentualnymi drobnymi artefaktami, można wykonać zamknięcie (il. 14.13e). Binaryzacja prowadząca do końcowego wyniku będzie już tylko formalnością (il. 14.13f).

Często obrazy binarne po binaryzacji wymagają jeszcze korekty. Mogą mieć otwory lub drobne obiekty, wynikające z obecności szumów i artefaktów, poszarpane brzegi, sklejone obiekty itp. Możliwości korekty tych błędów zostały omówione podczas analizy przekształceń morfologii matematycznej. Zwykle (co potwierdził nawet przykład z beleczkami kostnymi) lepiej jest dokonywać końcowej korekty na obrazie binarnym niż na obrazie szarym przed binaryzacją. Trzeba tutaj poszukiwać przysłowiowego złotego środka, bo dobra binaryzacja bez żadnej korekty obrazu szarego zwykle też się nie udaje.

W przypadku obrazów odpowiadających typowi B (il. 14.5b) pozostają w mocy wszystkie uwagi dotyczące obrazów typu A. Różnica polega przede wszystkim na potrzebie innego sposobu doprowadzenia do końcowej binaryzacji. Najczęściej sprawdzi się następujący schemat postępowania:

- Jeżeli obraz wyjściowy nie jest obrazem o dobrej jakości, można zacząć od prób eliminacji szumów i poprawy kontrastu (wyrównanie histogramu może dać kiepski efekt wizualny, ale ułatwi obracowanie wspólnego algorytmu dla całej serii podobnych obrazów), choć nie zawsze jest to niezbędne.
- Warto sprawdzić profil, co pomoże w poszukiwaniu odpowiednich narzędzi do przetworzenia lub korekty obrazu.
- Często dobre wyniki daje korekcja cienia. Czasami może być jednak niemożliwa do wykonania.
- Do detekcji obiektów dochodzimy często na podstawie binaryzacji, podobnie jak w przypadku obrazów typu A.
- Bardzo często detekcję prowadzimy z wykorzystaniem albo detekcji lokalnych ekstremów (*top hat*), albo detekcji działów wodnych.
- Analogicznie jak w obrazach typu A, zazwyczaj potrzebna jest korekta otrzymanych obrazów binarnych.

Skuteczność tego typu postępowania pokażemy na przykładzie detekcji prawej komory serca na ultrasonograficznych obrazach serc noworodków (il. 14.14). Zarys wyznaczony przez komputer nie budzi wątpliwości, że jest poprawny. Natomiast



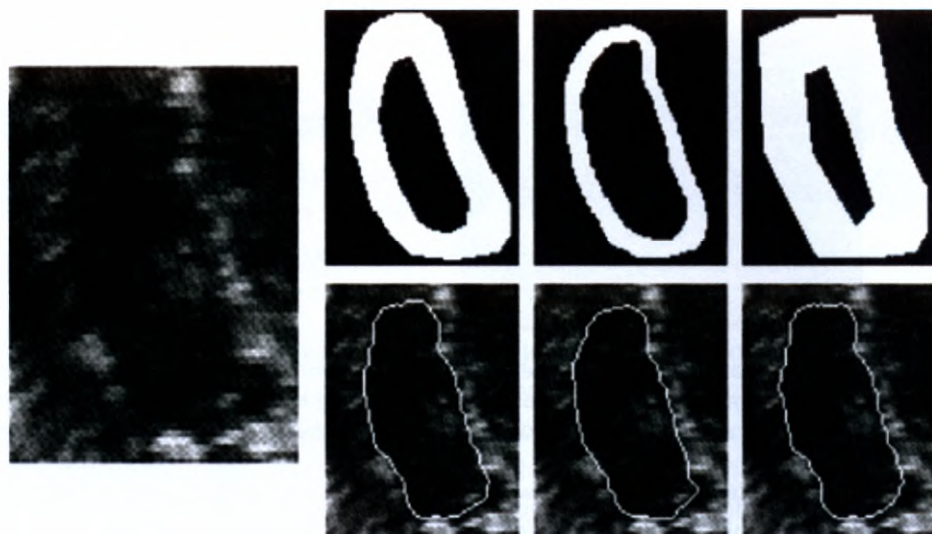
Il. 14.14. Obraz ultrasonograficzny prawej komory serca noworodka z naniesionym zarysem tej komory, wyznaczonym za pomocą programu analizy obrazu (z lewej), oraz 14 zarysów tej samej komory, wyznaczonych manualnie przez studentów czwartego roku medycyny

zadanie polegające na manualnym wyznaczeniu zarysu tej komory przez 14 studentów czwartego roku medycyny zakończyło się bardzo dużym rozrzutem wyników (pole powierzchni największego obiektu jest dwukrotnie większe od pola powierzchni najmniejszego obiektu).

W takim przypadku skuteczne okazuje się wykorzystanie detekcji działów wodnych. Ponieważ badania, których fragment tutaj przedstawiamy, ograniczały się do kilkunastu obrazów, nie było uzasadnienia dla tworzenia skomplikowanej procedury automatycznego generowania znaczników. Wykorzystano zatem znaczniki rysowane przez człowieka. Okazało się przy tym, że jeżeli znaczniki spełniały podstawowe wymagania zdefiniowane przez lekarza, to kształt i wielkość wykrytych komór serca pozostawała niemal niezmienna, mimo znacznych różnic w znacznikach (il. 14.15).

Najtrudniejsze do opracowania, a jednocześnie najbardziej spektakularne w przypadku sukcesu, są algorytmy do detekcji obiektów w obrazach typu C. W tym przypadku nie ma jednego uniwersalnego sposobu przetwarzania obrazu oraz detekcji obiektów. Jak zwykle punktem wyjścia może być próba binaryzacji oraz analiza profilu poziomów szarości.

Następnym krokiem powinno być zastanowienie się, na co zwracamy uwagę podczas obserwacji tych obrazów. Na tej podstawie można starać się wykryć te elementy, które sprawiają, że doszukujemy się w tych miejscach granic obiektów. Jeżeli ustalimy, na co zwraca uwagę nasz zmysł wzroku, to można następnie starać się zaprogramować ten proces. Zazwyczaj w takiej sytuacji szczególnie potrzebny jest wspomniany już wcześniej przełomowy pomysł, który sprawi, że rozwiązanie okaże się skuteczne. Dobrze jest skupić się na myśleniu, co potrzebujemy zrobić, ale bez nawiązywania do znanych nam przekształceń, bo to ograniczy naszą wyobraźnię.



Il. 14.15. Prawa komora serca noworodka (z lewej strony) oraz trzy różne znaczniki i odpowiadające im bardzo zbliżone wyniki detekcji działów wodnych

Liczba możliwych rozwiązań jest w tym przypadku praktycznie nieograniczona. Wiele osób zastanawia się, jak można nauczyć się rozwiązywać tego typu problemy. Niestety, niemal wszystkie podręczniki poruszające zagadnienia analizy obrazu omawiają po prostu pojedyncze przekształcenia. Nawet jeżeli stosują je w odniesieniu do jakiegoś konkretnego problemu, to nie wiadomo, w jaki sposób doprowadzono do obrazów, które wykorzystano do zilustrowania możliwości danej procedury. Można skorzystać z dostępnej bez ograniczeń na stronie internetowej Politechniki Krakowskiej książki *Praktyka analizy obrazu* lub opracowania J. Szali *Zastosowanie metod komputerowej analizy obrazu do ilościowej oceny struktury materiałów* (patrz Literatura). Zostały one, co prawda, opracowane dość dawno temu, ale pewne zasady nie ulegają zmianie, tak jak nie ulegają zmianie podstawy rachunku różniczkowego, fizyki czy też analizy statystycznej. Są to książki unikalne pod tym względem, że omawiają całe algorytmy przetwarzania obrazu. Zostały pomyślane jako swego rodzaju „książki kucharskie”. Trzeba znaleźć w nich obraz przypominający nasz problem i zobaczyć, jak został on rozwiązany. Opisane w książkach rozwiązanie może stanowić punkt wyjścia dla naszych własnych poszukiwań.

Zadaniem niniejszego rozdziału jest jedynie pokazanie pewnego podejścia do budowy algorytmów analizy obrazów. Nie jest on wystarczający do nauki ich projektowania. W tym celu należy analizować demonstracyjne rozwiązania, udostępniane przez producentów oprogramowania. Pomocny powinien być też drugi podręcznik,

autorstwa Adama Piwowarczyka, który był opracowywany równolegle, właśnie z myślą o nauce praktycznego wykorzystania wiedzy z zakresu podstaw analizy obrazu.

Należy pamiętać, że nawet w przypadku wykorzystania uczenia maszynowego analizowane obrazy powinny być wstępnie przygotowane do tego procesu. Często skutecznym rozwiązaniem jest użycie sztucznej inteligencji dopiero do analizy wyników pomiarów. Aby jednak doprowadzić do tego etapu, konieczne jest opracowanie odpowiedniego algorytmu i dlatego zagadnienie to jest bardzo ważne.

15. Podsumowanie

Zawarty w książce materiał obejmuje przegląd całości podstawowych narzędzi używanych w komputerowej analizie obrazu. Dziedzina ta jest jednak zbyt obszerna, aby na nieco ponad 350 stronach zawrzeć wszystkie informacje. Konieczne były więc decyzje, co wybrać, a co pominąć. Pomińnięto zostały między innymi następujące elementy, które wydawały się mniej istotne do zrozumienia istoty analizy obrazu:

- Przy omawianiu akwizycji obrazów pominięto zagadnienia zdolności rozdzielczej mikroskopów świetlnych i wynikających z niej zasad doboru wielkości rejestrowanego obrazu.
- Pominięto również wszelkie rozważania dotyczące parametrów kamer, gdyż bardzo szybko stają się one nieaktualne.
- Podczas omawiania filtrów zrezygnowano z szerszego omawiania filtrów definiowanych przez użytkownika.
- W ramach binaryzacji nie omówiono wszystkich istniejących metod binaryzacji, ograniczając się do najpopularniejszych.
- W przypadku operacji morfologicznych zrezygnowano z opisu niektórych złożonych operacji, które są sekwencją prostszych operacji, opisanych w całości w książce.
- W rozdziale dotyczącym obrazów kolorowych nie pokazano wszystkich metod binaryzacji, opartych na odległości pomiędzy poszczególnymi kolorami.
- W zaawansowanych metodach analizy obrazu ograniczono się do trzech grup przekształceń.
- Wprowadzono elementy stereologii, gdyż wiele osób nie zna ich i przez to popełnia błędy podczas analizy wyników. Ograniczono się jednak do podstaw i narzędzi, które można łatwo zaimplementować w komputerowej analizie obrazu.
- W przypadku pomiarów pominięto niemal w całości zagadnienia dalszej obróbki danych pomiarowych, zwłaszcza wykorzystujących wirtualne obiekty, gdyż nie są to zagadnienia podstawowe.
- W zakresie analizy obrazów 3D ograniczono się do zasygnalizowania problemu. Jego głębsza analiza wymaga bliskiego kontaktu z osobami dysponującymi sprzętem do akwizycji obrazów 3D i poznania specyfiki poszczególnych zastosowań.
- Budowa algorytmów została ograniczona do podania bardzo ogólnych wskazówek. W kilku miejscach omówiono dokładniej sekwencję przekształceń, ale celem książki jest wyjaśnienie, jak to działa, a nie nauka programowania systemów

analizy obrazu. Niemniej entuzjaści pracy z komputerem mogą dojść do bardzo dobrych wyników już po lekturze tej książki.

Skoro tak wiele opuszczono, to co znalazło się na kartach tej książki? Przede wszystkim zawiera ona dość dokładne omówienie poszczególnych operacji. Opisy są ścisłe i kompletne, a jednocześnie napisane w przystępny sposób. Bardzo ważną rolę pełnią ilustracje. W przypadku każdego rozdziału najpierw opracowywano ilustracje, co pozwalało na wizualizację treści, a dopiero w drugiej kolejności dodawany był tekst. Dzięki temu strona wizualna jest bardzo silnie zintegrowana z treścią. Duża liczba ilustracji powinna ułatwić lekturę. W większości przypadków ilustracje mają duże rozmiary, dzięki czemu powinny być bez trudu odczytane. Niemal wszystkie zdjęcia i wszystkie rysunki zostały wykonane przez autora, co zapewnia spójność z treścią książki.

16. Literatura

Handbook of 3D Machine Vision. Optical Metrology and Imaging, (Ed.) S. Zhang, CRC Press, Boca Raton, FL, 2013.

Hybrid Intelligence for Image Analysis and Understanding, (Eds.) S. Bhattacharyya, I. Pan, A. Mukherjee, P. Dutta, Wiley, Hoboken, JN, 2017.

Jähne B., *Practical Handbook on Image Processing for Scientific Applications*, CRC Press, Boca Raton, FL, 1997.

Kurzydłowski K.J., Ralph B., *The quantitative description of the microstructure of materials*, Boca Raton, FL, 1995.

Piwowarczyk A., *Analiza obrazu. Jak to się robi?*, Politechnika Krakowska, Kraków 2020 (w opracowaniu).

Practical Guide to Image Analysis, ASM international, Materials Park, OH, 2000.

Pratt W.K., *Introduction to Digital Image Processing*, CRC Press, Taylor & Francis Group, Boca Raton, FL, 2014.

Russ J.C., *The Image Processing Handbook*, Second Edition, CRC Press, Boca Raton, FL, 1995.

Ryś J., *Stereologia materiałów*, Fotobit-Degign, Kraków 1995.

Soille P., *Morphological Image Analysis. Principles and Applications*, Second Edition, Springer-Verlag, Berlin – Heidelberg 2003.

Sonka M., Hlavac V., Boyle R., *Image Processing, Analysis and Machine Vision*, Second Edition, Brooks/Cole Publishing Company, Pacific Grove, CA, 1999.

Szala J., *Zastosowanie metod komputerowej analizy obrazu do ilościowej oceny struktury materiałów*, Zeszyty Naukowe Politechniki Śląskiej, Hutnictwo, z. 61, Gliwice 2001.

Tadeusiewicz R., Korohoda P., *Komputerowa analiza i przetwarzanie obrazów*, Wydawnictwo Fundacji Postępu Telekomunikacji, Kraków 1997.

The Essential Guide to Image Processing, (Ed.) A.C. Bovik, Academic Press, Elsevier, 2009.

Underwood E.E., *Quantitative Stereology*, Addison-Wesley, Reading, Massachusetts, 1970.

Wojnar L., *Image analysis. Applications in materials engineering*, CRC Press, Boca Raton, FL, 1999.

Wojnar L., Kurzydłowski K.J., Szala J., *Praktyka analizy obrazu*, Polskie Towarzystwo Stereologiczne, Kraków 2002 (książka dostępna w wersji elektronicznej na stronie internetowej Biblioteki Politechniki Krakowskiej).

Dokumentacja systemu Aphelion, www.adcis.net.

40,00





Biblioteka Politechniki Krakowskiej



A-1154

Archiwum

Prof. dr hab. inż. **Leszek Wojnar**
- absolwent Politechniki
Krakowskiej oraz AGH w Krakowie,
profesor inżynierii materiałowej.
Popularyzator wiedzy, specjalista
z zakresu wykorzystania
komputerowej analizy obrazu.
Autor wydanej w USA książki
pt. *Image analysis. Applications in
materials engineering*.
Wieloletni pracownik
Politechniki Krakowskiej,
dziekan Wydziału Mechanicznego
w latach 2008-2016.
Pasjonat oraz instruktor
narciarstwa, żeglarz regatowy.
Niespokojny duch, łączący cechy
romantyka i artysty
z pragmatycznym, technicznym
rozwiązywaniem problemów.
Grafik, rysownik i fotograf.
Autor satyrycznych rysunków
komentujących absurdy otaczającej
nas rzeczywistości oraz tomiku
poezji pt. *Uklucie uczucia*.
Znany z niekonwencjonalnych
pomysłów, ciętego języka
i upartej walki o wyznawane ideały.

Biblioteka Politechniki Krakowskiej



100000337767

ISBN 978-83-66531-52-9