Przetwarzanie informacji obrazowej

Zajęcia 1

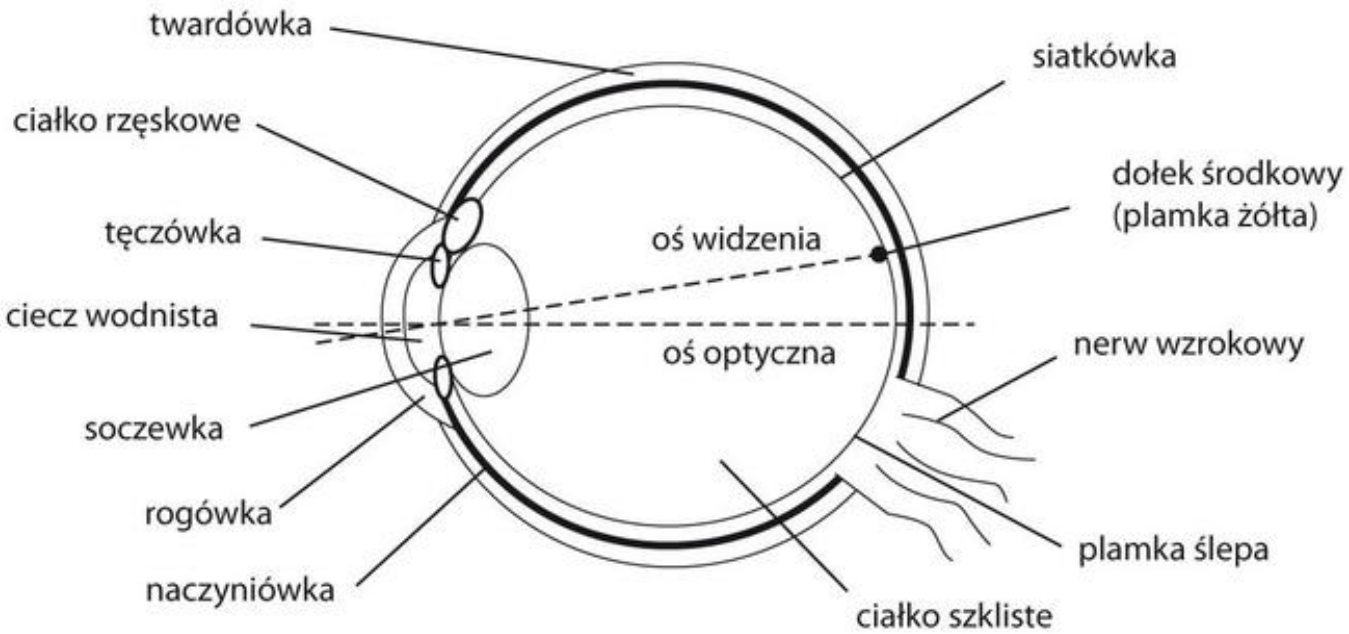
Piotr Syga

04.03.2022

Motywacja

- poprawa jakości zdjęć i filmów
- modyfikacja obrazów, dodawanie informacji pomocniczych
- wykrywanie, rozpoznawanie, klasyfikacja obiektów
- biometria
- ochrona prywatności, ochrona praw autorskich

Budowa oka


twardówka
siatkówka
ciałko rzęskowe
dołek środkowy
(plamka żółta)
tęczówka
oś widzenia
ciecz wodnista
oś optyczna
nerw wzrokowy
soczewka
rogówka
plamka ślepa
naczyniówka
ciałko szkliste

Postrzeganie barw i ruchu

Barwy

Trzy rodzaje czopków odpowiedzialnych za widzenie barw

- 430 ± 10 nm
- 540 ± 5 nm
- 575 ± 15 nm

Widzenie nocne

Pręciki – fotoreceptory o wysokiej czułości na światło (w tym rozproszone), niska ostrość, wrażliwość na ruch (odwrócone sprzężenie zwrotne)

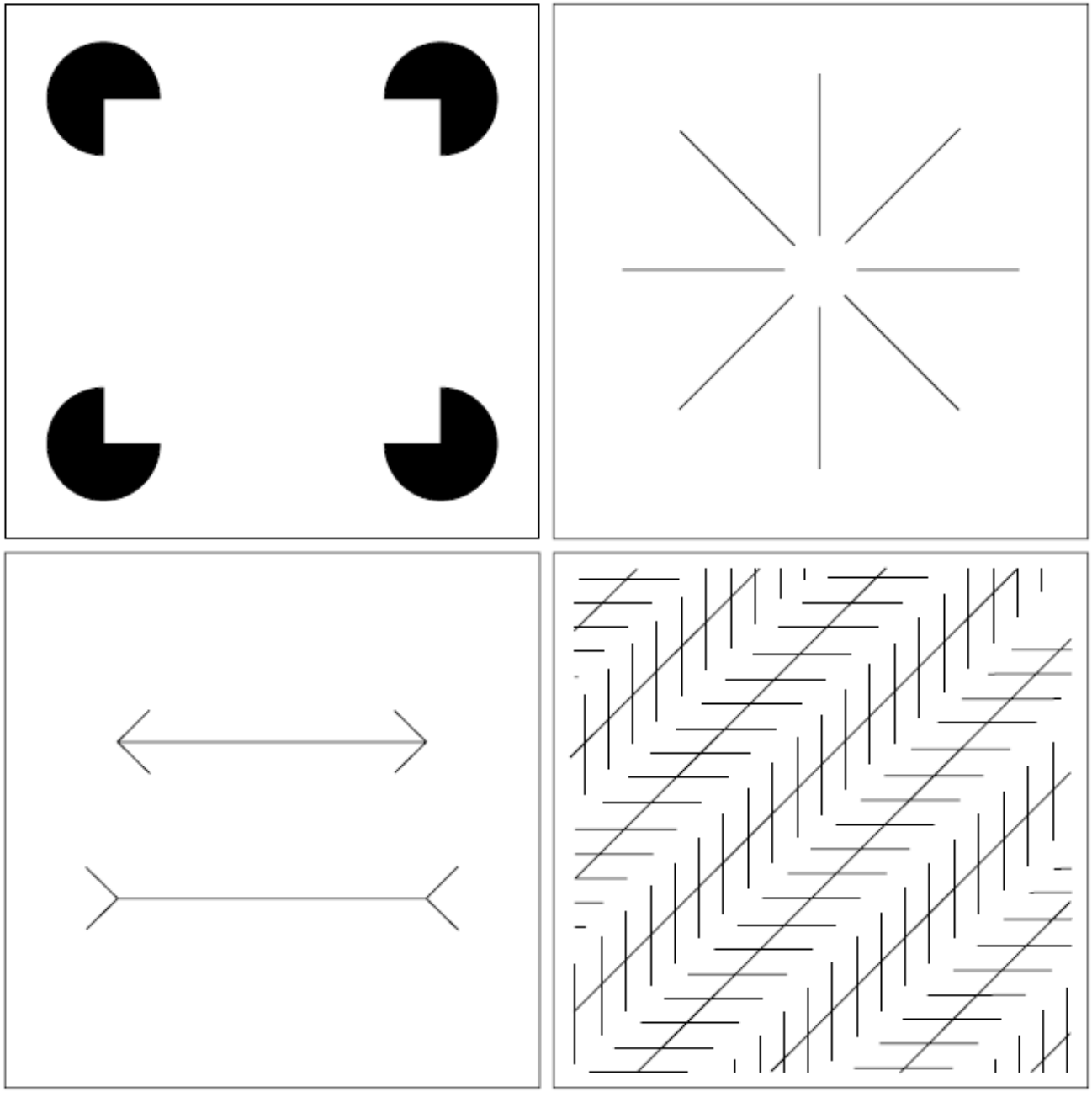
Percepcja

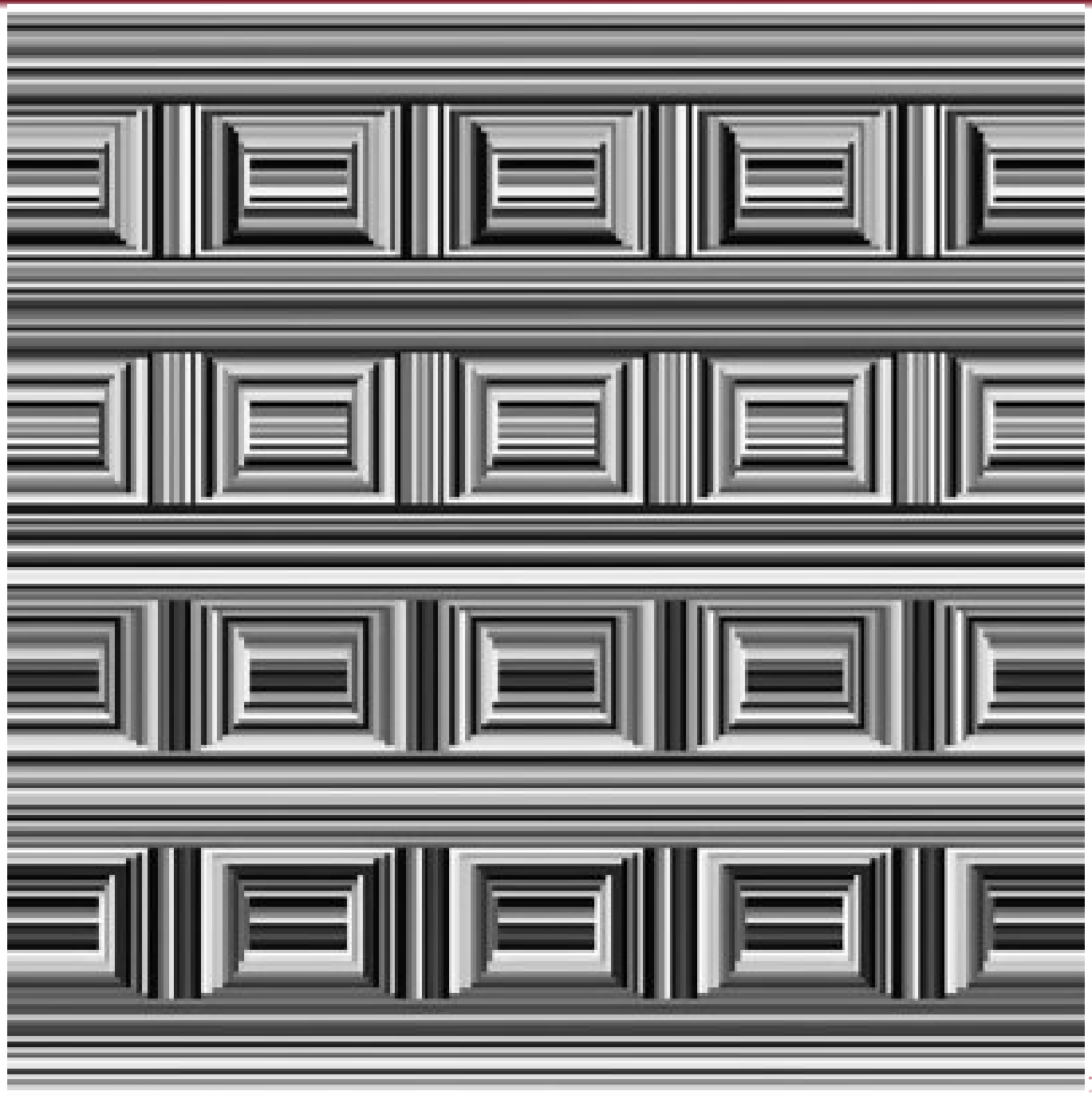


Percepcja

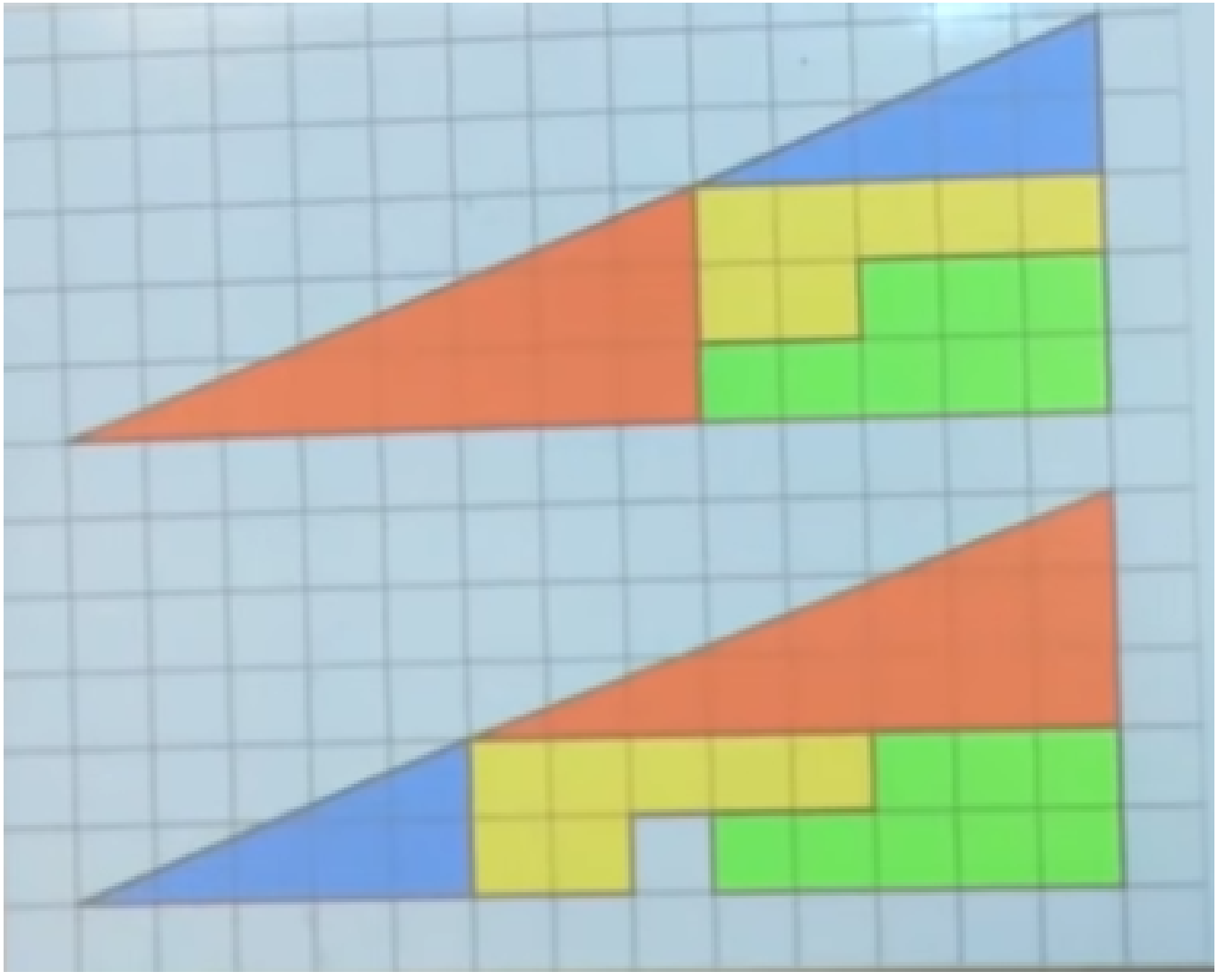


Postrzeganie ograniczone jest nie tylko przez intensywność światła oraz długość fali. Czas ekspozycji, kontrast oraz inne czynniki wpływają na interpretację barw i kształtów.

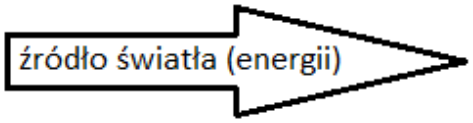


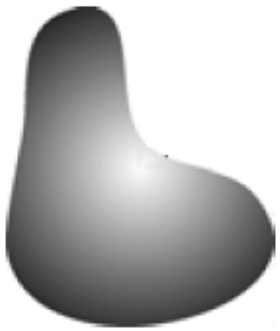




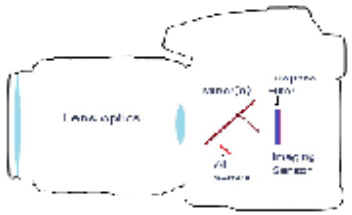


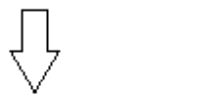
Obrazowanie

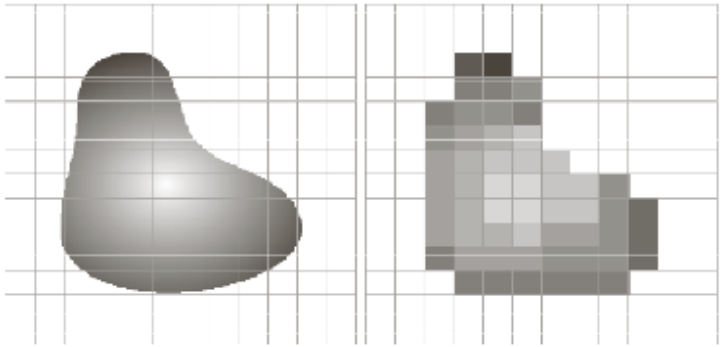




obiekt





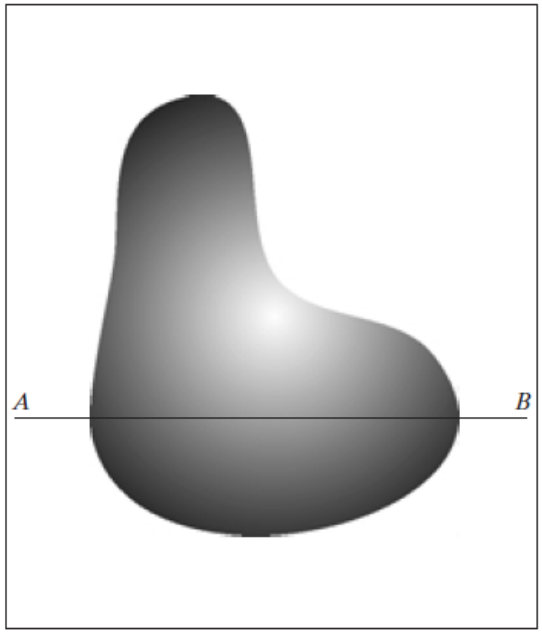


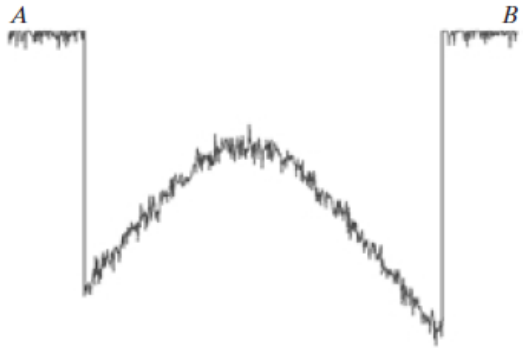
Obraz jako funkcja

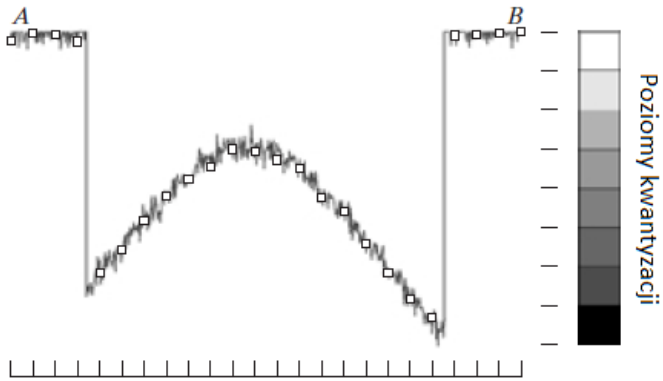
$$f(x, y) = i(x, y)r(x, y) \ ,$$

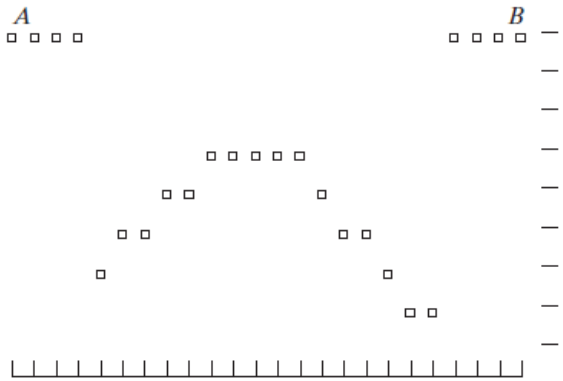
- $0 < i(\cdot, \cdot) < \infty$, jest funkcją określającą „natężenie światła”
- $0 < r(\cdot, \cdot) < 1$, jest funkcją określającą zdolność obiektu do odbijania światła

Próbkowanie, kwantyzacja









Reprezentacja, piksele

Będziemy zakładać, że operujemy na przestrzeni dyskretnej, rozmiaru $N \times M$.

$$f(x, y) = \begin{bmatrix} f(0, 0) & f(0, 1) & \dots & f(0, M-1) \\ f(1, 0) & f(1, 1) & \dots & f(1, M-1) \\ \vdots & \vdots & \ddots & \vdots \\ f(N-1, 0) & f(N-1, 1) & \dots & f(N-1, M-1) \end{bmatrix}$$

Uwaga: Zależnie od użytego oprogramowania, zakres może być przesunięty o 1.
Pojedynczy element macierzy będziemy nazywać pikselem.



$N = 4032,\ M = 2268$



$N = 202,\ M = 113$

Przestrzenie barw

Poza dyskretyzacją przestrzenną, należy poradzić sobie z dyskretyzacją wartości funkcji $f(\cdot, \cdot)$.

- obraz binarny: $f(x, y) \in \{0, 1\}$
- obraz w skali szarości: $f(x, y) \in \{0, \ldots, L-1\}$ $(L = 2^k,\ k \in \mathbb{N}^+)$
- obraz wielokanałowy: $f(x, y) \in \{0, \ldots, L-1\}^c$ $(c \in \mathbb{N}^+)$

Uwaga:
Zamiast $f(x, y) = \langle i_0, i_1, \ldots, i_{c-1} \rangle$ będziemy przyjmować $f(x, y) = \langle f_0(x, y), \ldots, f_{c-1}(x, y) \rangle$, $f_j(x, y) = i_j$.



$k = 8,\ c = 3$



$k = 8,\ c = 1$



$k = 6,\ c = 1$



$k = 3,\ c = 1$

RGB

Historia

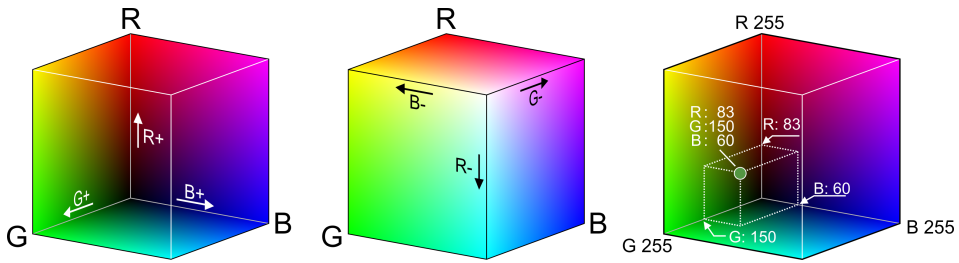
W związku z trzema rodzajami czopków u naczelnych, już w 19. wieku powstała teoria trichromatyczna (Young–Helmholtz), opierając się na niej powstał i rozwijał się addytywny system reprezentowania barw za pomocą składowej czerwonej, niebieskiej i zielonej (koła kolorystyczna dla CRT). Początkowo 'ilość koloru' w danym kanale była określana procentowo.

Warianty

Standardem jest identyfikacja koloru za pomocą 8-bitów na kanał. Można spotkać się z wariantami zawężonymi (4b/kanał, 4kanały) oraz, częściej, rozszerzonymi – 12, 16, 24 i 32 bitowymi kanałami, dla dużych gamutów UHD.

Uwaga: W `OpenCV` można się spotkać z porządkiem BGR.

RGB

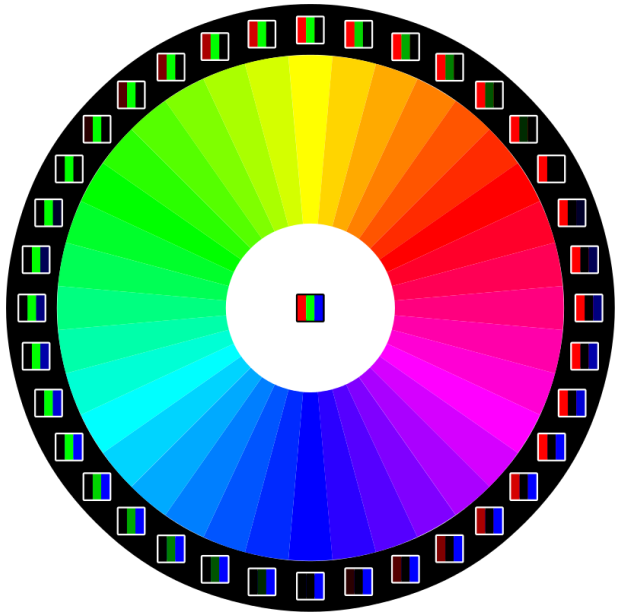


Warianty

Standardem jest identyfikacja koloru za pomocą 8-bitów na kanał. Można spotkać się z wariantami zawężonymi (4b/kanał, 4kanały) oraz, częściej, rozszerzonymi – 12, 16, 24 i 32 bitowymi kanałami, dla dużych gamutów UHD.

Uwaga: W `OpenCV` można się spotkać z porządkiem BGR.

RGB

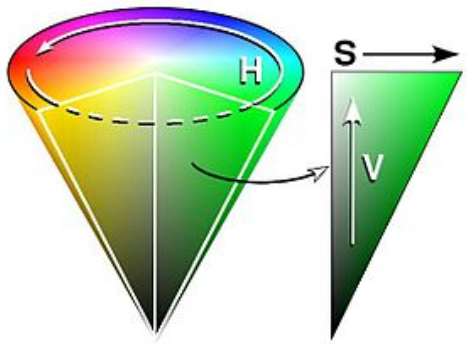


Warianty

Standardem jest identyfikacja koloru za pomocą 8-bitów na kanał. Można spotkać się z wariantami zawężonymi (4b/kanał, 4kanały) oraz, częściej, rozszerzonymi – 12, 16, 24 i 32 bitowymi kanałami, dla dużych gamutów UHD.

Uwaga: W `OpenCV` można się spotkać z porządkiem BGR.

HSV



Model opierający się na trzech kanałach: odcieniu (*hue*), nasyceniu (*saturation*) oraz 'ilości' zmieszanych barw (*value*).

Odcień – odnosi się do długości fali, określa stopień podobieństwa do barw (czerwona, niebieska, zielona, żółta, cyjan i magenta), wyrażony jest kątowo

Nasycenie – określa intensywność koloru, od jego braku, do maksymalnej intensywności, wyrażony jest wartością z przedziału $[0, 1]$

Wartość – określa 'ilość' zmieszanych barw lub też cienie, wyrażony jest wartością z przedziału $[0, 1]$

Konwersje

RGB – HSV

Dla $R, G, B \in [0, 255]$:

$\hat{R} = \frac{R}{255}, \hat{G} = \frac{G}{255}, \hat{B} = \frac{B}{255}$

$M = \max\left\{\hat{R}, \hat{G}, \hat{B}\right\}, m = \min\left\{\hat{R}, \hat{G}, \hat{B}\right\}$

$$V = M$$

$$S = \begin{cases} 0, & M = 0 \\ \frac{M-m}{M} & \text{w p.p} \end{cases}$$

$$H = \begin{cases} 0, & M = m \\ 60\left(\frac{\hat{G}-\hat{B}}{M-m}\right), & M = \hat{R} \\ 60\left(2 + \frac{\hat{B}-\hat{R}}{M-m}\right), & M = \hat{G} \\ 60\left(4 + \frac{\hat{R}-\hat{G}}{M-m}\right), & M = \hat{B} \end{cases}$$

Konwersje

RGB – HSV

Dla $H \in [0, 360]$, $S \in [0, 1]$, $V \in [0, 1]$:

$\hat{H} = \frac{H}{60}$, $X = V \cdot S \cdot \left(1 - \left|(\hat{H} \mod 2) - 1\right|\right)$

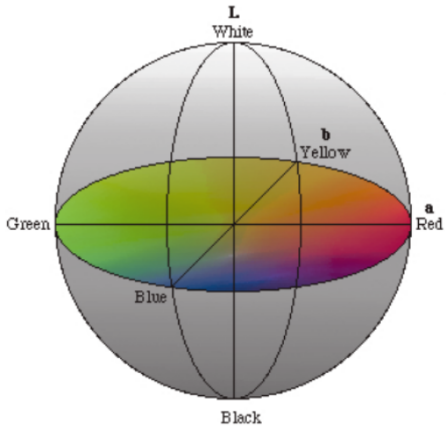
$$\hat{R} = \begin{cases} 0, \text{ H – niezdefiniowane} \vee 2 \leq \hat{H} \leq 4 \\ VS, \ 0 \leq \hat{H} \leq 1 \vee 5 < \hat{H} \leq 6 \\ X, \ 1 < \hat{H} \leq 2 \vee 4 < \hat{H} \leq 5 \end{cases}$$

$$\hat{G} = \begin{cases} 0, \text{ H – niezdefiniowane} \vee 4 \leq \hat{H} \leq 6 \\ VS, \ 1 < \hat{H} \leq 3 \\ X, \ 0 \leq \hat{H} \leq 1 \vee 3 < \hat{H} \leq 4 \end{cases}$$

$$\hat{R} = \begin{cases} 0, \text{ H – niezdefiniowane} \vee 0 \leq \hat{H} \leq 2 \\ VS, \ 3 < \hat{H} \leq 5 \\ X, \ 2 < \hat{H} \leq 3 \vee 5 < \hat{H} \leq 6 \end{cases}$$

$$(R, G, B) = (\hat{R} + V(1 - S), \hat{G} + V(1 - S), \hat{B} + V(1 - S))$$

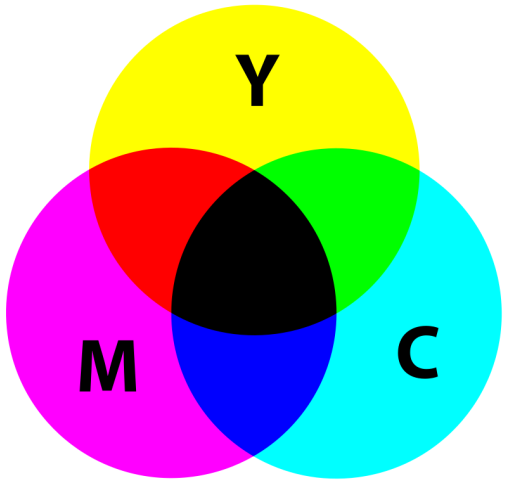
CIELAB



Przestrzeń barw określona za pomocą trzech wartości:

- L^* – jasność barwy wyrażona w skali od 0 (czerń) do 100 (biel)
- a^* – odległość od barwy zielonej (-2^{K-1}) do czerwonej $(2^{K-1}-1)$
- b^* – odległość od barwy niebieskiej (-2^{K-1}) do żółtej $(2^{K-1}-1)$

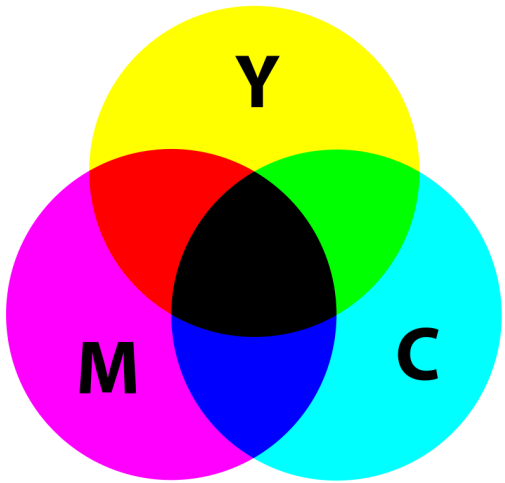
CMYK



Model oparty na odejmowaniu kolorów, który pierwotnie był używany do warstwowego nakładania farby drukarskiej. Model opiera się na 4 kanałach – cyjan (C), magenta (M), żółty (Y) oraz czarny (K). Dla urządzeń drukujących każdy kanał ma (próbkowaną zgodnie z rozdzielczością urządzenia) wartość z przedziału [0, 1], standardem cyfrowym jest używanie 8 bitów na kanał.

Uwaga: Dla różnych urządzeń wyświetlających/drukujących, ale również w części bibliotek, skale C, M, Y mogą się różnić.

CMYK



Przykład najprostszej konwersji (dla RGB znormalizowanego dla $[0, 1]$):

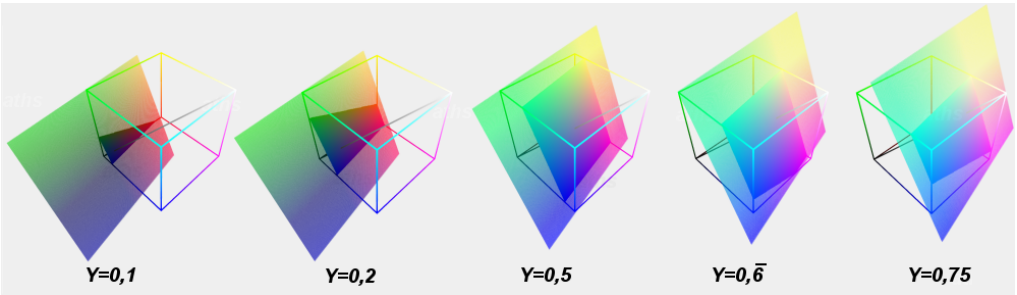
$$K = 1 - \max\{R, G, B\}$$

$$C = \frac{1 - (R + K)}{1 - K}$$

$$M = \frac{1 - (G + K)}{1 - K}$$

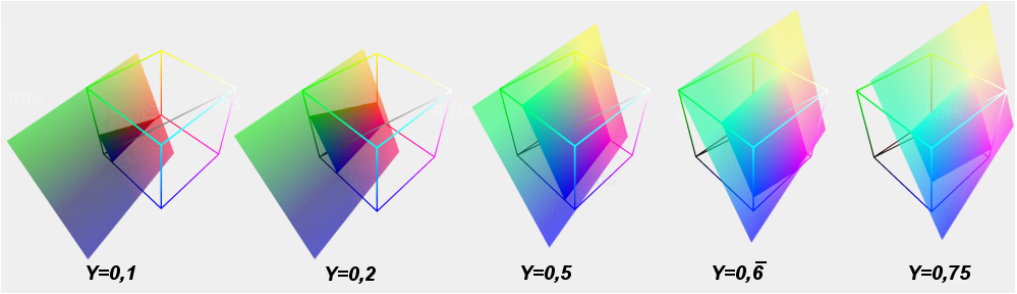
$$Y = \frac{1 - (B + K)}{1 - K}$$

YCbCr



Model opiera się na 3 kanałach: luminancji (Y) oraz dwóch kanałów chrominancji (Cb, Cr). Stosowany jest przy przetwarzaniu obrazów i filmów w postaci cyfrowej, gdyż pozwala dostosować się do możliwości urządzenia odbierającego, co pozwala na efektywną kompresję. Zależnie od domyślnego urządzenia wyświetlającego dobierane są różne stałe kolorystyczne (dla czerwieni, zieleni i niebieskiego) oraz zauważalna jest różnica w kwantyzacji.

YCbCr



Przykład konwersji (dla RGB znormalizowanego dla $[0, 255]$), przed zaokrąglaniem:

$$Y = 16 + \frac{65.738}{256}R + \frac{129.057}{256}G + \frac{25.064}{256}B$$
$$Cb = 128 - \frac{37.945}{256}R - \frac{74.494}{256}G + \frac{112.439}{256}B$$
$$Cr = 128 + \frac{112.439}{256}R - \frac{94.154}{256}G - \frac{18.285}{256}B$$

Konwersja YCbCr do RGB:

$$R = Y + 1.402525\,Cr$$
$$G = Y - 0.34373\,Cb - 0.714401\,Cr$$
$$B = Y + 1.769905\,Cb + 0.000013\,Cr$$

Obrazy w skali szarości

- W wielu zastosowaniach interesować nas będzie jedynie intensywność (poziom naświetlenia) obrazu.
- Ograniczamy się do jednego kanału.
- Można zauważyć wyraźną różnicę pomiędzy analizą pojedynczo każdego z kanałów RGB lub HSV, a obrazem w skali szarości.
- Przy konwersji uśredniającej wartości kanałów powstają obrazy o dużym (niezamierzonym) kontraście oraz niskiej intensywności.
- Konwersja RGB do pojedynczego kanału intensywności:

$$I = 0.299R + 0.587G + 0.114B$$

Binaryzacja

Dla części z algorytmów, nawet informacja o poziomach szarości jest nadmiarowa (operacje morfologiczne, wyodrębnianie fragmentu obrazu, ...). W takich przypadkach obraz w skali szarości konwertowany jest na macierz binarną.

- thresholding – ustawienie wartości funkcji zgodnie z zadanym progiem T:

$$I_B(x, y) = \begin{cases} 0, & I(x, y) < T \\ 1, & \text{w p.p} \end{cases}$$

- metoda Otsu – pozwala dynamicznie dobrać próg poprzez rozważenie dwóch problemów: minimalizacji wariancji wewnątrz klasy oraz maksymalizacji odseparowania obu klas.

Piksele sąsiadujące, sąsiedztwo

- Dla piksela p o współrzędnych (x, y) określamy 4-sąsiedztwo, jako zbiór:
  $N_4(p) = \{(x-1, y), (x+1, y), (x, y-1), (x, y+1)\}$
- Dla piksela p o współrzędnych (x, y) określamy sąsiedztwo przekątniowe, jako zbiór:
  $N_D(p) =$
  $\{(x-1, y-1), (x-1, y+1), (x+1, y-1), (x+1, y+1)\}$
- Dla piksela p o współrzędnych (x, y) określamy 8-sąsiedztwo, jako zbiór:
  $N_8(p) = N_4(p) \cup N_D(p)$

Co jeśli, któryś z pikseli leży poza obrazem?

Piksele sąsiadujące, sąsiedztwo

- Dla piksela p o współrzędnych (x, y) określamy 4-sąsiedztwo, jako zbiór:
  $N_4(p) = \{(x-1, y), (x+1, y), (x, y-1), (x, y+1)\}$
- Dla piksela p o współrzędnych (x, y) określamy sąsiedztwo przekątniowe, jako zbiór:
  $N_D(p) =$
  $\{(x-1, y-1), (x-1, y+1), (x+1, y-1), (x+1, y+1)\}$
- Dla piksela p o współrzędnych (x, y) określamy 8-sąsiedztwo, jako zbiór:
  $N_8(p) = N_4(p) \cup N_D(p)$

Co jeśli, któryś z pikseli leży poza obrazem?

- pominięcie
- padding

Garść definicji

Przyległość

Mówimy, że dwa piksele p_0 i p_1 są x-przyległe, jeśli $p_1 \in N_x(p_0)$.

Ścieżka

Ścieżką (w danym sąsiedztwie) z piksela p_0 do piksela p_n nazywamy taki ciąg pikseli $p_0, p_1, \ldots, p_n$, że każda para pikseli p_i, p_{i+1} jest przyległa.

Połączenie

Niech V będzie zbiorem wartości. Mówimy, że dwa piksele p_0, p_n o intensywnościach z V są połączone, jeśli istnieje ścieżka z p_0 do p_n i wszystkie piksele na ścieżce mają intensywność ze zbioru V.

Regiony i kontury

Komponenty spójne

Dla dowolnego piksela p zbiór pikseli połączonych z p ($C(p)$) nazywamy komponentą spójną. Jeśli dla ustalonego podzbioru S, $C(p) = S$, to S nazywamy zbiorem połączonym lub regionem.

Kontur

Niech S będzie regionem, podzbiór $B(S) = \{p \in S : \exists_{q \in N_8(p)} q \notin S\}$ nazywamy konturem. W przypadku niepełnego sąsiedztwa, piksele skrajne S należą do konturu.

Splot macierzy

Splot

Niech A i B będą macierzami. Wtedy splot (konwolucję) macierzy określamy wzorem

$$C(j,\, k) = A * B = \sum_{p} \sum_{q} A(p,\, q) B(j - p, k - q)$$