

Rozpoznawanie obrazów: Wprowadzenie

Krzysztof Krawiec

Przetwarzanie i Rozpoznawanie Obrazów

February 3, 2015

Kwestie organizacyjne

- Nazwa przedmiotu: *Przetwarzanie i rozpoznawanie obrazów*
- Semestr 1 studiów II stopnia, Inteligentne Systemy Wspomagania Decyzji
- Wymiar: 2W, 2L
- Zaliczenie Lab + Egzamin
- Strona WWW przedmiotu: <http://www.cs.put.poznan.pl/kkrawiec/>
- Prowadzący:
 - dr hab. inż. K.Krawiec, krawiec@cs.put.poznan.pl
 - dr inż. Bartosz Wieloch, bwieloch@cs.put.poznan.pl
- Wykład: prezentacja zagadnień teoretycznych i praktycznych przeplatana studiami przypadków.
- Powiązania z przedmiotami: Uczenie Maszynowe

Przykładowe pojęcia których znajomość może pomóc:

- Metoda najmniejszych kwadratów
- Transformata Fouriera
- Splot (convolution)
- Algorytm A^*
- Analiza głównych składowych (PCA)
- Dodawanie Minkowskiego

Główne punkty ciężkości wykładu

- teoria: przetwarzanie jakościowe, analiza, opis: ekstrakcja cech, rozpoznawanie, adaptacja i uczenie,
- praktyka: aspekty informatyczne: algorytmy i struktury danych, złożoność obliczeniowa i pamięciowa, stopień trudności implementacji, zastosowania.

Literatura podstawowa:

1. Gonzalez, Wintz, Digital Image Processing. Addison-Wesley 2008 (lub wcześniejsze wydania)
2. Domański, M., Obraz cyfrowy. WKŁ 2010.
3. Zieliński, T.P., Cyfrowe przetwarzanie sygnałów. WKŁ 2009.

Literatura uzupełniająca:

1. Cyganek, B., Komputerowe przetwarzanie obrazów trójwymiarowych. EXIT 2002.
2. Stąpor, K., Automatyczna klasyfikacja obiektów. EXIT 2005.
3. Mark Owen, Przetwarzanie sygnałów w praktyce. WKŁ 2009.
4. Choraś, R. Komputerowa wizja. Metody interpretacji i identyfikacji obiektów. EXIT, 2006.
5. Watkins, C.D., Sadun, A., Marenka, S.: Nowoczesne metody przetwarzania obrazu. WNT 1995.
6. Wojciechowski, K., Rozpoznawanie obrazów, Politechnika Śląska, Gliwice 1997.

Literatura starsza i uzupełniająca/powiązana

W języku polskim:

- Pavlidis, T., Grafika i przetwarzanie obrazów. Algorytmy, (Seria: Biblioteka Inżynierii Oprogramowania), WNT, Warszawa 1987.
- Tadeusiewicz, R., Systemy wizyjne robotów przemysłowych, WNT, Warszawa 1992.
- Tadeusiewicz, R., Flasiński, M., Rozpoznawanie obrazów, (Seria: Współczesna Nauka i Technika. Informatyka) PWN, Warszawa 1991.
- Ostrowski, M. (red.), Informacja obrazowa. WNT, Warszawa, 1992.
- Nałęcz, M., Problemy biocybernetyki i inżynierii biomedycznej, WKiŁ, Warszawa 1990.
- Nieniewski, M., Morfologia matematyczna w przetwarzaniu obrazów, Akademicka Oficyna Wydawnicza, Warszawa 1998.
- Zieliński, K.W., Strzelecki, M.: Komputerowa analiza obrazu biomedycznego. Wstęp do morfometrii i patologii ilościowej, Wydawnictwo Naukowe PWN, 2002.

W języku angielskim:

- Duda, R., Hart, P.: Pattern Classification and Scene Analysis, Wiley,

- Biblioteka OpenCV
- ImageJ
- ImageMagick
- Matlab / Octave
- Pakiet R

Wprowadzenie

Przecięcie:

- Sztucznej inteligencji (ang. Artificial Intelligence),
- Inteligencji obliczeniowej (ang. Computational Intelligence),
- Przetwarzania sygnałów (ang. Signal Processing).

Inne powiązane dziedziny:

- grafika komputerowa (ang. Computer Graphics)
- uczenie maszynowe (ang. Machine Learning)
- psychologia poznawcza (ang. Cognitive Science)

Próba definicji dziedziny

- Fukunaga (1972) „Pattern recognition consists of two parts, feature selection and classifier design”
- Duda & Hart (1973) „Pattern recognition, a field concerned with machine recognition of meaningful regularities in noisy or complex environments”
- Gonzalez and Thomason (1978) „Pattern recognition can be defined as the categorization of input data into identifiable classes via the extraction of significant features or attributes of the data from a background of irrelevant detail”
- Bezdek (1981) „Pattern recognition is a search for structure in data”
- Schalkoff (1992) „Pattern recognition is the science that concerns the description or classification (recognition) of measurements”

Uwarunkowania historyczne:

- Problemy terminologiczne: w tradycji PR “pattern” – obraz
- W ramach tego wykładu: Rozważamy obrazy w sensie wizyjnym - wykluczamy np. przebiegi czasowe
- ‘Rozwód’ Sztucznej Inteligencji i Rozpoznawania obrazów w latach

- > 80% informacji wpływa do mózgu drogą wzrokową
- “cywilizacja wizualna”: multimedia, Internet,
- rosnąca dostępność sprzętu,
- coraz szersze spektrum potencjalnych i realnych zastosowań, zwłaszcza w zadaniach/aktywnościach uciążliwych i/lub trudnych dla człowieka,
- rosnąca dostępność gotowych rozwiązań programistycznych (np. API MS Kinect).

Klasy trudności praktycznych zadań rozpoznawania obrazów:

- T_0 : zadania trudne
- T_1 : zadania bardzo trudne
- T_2 : zadania za trudne
- T_∞ : zadania 'zapomnij. . . '

Dlaczego to jest trudne?

- różne formy niedoskonałości obrazu: niepełność/niepewność informacji wynikająca z:
 - niedoskonałości procesu akwizycji obrazu (szumy, zniekształcenia optyczne, zniekształcenia chromatyczne, etc.) – może podlegać zmniejszaniu w wyniku udoskonaleń technicznych
 - natury procesu akwizycji obrazu (mapowanie 3D->2D, oświetlenie sceny) – z tym trzeba się pogodzić
- konieczność uwzględnienia wiedzy
 - dziedzinowej (ang. domain knowledge)
 - zdroworozsądkowej (ang. common sense knowledge) zwłaszcza w uczeniu z informacji obrazowej.
- znaczna ilość przetwarzanej informacji => złożoność obliczeniowa/pamięciowa procesów związanych z analizą/rozpoznawaniem

Dlaczego to jest trudne?

Despite this significant level of activity, the field [of image recognition and interpretation] remains a challenge. In particular, solutions of problems in image analysis are characterized by task-specific formulations, thus limiting the capability for advancement using the time-tried method of building a generalized body of results based on preceding accomplishments. For the foreseeable future, the design of image analysis systems will continue to require a mixture of art and science. (Gonzalez & Woods, str.657)

With no effort we scan a scene by directing our gaze at specific objects, discerning them individually despite the background of other objects, contours, shadows, and changes in illumination. The process is partially intentional, partially automatic, and entirely amazing: no machine can accomplish this, but the simplest insect can. A single glance captures megabytes of data; we reduce this flood by singling out specific objects for attention (...)

(Hemmen, Domany, Cowan, 2001)

- przetwarzanie sygnałów (ang. Signal Processing): obejmuje podejścia i metody modelujące obraz jako dwu- lub więcej-wymiarowy obraz w dziedzinie przestrzeni,
- widzenie komputerowe (ang. Computer Vision): zorientowana na analizę scen i obiektów 3D
- widzenie maszynowe (ang. Machine Vision): dotyczy głównie zastosowań przemysłowych, głównie kontrola jakości (QC)
- rozumienie obrazów (ang. Image Understanding): koncentruje się na interpretacji *całych* scen

Schemat wnioskowania z informacji obrazowej

Źródło promieniowania	Lighting
Scena (obrazowany obiekt)	Scene
<i>Oddziaływanie fizyczne</i>	<i>Reflection/transmission/emission</i>
Obraz analogowy (fizyczny)	Analog image
<i>Pozyskiwanie obrazu</i>	<i>Image acquisition</i>
Obraz cyfrowy	Digital image
<i>Przetwarzanie obrazu</i>	<i>Image processing</i>
Obraz cyfrowy	Digital image
<i>Analiza obrazu (ekstrakcja cech)</i>	<i>Image analysis (feature extraction)</i>
Opis obrazu (np. wektor cech)	Image/Scene description (feature vector)
<i>Interpretacja/podejmowanie decyzji</i>	<i>Interpretation/decision making</i>
Rozpoznanie/decyzja/pomiar	Recognition/decision/measurement

Typowe zadania przetwarzania i analizy obrazów (rozumiane pod kątem funkcjonalnym):

- Rozpoznawanie (ang. recognition)
- Identyfikacja (ang. identification)
- Analiza obrazu (ang. image analysis) - ekstrakcja innej informacji z obrazu (niekoniecznie przypominającej cechy używane przez człowieka)
- Polepszanie jakości obrazu dla potrzeb dalszej analizy
 - zorientowane na percepcję człowieka (ang. image enhancement),
 - zorientowane na możliwe wierne odtworzenie obrazu fizycznego (ang. image restoration)

Zadania typowe dla obrazów scen 3D:

- Interpretacja scen (ang. scene interpretation)
 - W tym: rozumienie obrazu (ang. image understanding) – konieczny udział reprezentacji symbolicznej
- Estymacja położenia (ang. pose estimation),
- Zadania typowe dla sekwencji wideo:
 - Detekcja ruchu (ang. motion detection),
 - Śledzenie obiektów (ang. object tracking),
 - Nawigacja (ang. navigation)

Kryteria oceny systemów RO:

- złożoność obliczeniowa i pamięciowa,
- jakość obrazu (wizualizacja)
- trafność (rozpoznawania),
- czułość, specyficzność (vide krzywa ROC)
- dokładność (np. lokalizacji),
- niezawodność,
- odporność na dane (obrazy) zaszumione/zniekształcone/uszkodzone,
- niezmienniczość T,S,R (+3D).

Zastosowania w naukach podstawowych:

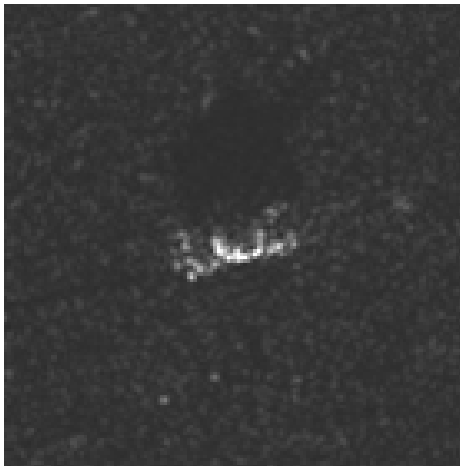
- Kartografia, geografia, geologia, meteorologia: analiza zdjęć satelitarnych i lotniczych
- Medycyna, np. badania przesiewowe zdjęć rentgenowskich, analiza obrazów mikroskopowych,
- Archeologia: odtwarzanie zniszczonych eksponatów,
- Astronomia: odsiewanie artefaktów i automatyczna pre-klasyfikacja ciał niebieskich,
- Fizyka: polepszanie jakości obrazów eksperymentów w fizyce plazmowej i mikroskopii elektronowej,

- Bioidentyfikacja: klasyfikacja i identyfikacja na podstawie: odcisków palców, obrazu tęczówki oka, obrazu siatkówki oka, kształtu dłoni, sposobu chodzenia, odcisku ust, i innych cech biometrycznych
- Bezpieczeństwo: wspomaganie monitorowania obiektów (CCTV), wykrywanie zapalników w prześwietleniach bagażu, sterowanie ruchem miejskim,
- Przemysł: kontrola jakości produktów,
- Robotyka (w tym sterowanie pojazdami),
- Rozpoznawanie znaków: maszynowych i ręcznie pisanych (on/off-line),
- Bazy danych: realizacja zapytań w obrazowych,
- Rozrywkowe: np. Sony AIBO dog robot
- wojskowe,

Obecnie dynamicznie rozwijające się nurty zastosowań:

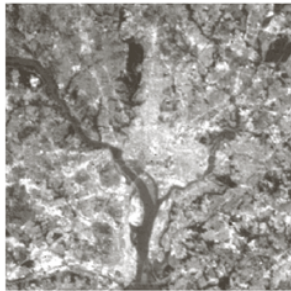
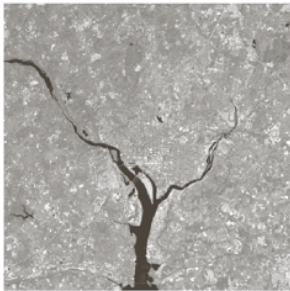
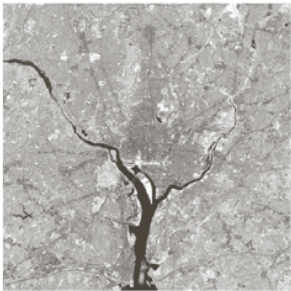
- Systemy monitorowania
- Kontrola jakości
- Medycyna
- Zastosowania wojskowe



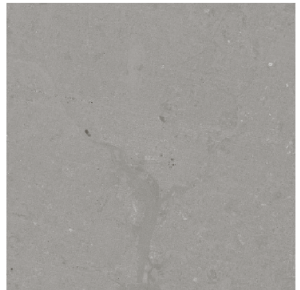
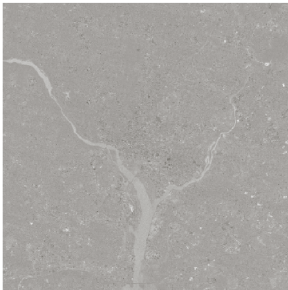
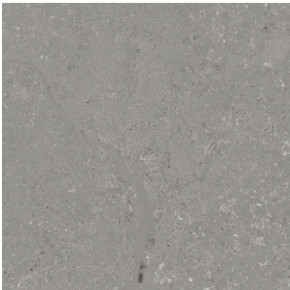
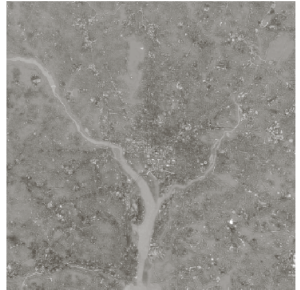




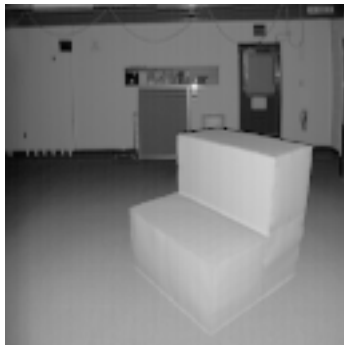
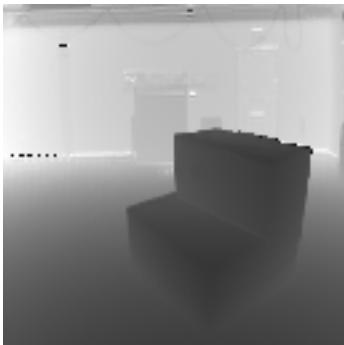
Przykładowe obrazy



Przykładowe obrazy



Przykładowe obrazy



Przykładowe obrazy



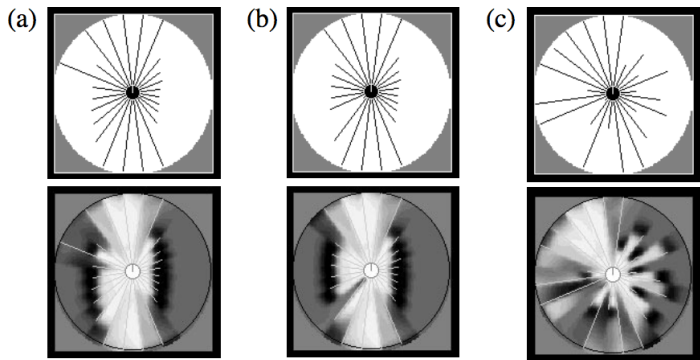
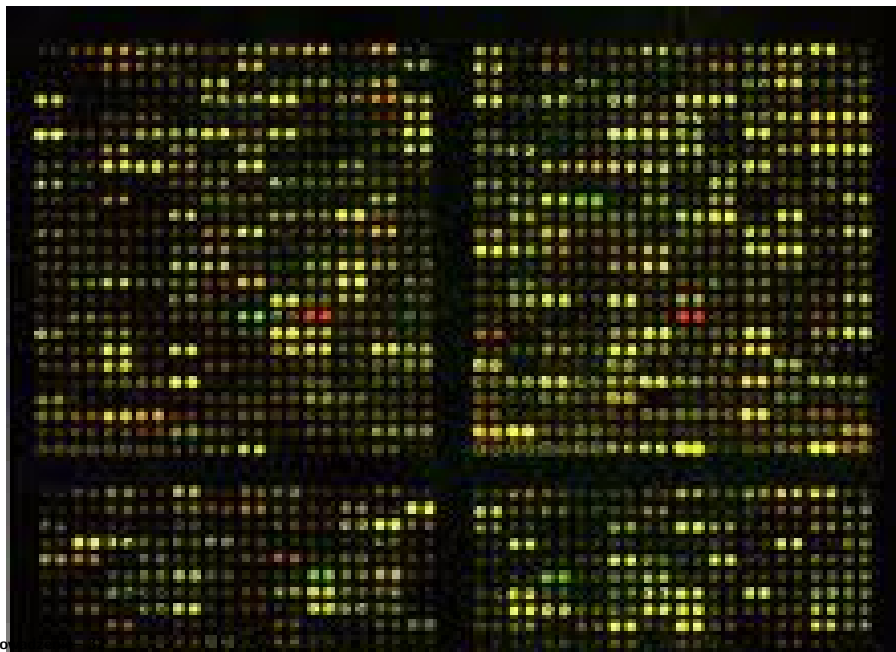


Figure 2: Sensor interpretation: Three example sonar scans (top row) and local occupancy maps (bottom row), generated by the neural network.

Przykładowe obrazy



- Głównie programowe; sprzętowe dominowały do późnych lat 80tych, obecnie tylko w zastosowaniach o wysokich wymaganiach dot. prędkości przetwarzania.
- Przerzucenie się na oprogramowanie zredukowało koszty i "ustawia punkt wyjścia w rozwiązywaniu różnych problemów dużo dalej niż przed dekadą" ([GW] s. 16).
- Nowe architektury obliczeniowe: np. karty graficzne (GPU, GPGPU)

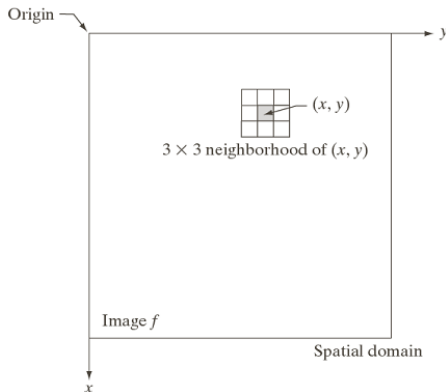
Metody cyfrowej reprezentacji obrazów

Obraz rastrowy

Obraz rastrowy (ang. *raster image*, *~bitmap*)

Podstawowe właściwości:

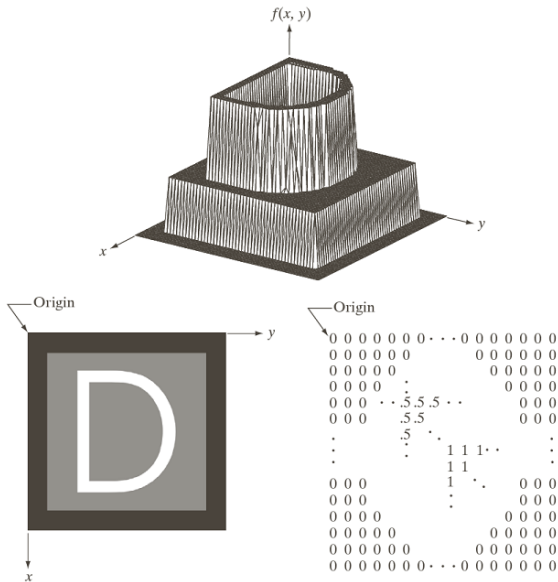
- szerokość (*width*)
- wysokość (*height*)
- [głębokość (*depth*)]



Zamienne interpretacje:

- macierz,
- funkcja f zdefiniowana na kartezjańskim układzie współrzędnych (tzw. funkcja charakterystyczna)
 - dyskretna w dziedzinie i przeciwdziedzinie,
 - stanowi pewne przybliżenie obrazu rzeczywistego (fizycznego),

Zamienne reprezentacje: macierz i funkcja



- 2D (planarne)
 - Pojedynczy element dziedziny: *pixel*, pel = picture element, *piksel*, *punkt obrazu*
 - $f(x, y)$
- 3D
 - $f(x, y, z)$: Pojedynczy element dziedziny: voxel = volume element,
 - $f(x, y, t)$: Sekwencje wideo
- 4D, np. $f(x, y, z, t)$, np. sekwencje z fMRI (czynnościowy magnetyczny rezonans jądrowy)
- Uwaga: W przeciwieństwie do ścisłego ujęcia matematycznego, pojedynczy punkt dziedziny f (np. (x, y)) ma niezerowe rozmiary.

Liczba kanałów:

- 1: wartość skalarna:
 - przeciwdziedzina binarna: obraz *czarno-biały* (*bitmapa*)
 - przeciwdziedzina wielowartościowa: obraz *monochromatyczny* (*greyscale*).
- >1 : wektor:
 - obraz wielokanałowy; najczęściej kolorowe (np. RGB),
 - obraz wielomodalny (np. tomograf MRI, obrazy satelitarne),

Przedział wartości:

- dogodny technicznie: $[0, 255]$, wartości dyskretne
- dogodny matematycznie: $[0, 1]$, wartości dyskretne lub ciągłe
- inne, np. $[0, 4095]$

Alternatywnie przedział wartości charakteryzuje się czasami liczbą bitów przeznaczonych na przechowanie informacji o pojedynczym pikselu (*depth* lub bpp, ang. *bits per pixel*):

- 1bpp: obrazy binarne, bitmapy
- 8bpp: obrazy monochromatyczne (jednokanałowe)
 - czasami także 12 bpp lub 14 bpp
- 24bpp: obrazy barwne (RGB, $3 \times 8\text{bpp} = 24\text{bpp}$)
- 32 bpp i więcej: dodatkowe kanały

Dwa procesy dyskretyzacji:

- Dyskretyzacja przestrzenna
- Dyskretyzacja wartości sygnału

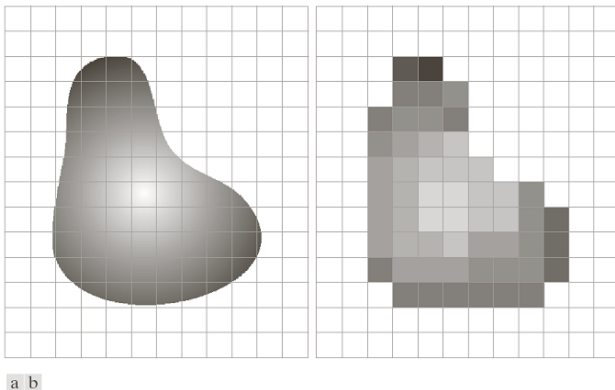
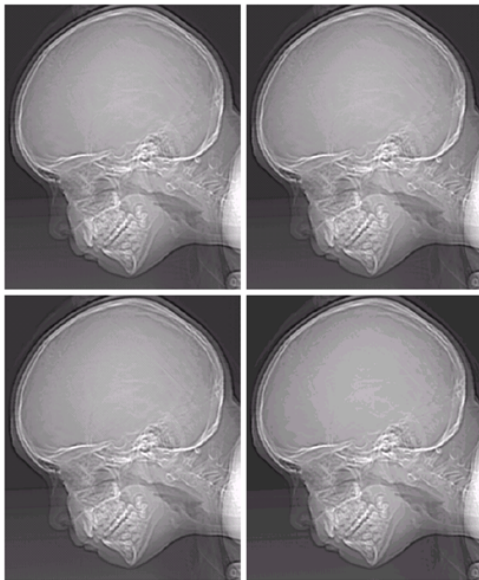


FIGURE 2.17 (a) Continuous image projected onto a sensor array. (b) Result of image sampling and quantization.

Artefakty dyskretyzacji: fałszywe kontury



a b
c d

FIGURE 2.21

(a) 452×374 , 256-level image. (b)–(d) Image displayed in 128, 64, and 32 gray levels, while keeping the spatial resolution constant.

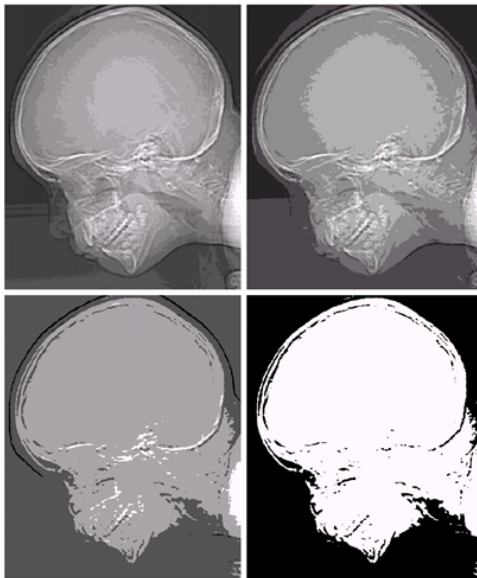
Artefakty dyskretyzacji: fałszywe kontury

c f
g h

FIGURE 2.21

(Continued)

(e)–(h) Image displayed in 16, 8, 4, and 2 gray levels. (Original courtesy of Dr. David R. Pickens, Department of Radiology & Radiological Sciences, Vanderbilt University Medical Center.)



Jakość obrazu a percepcja

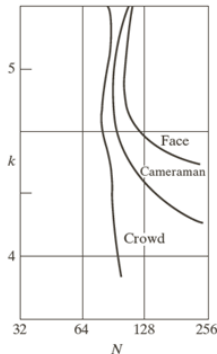
Jak obiektywne parametry obrazu (liczba bitów na piksel k i rozmiar/rozdzielczość N) przekładają się na postrzegana jakość obrazu?

- Krzywe izopreferencji
- Konkluzja: obrazy o większej liczbie detali implikują niższe wymagania odnośnie liczby poziomów szarości.



a b c

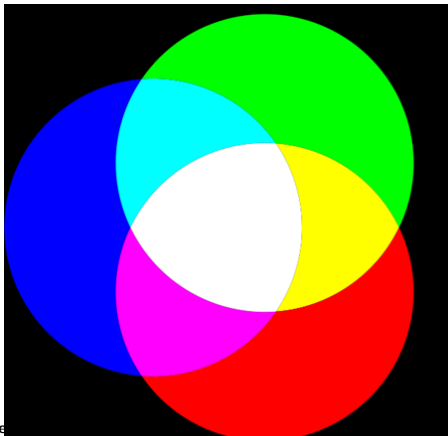
FIGURE 2.22 (a) Image with a low level of detail. (b) Image with a medium level of detail. (c) Image with a relatively large amount of detail. (Image (b) courtesy of the Massachusetts Institute of Technology.)



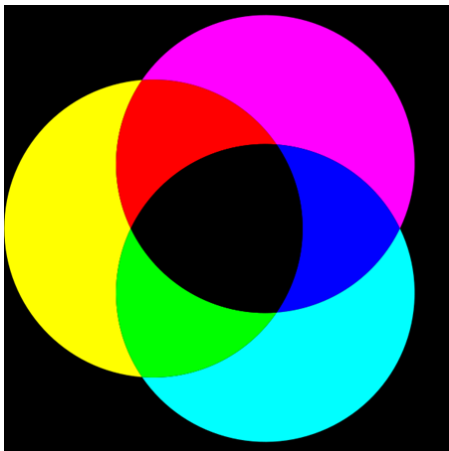
- Co to jest kolor?
- Przestrzenie reprezentacji barw: RGB, HSI, HSV, CMYK,...
- Motywacje dla utrzymywania wielu przestrzeni reprezentacji barw:
 - zgodność z percepcją wzrokową człowieka,
 - zgodność z pewnymi procesami/standardami technologicznymi,
 - zróżnicowanie reprezentacji barw przydatne w przetwarzaniu i wnioskowaniu.

Przestrzeń RGB

Model addytywny. Powszechnie stosowany w urządzeniach projekcyjnych (wrażenie koloru powstaje w efekcie projekcji/emisji światła).



Model subtraktywny. Wykorzystywany głównie w druku (wrażenie koloru powstaje w efekcie odbicia światła).



Konwersja z RGB do CMY

- $C = 1 - R$, $M = 1 - G$, $Y = 1 - B$

Wprowadzenie kanału K (K – key (czern)) redukuje ilość niezbędnego barwnika.

Uwaga: konwersja nieodwracalna (stratna).

if $\min\{C', M', Y'\} = 1$

then

$$t_{CMYK} = \{0, 0, 0, 1\}$$

else

$$K = \min\{C', M', Y'\}$$

$$t_{CMYK} = \left\{ \frac{C' - K}{1 - K}, \frac{M' - K}{1 - K}, \frac{Y' - K}{1 - K}, K \right\}$$

CMYK vs. CMY

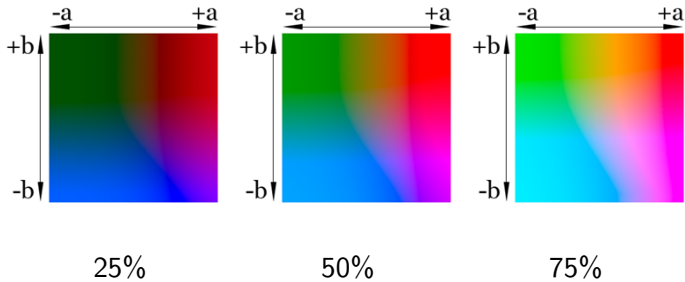


- Przestrzenie RGB i CMY/CMYK nie są przestrzeniami absolutnymi, bo odwołują się do nieustandaryzowanych barw podstawowych – rezultat na monitorze/w wydruku może być różny.
- Inaczej mówiąc: co to jest R? (w sensie fizycznym)

Absolutna przestrzeń reprezentacji barw: L^*a^*b (LAB, CIELAB). Trzy kanały:

- L – jasność (lightness),
- a – pozycja pomiędzy M a G
- b – pozycja między Y a B
- a i b abstrahują od L

Założenie: takie same subiektywne różnice w percepcji barw powinny przekładać się na zbliżone różnice wartości składowych (liniowość percepcyjna, perceptual linearity)



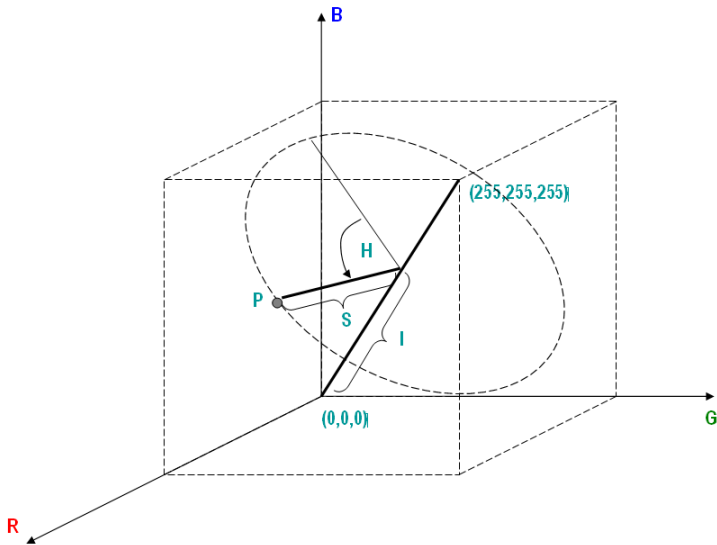
Motywacja: lepsza zgodność z cechami obrazu postrzeganymi przez człowieka (w odróżnieniu od np. RGB)

- *Hue* (barwa)
- *Saturation* (nasycenie, wysycenie)
- *Intensity* [Value] (jasność)

Bardzo popularna w metodach IP i PR, także w grafice komputerowej (tzw. *color pickers*).

Zbliżone przestrzenie: HSV, HSL, HSB

Konwersja RGB $\leftrightarrow$ HSI



Wymagane przejście ze współrzędnych kartezjańskich do biegunowych:

$$\begin{aligned}\alpha &= \frac{1}{2}(2R - G - B) \\ \beta &= \frac{\sqrt{3}}{2}(G - B) \\ H(\beta, \alpha) &= \text{atan2}(\beta, \alpha)\end{aligned}$$

W praktyce używa się uproszczonych wzorów które nie angażują trygonometrii:

$$H' = \begin{cases} \text{undefined}, & \text{if } C = 0 \\ \frac{G-B}{C} \bmod 6, & \text{if } M = R \\ \frac{B-R}{C} + 2, & \text{if } M = G \\ \frac{R-G}{C} + 4, & \text{if } M = B \end{cases}$$

$$H = 60^\circ \times H'$$

gdzie C to *chroma*, tj.

$$C = \max(R, G, B) - \min(R, G, B)$$

Wzory przybliżone generują wartości Hue różniące się od wartości dokładnych o nie więcej niż ok. 1° .



Przekształcenia mogą prowadzić do częściowej utraty informacji z powodu:

- dyskretyzacji,
- projekcyjnego (zamiast bijekcyjnego) mapowania (np. RGB- $\rightarrow$ HSI dla obrazów mało intensywnych, np. słabo wybarwionych preparatów mikroskopowych)

Quizz:

- Czym z teorii informacyjnego punktu widzenia różnią się poszczególne metody reprezentacji barw?

Najczęściej:

- wiersz po wierszu, kolumna po kolumnie

Uwagi techniczne

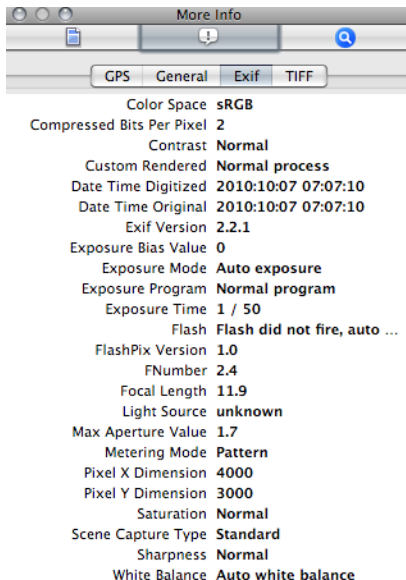
- preferowane przeglądanie zgodne z porządkiem pikseli (ze względu na pamięć podręczną cache),
- w przypadku obrazów wielokanałowych: kanały przeplatane; niekiedy dla przyspieszenia operacji warto odseparować od siebie kanały w pamięci,
- reprezentacja w pamięci często de facto 32 bitowa
- w przypadku nieparzystej szerokości obrazu, wiele bibliotek rezerwuje o jeden bajt na linię obrazu więcej, tak aby każda linia zaczynała się od parzystego adresu.

Inne parametry obrazu opisujące sposób jego pozyskiwania:

- rozdzielczość (DPI),
- długość linii w bajtach (np. BMP)

Standard EXIF (Exchangeable Image File Format) - format przewidujący pewien zestaw metatagów dodawanych do istniejących formatów plików graficznych, np. czas ekspozycji, ustawienia równowagi bieli. Stosowany m.in. w JPEG, TIFF, RIFF, nie używany w PNG, JPEG2000, GIF.

Przykład rekordu EXIF



Formaty plików obrazów rastrowych

Stratne:

- JPEG (Joint Photographic Experts Group) – zasadniczo metoda kompresji a nie format pliku, nazwa formatu to JFIF (JPEG File Interchange Format). Sukcesywne kompresje i dekompresje degradują jakość obrazu.

Bezstratne:

- GIF (Graphics Interchange Format) – 256 kolorów + multiframe, opatentowany (z powodu LZW), obecnie patenty wygasły w większości krajów
- PNG (Portable Network Graphics) – następca GIF, open source
- BMP – Windows bitmap
- RAW – raczej metaformat, nieustandaryzowany

Mieszane:

- TIFF/TIF (Tagged Image File Format) – szeroki standard, może angażować LZW
- JPEG2000 – lepsze właściwości niż JPEG, ale wyższy koszt obliczeniowy kompresji i dekompresji; zastosowaniach profesjonalne (np. niektóre kina)

Inne: PCX, PPM, PGM

Przykład artefaktów wynikających z kompresji obrazu



Przykład artefaktów wynikających z kompresji obrazu



Materiały dodatkowe

Obrazy rastrowe to nie jedyna forma reprezentacji informacji obrazowej. Klasy obrazów (za: Pavlidis):

- Klasa 1: obrazy o pełnej gradacji i kolorowe ("true color"); najbliższe obrazowi rzeczywistemu.
- Klasa 2: obrazy dwupoziomowe lub kilkukolorowe ("czarno-białe"), 1bit/element
- Klasa 3: krzywe ciągłe i linie proste (różnice dx, dy lub kody łańcuchowe)
- Klasa 4: punkty lub wieloboki (powiązanie z grafiką komputerową; modele do PR)

Transformacje pomiędzy klasami obrazów

