

Wybrane narzędzia i metody przetwarzania i rozpoznawania obrazów

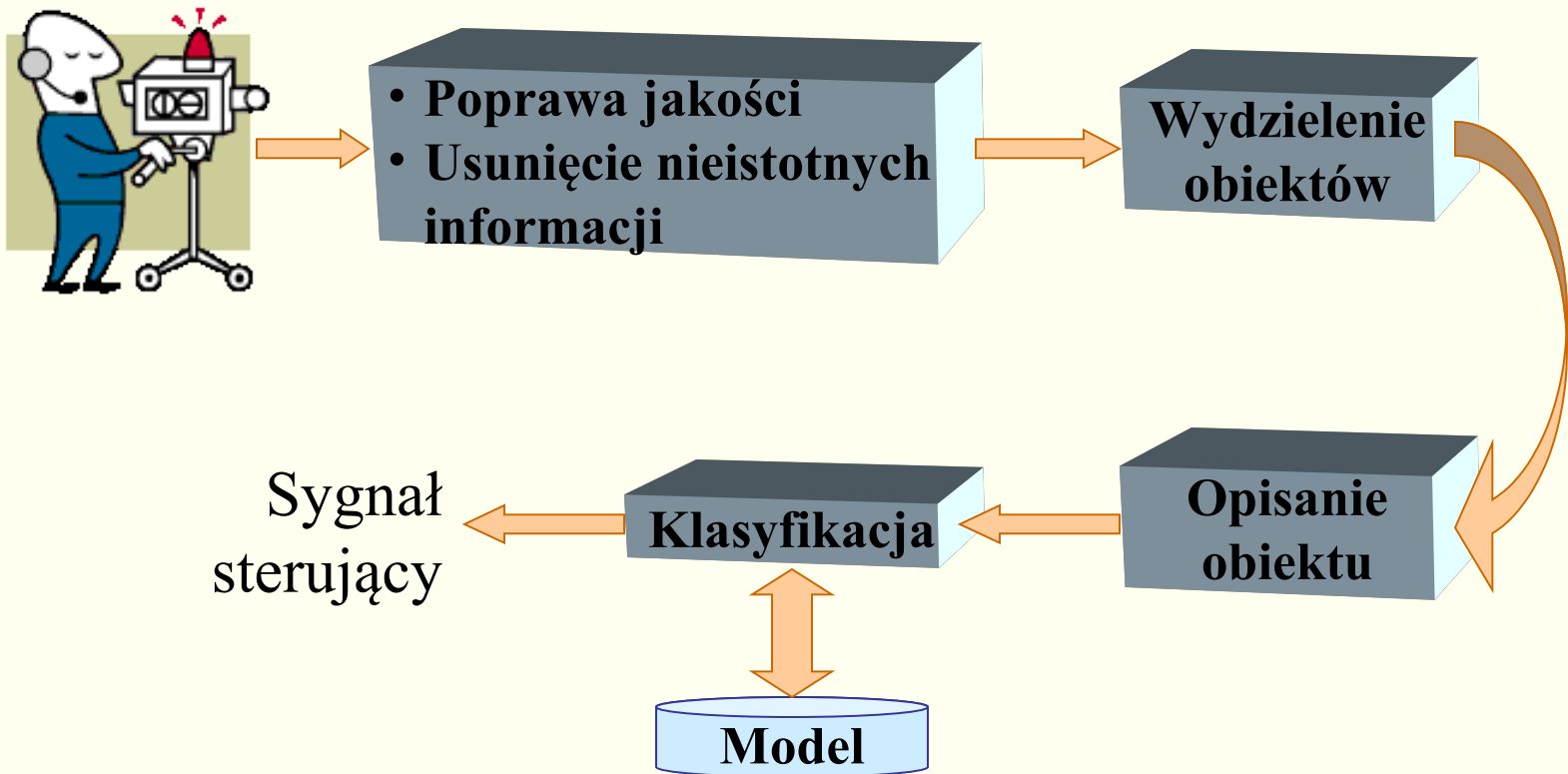
Joanna Marnik

<https://jmarnik.v.prz.edu.pl>

Plan wykładu

1. Budowa systemu wizyjnego
2. Poprawa jakości obrazu
3. Metody filtracji zakłóceń
4. Metody pozwalające na usunięcie nieistotnych informacji
5. Wydzielanie obiektów
6. Klasyfikacja
7. Przykładowe systemy wizyjne i sposób ich rozwiązania
 - i. Rozpoznawanie języka migowego
 - ii. Rozpoznawanie akcji człowieka w oparciu o obrazy historii ruchu

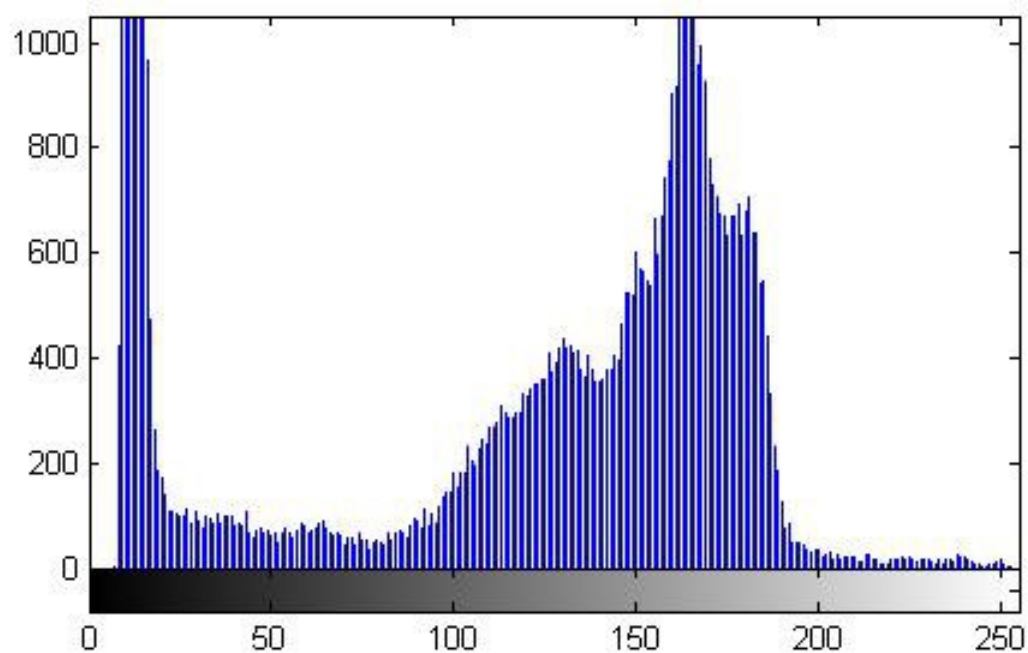
Schemat systemu wizyjnego



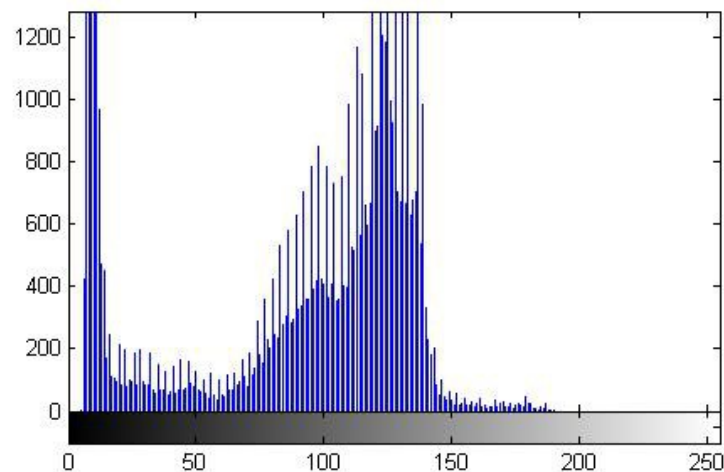
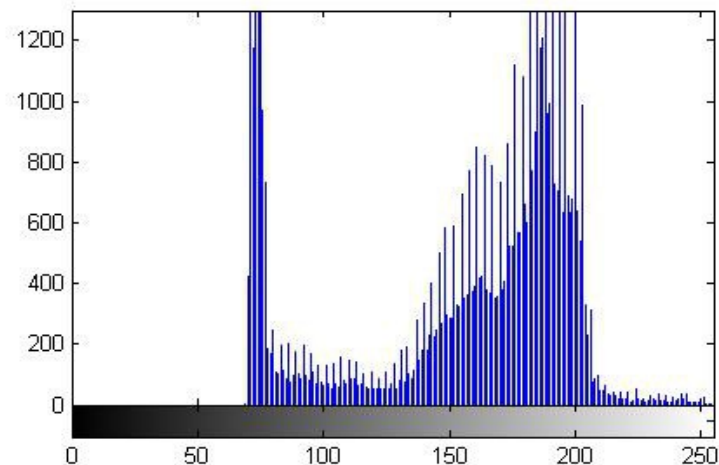
Poprawa jakości obrazu

1. Ocena jakości obrazu w oparciu o histogram
2. Poprawa kontrastu
3. Filtracja zakłóceń

Obraz i jego histogram

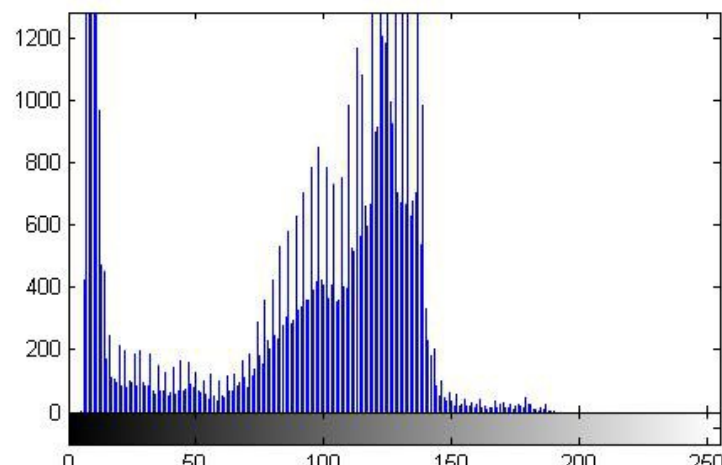


Ocena jakości obrazu na podstawie histogramu

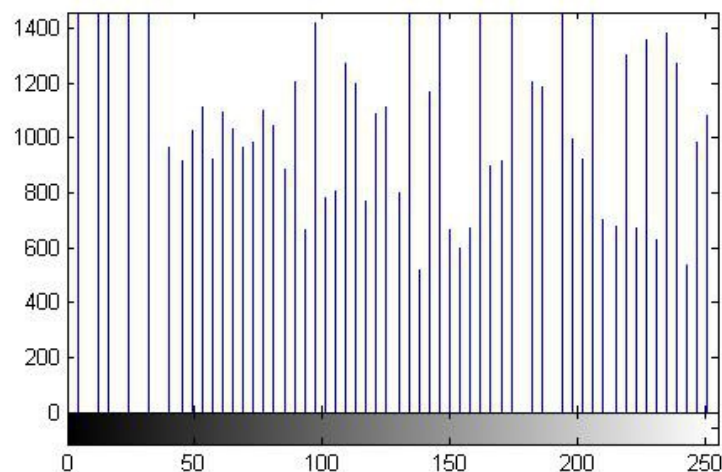


Poprawa kontrastu poprzez wyrównywanie histogramu

przed



po



Metody filtracji zakłóceń

■ Metody przestrzenne

- ◆ Filtracja uśredniająca
- ◆ Filtracja medianowa

■ Metody częstotliwościowe

- ◆ Filtry dolnoprzepustowe
- ◆ Filtry pasmowe

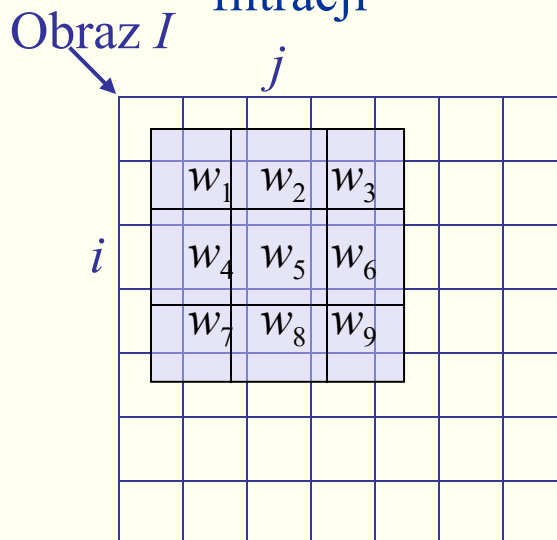
■ Filtry morfologiczne

- ◆ Filtry OC i CO

Filtracja przestrzenna uśredniająca

■ Filtracja uśredniająca

- ◆ Polega na wyznaczeniu nowej wartości dla każdego piksela jako średniej ważonej z wartości znajdujących się w obszarze maski filtracji



$$J(i,j) = [w_1 \cdot I(i-1,j-1) + w_2 \cdot I(i-1,j) + w_3 \cdot I(i-1,j+1) \\ + w_4 \cdot I(i,j-1) + w_5 \cdot I(i,j) + w_6 \cdot I(i,j+1) \\ + w_7 \cdot I(i+1,j-1) + w_8 \cdot I(i+1,j) + w_9 \cdot I(i+1,j+1)] / \\ (w_1 + w_2 + w_3 + w_4 + w_5 + w_6 + w_7 + w_8 + w_9)$$

- ◆ Ze względu na uśrednianie obraz wynikowy jest rozmyty, przy czym większe maski powodują większe rozmycie

Filtracja uśredniająca - przykłady

Przykładowe maski

$$h = \frac{1}{9} \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix}$$
$$h = \frac{1}{10} \begin{bmatrix} 1 & 1 & 1 \\ 1 & 2 & 1 \\ 1 & 1 & 1 \end{bmatrix}$$
$$h = \frac{1}{16} \begin{bmatrix} 1 & 2 & 1 \\ 2 & 4 & 2 \\ 1 & 2 & 1 \end{bmatrix}$$

obraz oryginalny (320×200)



obraz zakłócony szumem Gaussa



obraz po filtracji z maską 3×3

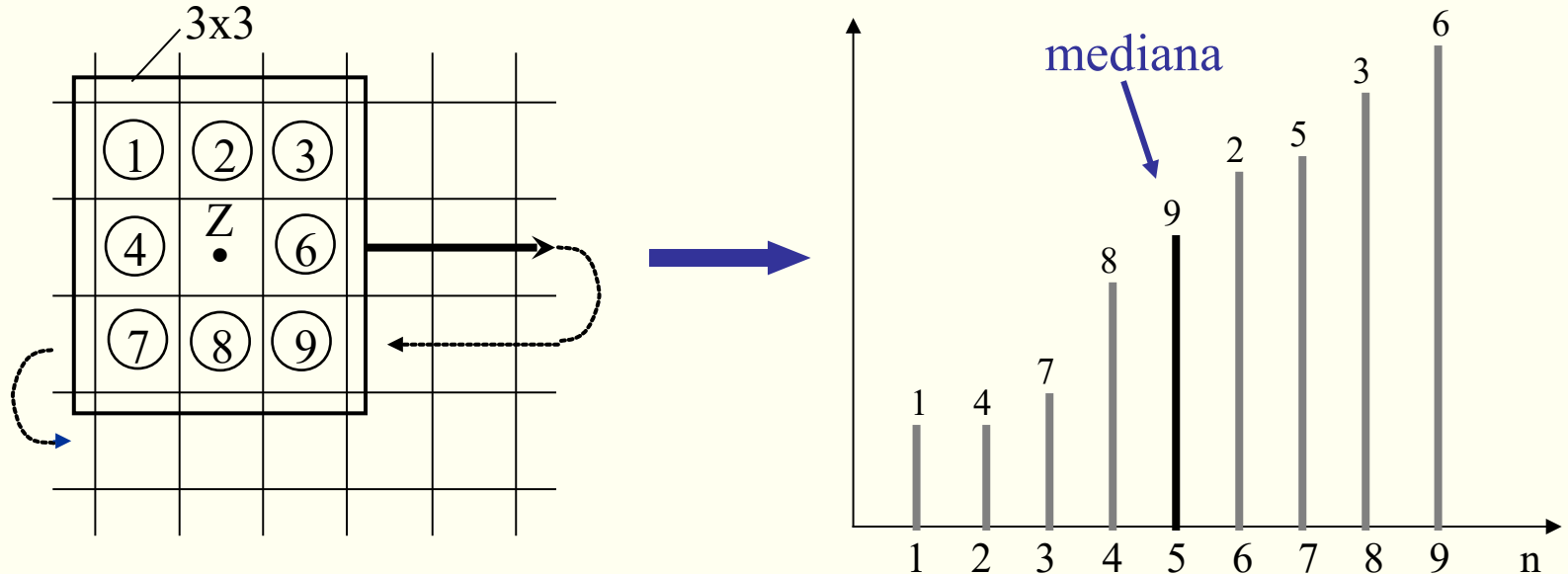


obraz po filtracji z maską 7×7



Filtracja medianowa

- Wartość wynikowa przetwarzanego piksela ustawiana jest na wartość równą medianie wartości obrazu „przykrytych” maską
- Mediana jest to wartość środkowa w uporządkowanym ciągu wartości



Filtracja medianowa - przykłady

obraz oryginalny



obraz zakłócony szumem Gaussa



filtracja medianowa z maską 3×3

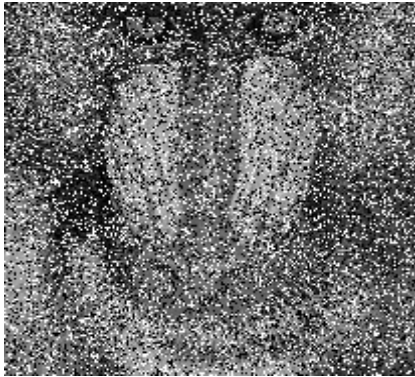


filtracja medianowa z maską 7×7



Filtracja zakłócenia *sól i pieprz*

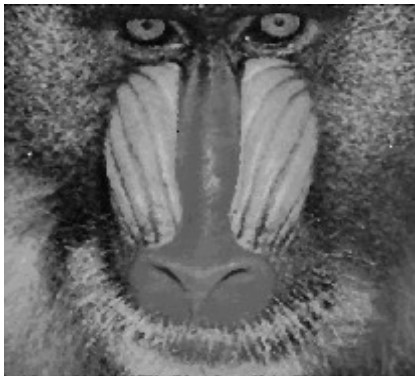
$d=0.3$



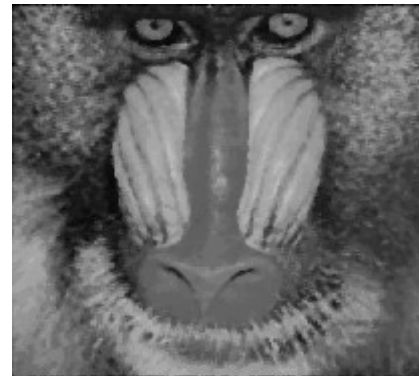
3×3



5×5

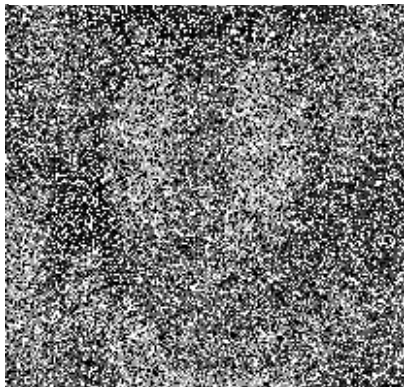


7×7

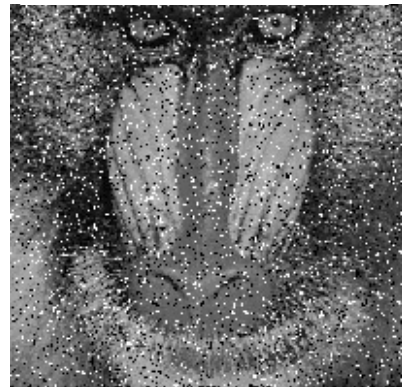


Filtracja zakłócenia *sól i pieprz*

$d=0.5$



3×3



5×5

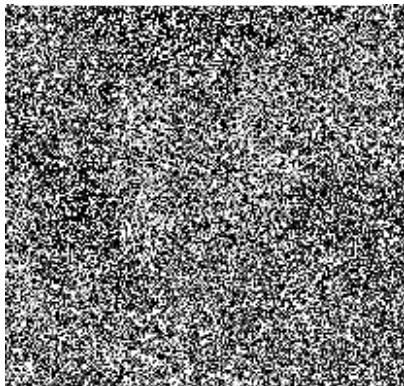


7×7

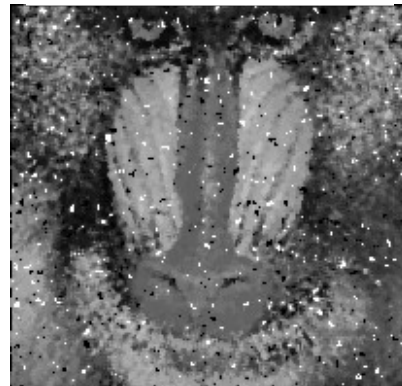


Filtracja zakłócenia *sól i pieprz*

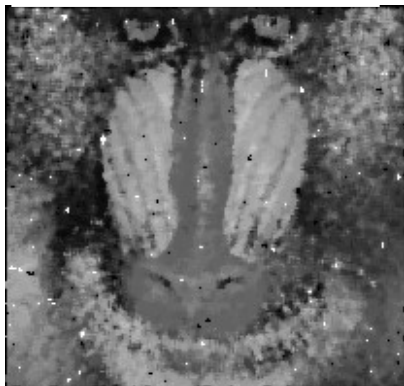
$d=0.7$



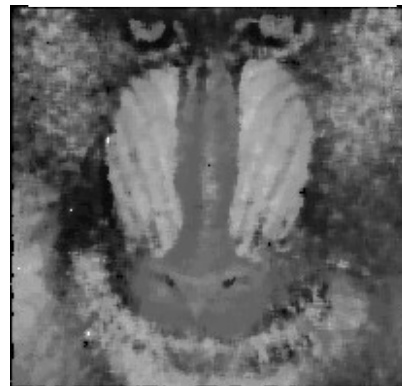
7×7



9×9



11×11

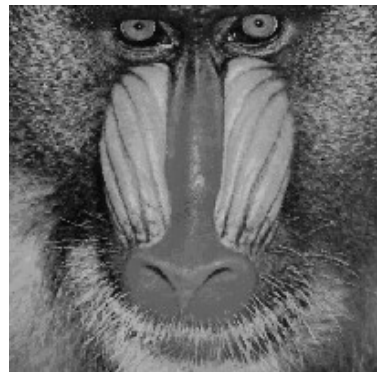


Porównanie metod filtracji przestrzennej

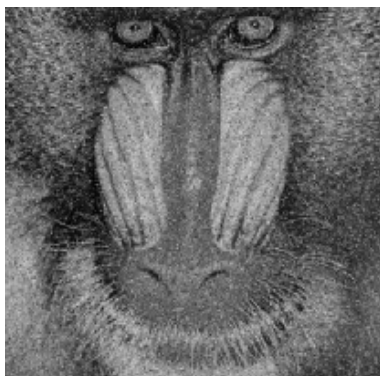
Gęstość zakłócenia $d=0.1$



filtracja medianowa 3×3



filtracja uśredniająca 3×3



filtracja uśredniająca 5×5



Filtracja częstotliwościowa

1. Przekształcić obraz f do postaci widma F z użyciem 2-wymiarowej transformacji Fouriera

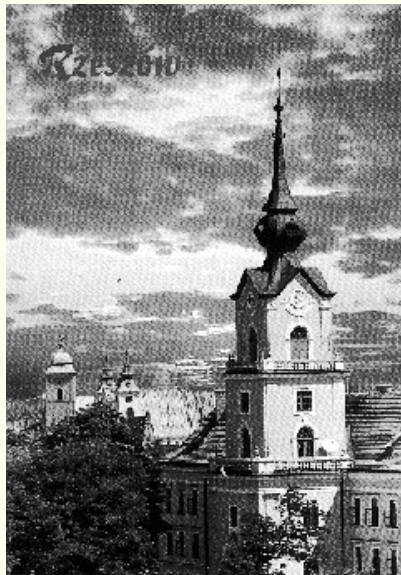
$$F(u, v) = \frac{1}{NM} \sum_{x=0}^{N-1} \sum_{y=0}^{M-1} f(x, y) e^{-2\pi i (xu / N + yv / M)}$$

1. Usunąć z widma te częstotliwości, które zawierają zakłócenia
2. Zmodyfikowane widmo przekształcić do postaci przestrzennej z użyciem odwrotnej transformacji Fouriera

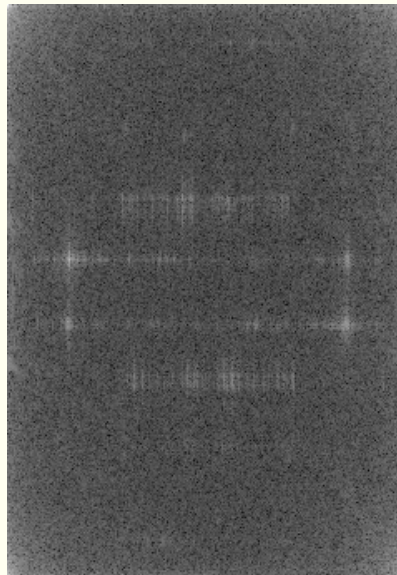
$$f(x, y) = \sum_{u=0}^{N-1} \sum_{v=0}^{M-1} F(u, v) e^{2\pi i (xu / N + yv / M)}$$

Filtracja częstotliwościowa - uwagi

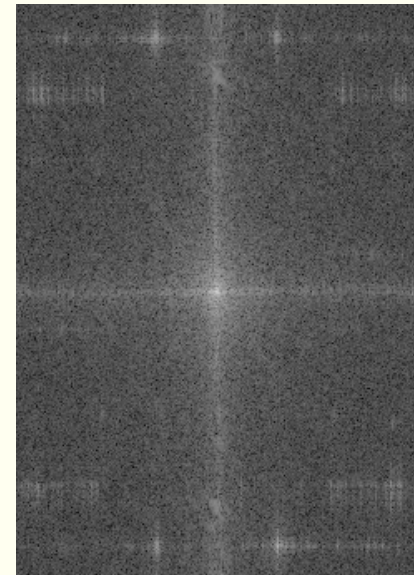
- Otrzymane po transformacji widmo F wygodnie jest przedstawić centryczne, mnożąc przed transformacją $f(x,y)$ przez $(-1)^{x+y}$, co powoduje przesunięcie spektrum.
- Dla lepszej wizualizacji przyjmuje się $\log(1+|F(u, v)|)$



obraz oryginalny



widmo

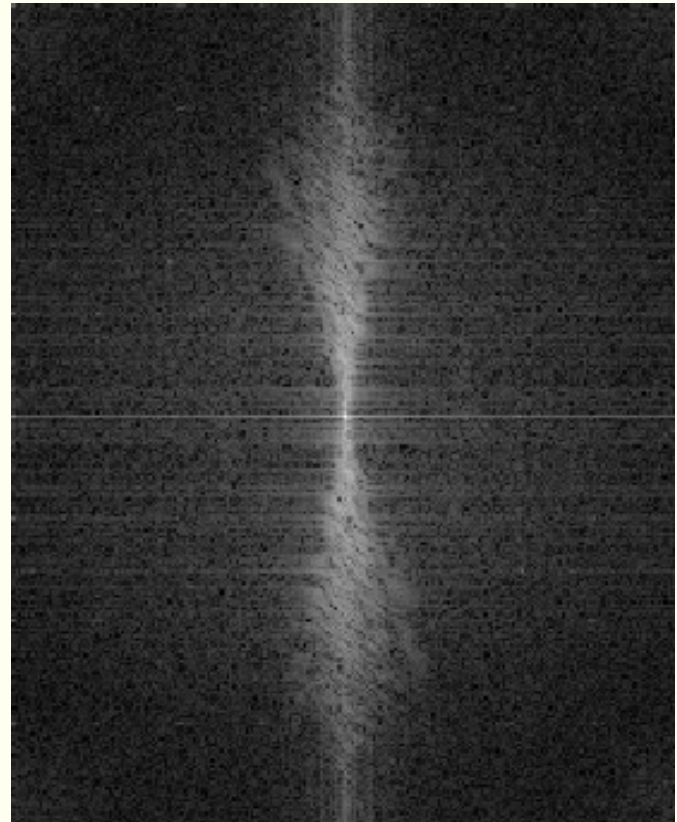


widmo przedstawione
centrycznie

Obraz i jego widmo

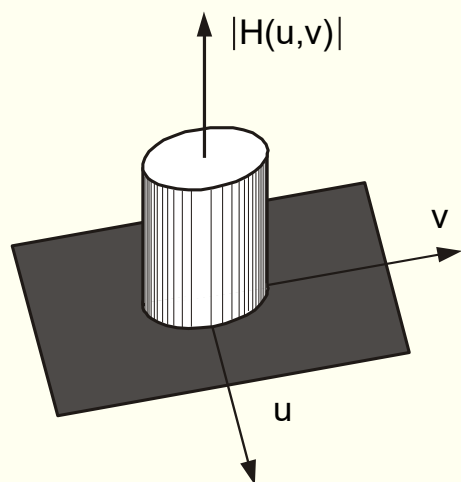


obraz



widmo

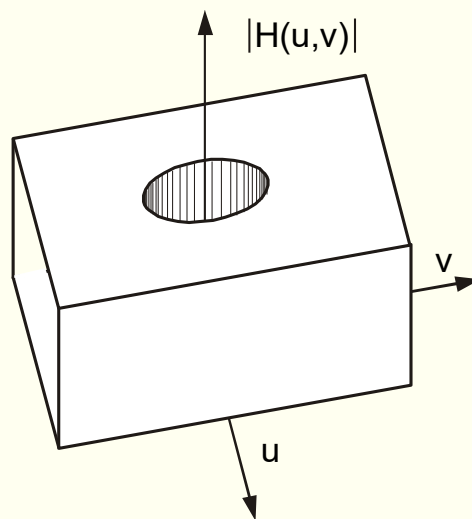
Filtry częstotliwościowe



filtr

dolnoprzepustowy

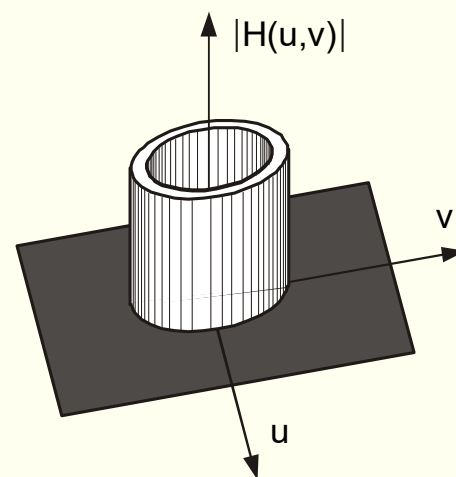
$$H(u, v) = \begin{cases} 1 & \text{dla } \omega \leq \omega_0 \\ 0 & \text{dla } \omega > \omega_0 \end{cases}$$



filtr

górnoprzepustowy

$$H(u, v) = \begin{cases} 0 & \text{dla } \omega < \omega_0 \\ 1 & \text{dla } \omega \geq \omega_0 \end{cases}$$

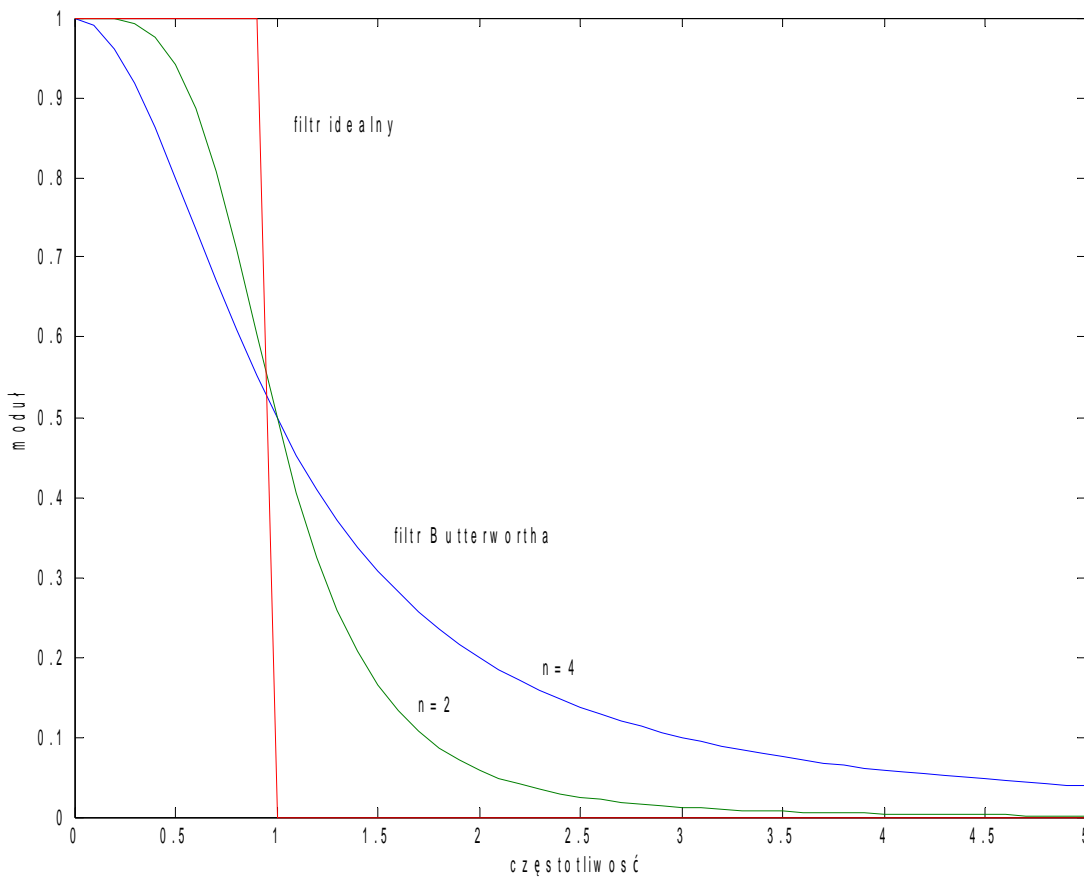


filtr pasmowy

$$H(u, v) = \begin{cases} 1 & \text{dla } \omega_1 \leq \omega \leq \omega_2 \\ 0 & \text{dla } \omega < \omega_1, \omega > \omega_2 \end{cases}$$

Przy czym $\omega^2 = u^2 + v^2$ to częstotliwość odcięcia filtru

Dolnoprzepustowy filtr Butterwortha



$$H(u, v) = \frac{1}{1 + \left(\frac{\omega}{\omega_0} \right)^{2n}},$$

$n = 1, 2, \dots$ - rząd filtru

Filtracja dolnoprzepustowa – filtr idealny

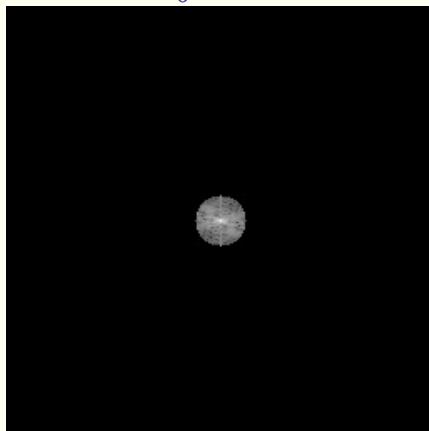
obraz

wejściowy



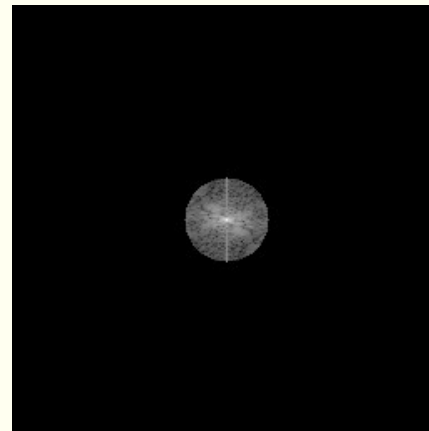
$$|G(u, v)H(u, v)|$$

$$\omega_0 = 30$$



$$|G(u, v)H(u, v)|$$

$$\omega_0 = 50$$



obrazy wynikowe

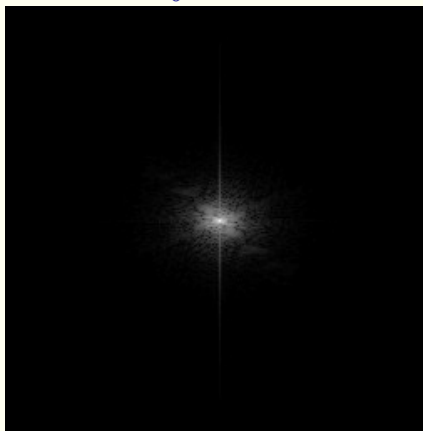


Filtracja dolnoprzepustowa – filtr Butterwortha

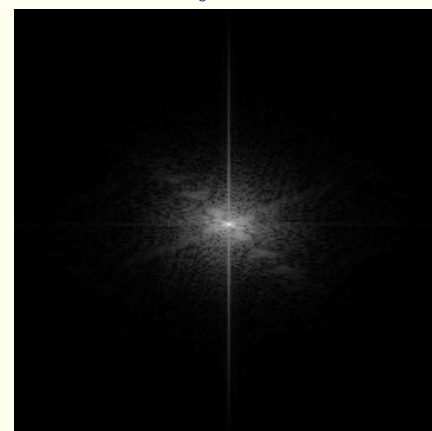
obraz wejściowy



$$|G(u, v)H(u, v)|$$
$$\omega_0 = 10$$



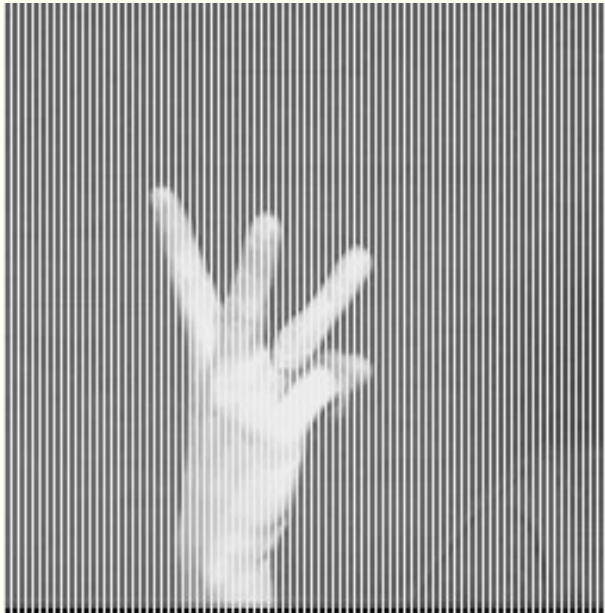
$$|G(u, v)H(u, v)|$$
$$\omega_0 = 30$$



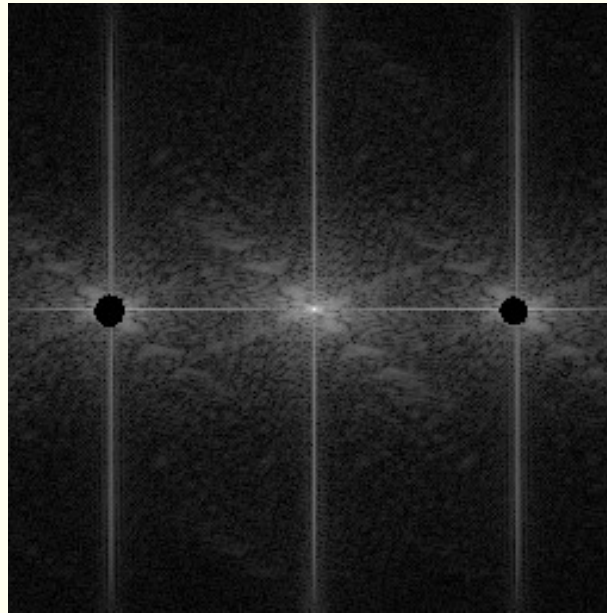
obrazy wynikowe



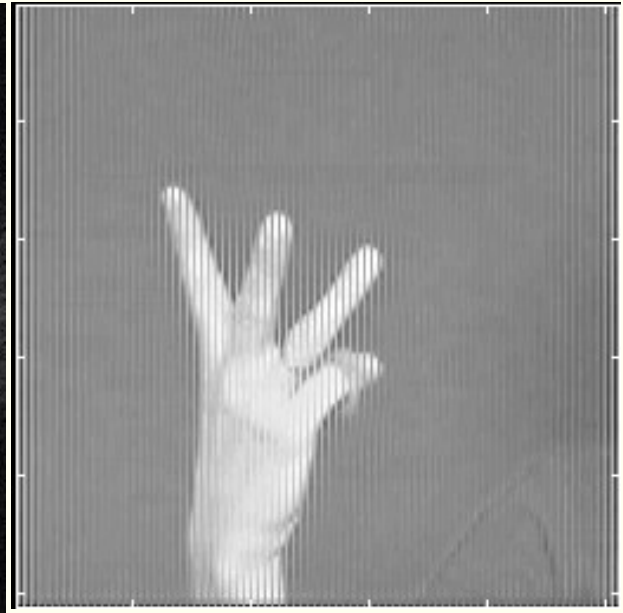
Przykład zastosowania filtru zaporowego



Obraz wyjściowy
zakłócony
periodycznie



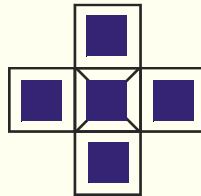
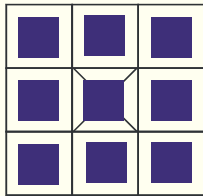
Widmo obrazu po
filtracji



Obraz po filtracji

Filtracja morfologiczna

- Transformacje morfologiczne wykorzystują element strukturalny ze zdefiniowanym dlań punktem odniesienia.



- Dla wykonania operacji morfologicznej punkt odniesienia elementu strukturalnego przykładany jest do każdego punktu obrazu
- Lokalny obszar obrazu pokryty elementem strukturalnym wykorzystywany jest do obliczenia nowej wartości punktu obrazu wskazywanego przez punkt odniesienia w sposób specyficzny dla danej operacji morfologicznej

Filtry morfologiczne

■ Dylatacja

Punktowi obrazu pokrywającemu się z punktem odniesienia elementu strukturalnego przypisuje się maksymalną spośród wartości pikseli obrazu, którym w elemencie strukturalnym odpowiada 1

■ Erozja

Punktowi obrazu pokrywającemu się z punktem odniesienia elementu strukturalnego przypisuje się minimalną spośród wartości pikseli obrazu, którym w elemencie strukturalnym odpowiada 1

■ Opening

Erozja, po której wykonywana jest dylatacja z tym samym elementem strukturalnym

■ Closing

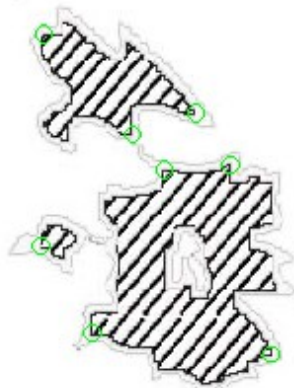
Dylatacja, po której wykonywana jest erozja z tym samym elementem strukturalnym

■ OC, CO – opening (closing), potem closing (opening) z tym samym elementem strukturalnym

Działanie filtrów morfologicznych na obiekty binarne



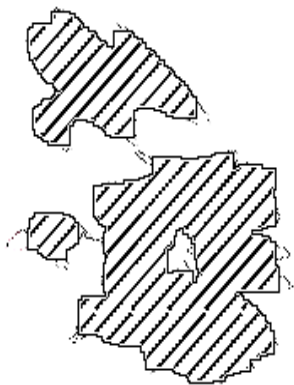
EROZJA: $X \ominus B$



DYLATACJA: $X \oplus B$



OTWARCIE: $X \circ B$



ZAMKNIĘCIE: $X \bullet B$

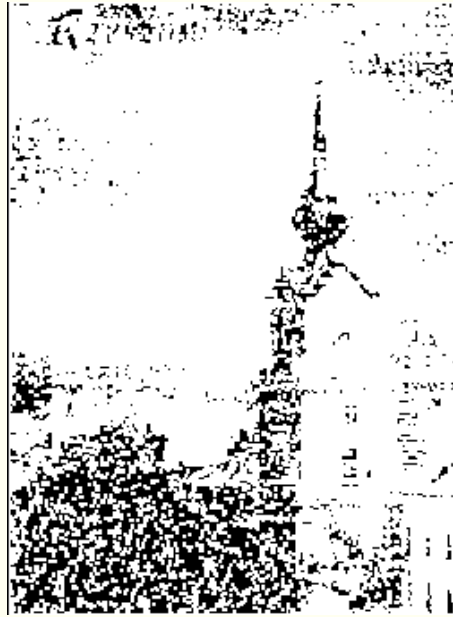


X - zbiór poddawany filtracji
 B - element strukturalny
zacieniowany obszar - obiekt,
ciemna linia - linie brzegowe
obiektu po filtracji,
jasna linia - linie brzegowe
obiektu przed filtracją

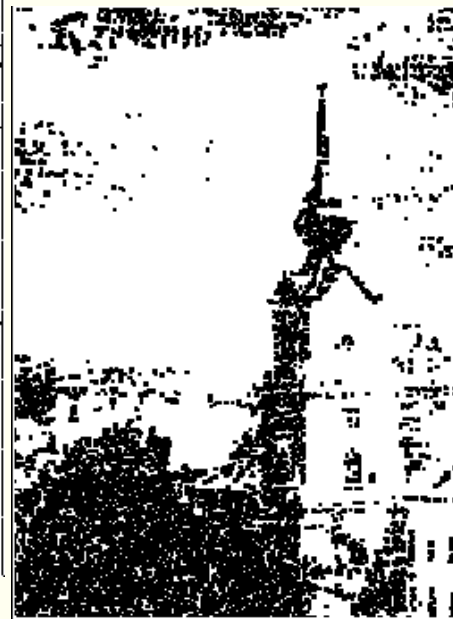
Przykład filtracji morfologicznej



(a)



(b)



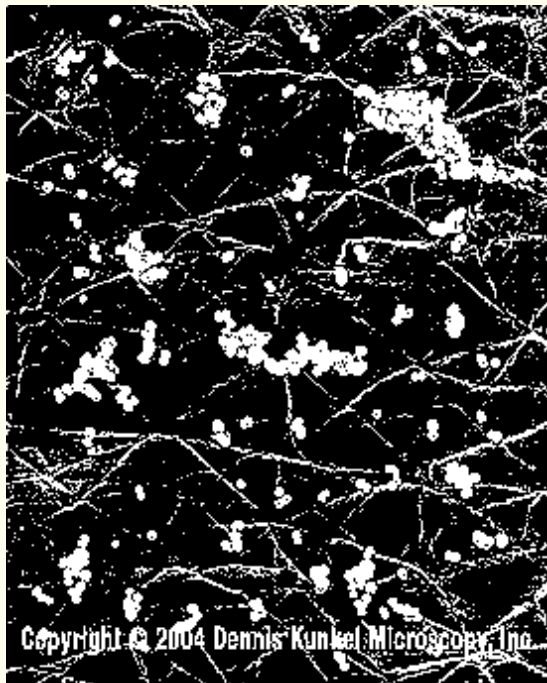
(c)



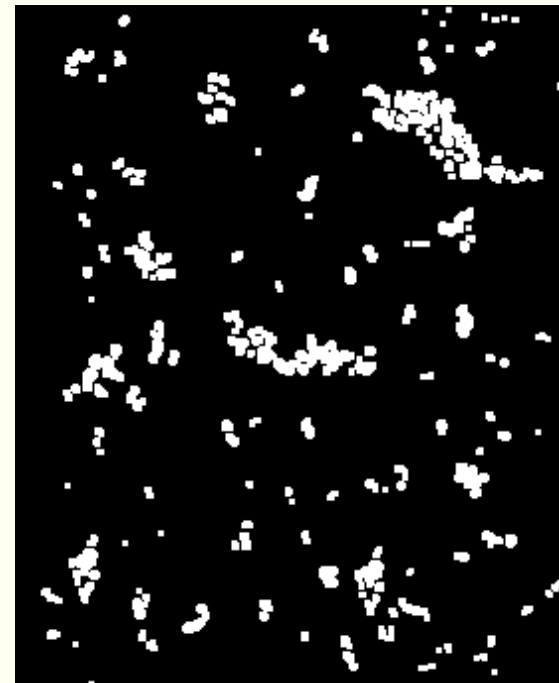
(d)

- a) obraz binarny z zakłóceniem typu „sól i pieprz” ($d = 0.1$)
- b) obraz (a) po dylatacji z elementem strukturalnym 3x3
- c) obraz (a) po closingu z elementem strukturalnym 3x3
- d) obraz (a) po filtracji CO z elementem strukturalnym 3x3

Zastosowanie openingu do eliminacji linii na obrazie



a) obraz binarny



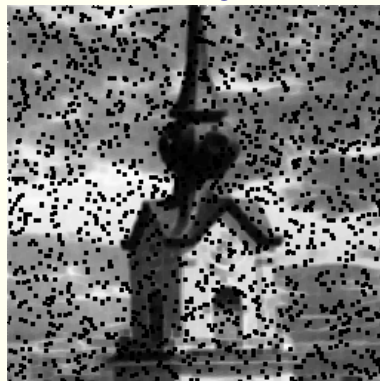
b) obraz (a) po otwarciu

Przykład filtracji morfologicznej obrazu szarego

oryginał



erozja



otwarcie



OC



dylatacja



zamknięcie



CO



Metody pozwalające na usunięcie nieistotnych informacji

1. Binarizacja
2. Wydzielanie krawędzi

Binaryzacja przez progowanie

Polega na przekształceniu obrazu oryginalnego do obrazu czarno-białego, w którym piksele o wartości przekraczającej ustalony próg T zastępowane są kolorem białym, pozostałe czarnym:

$$BW(i, j) = \begin{cases} 1 & \text{gdy } I(i, j) \geq T \\ 0 & \text{gdy } I(i, j) < T \end{cases}$$



progowanie



Progowanie - przykłady



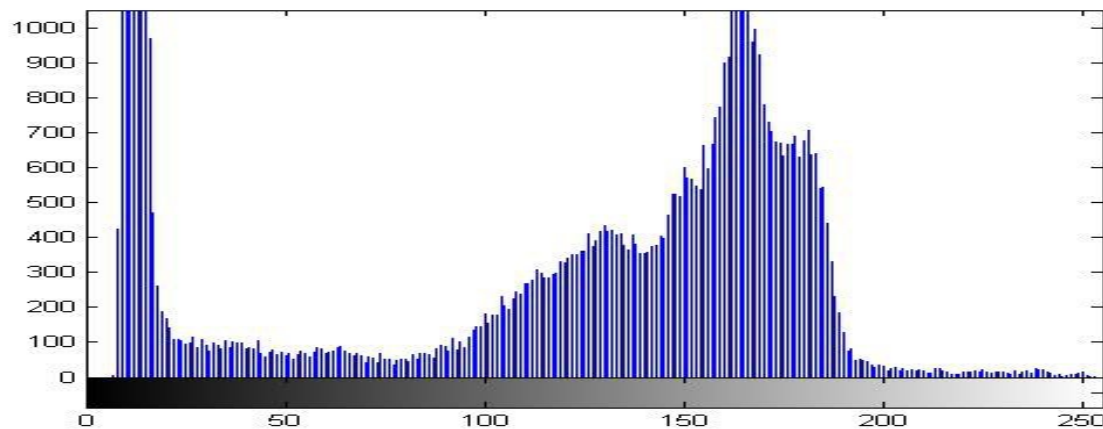
$T = 88$



$T = 30$

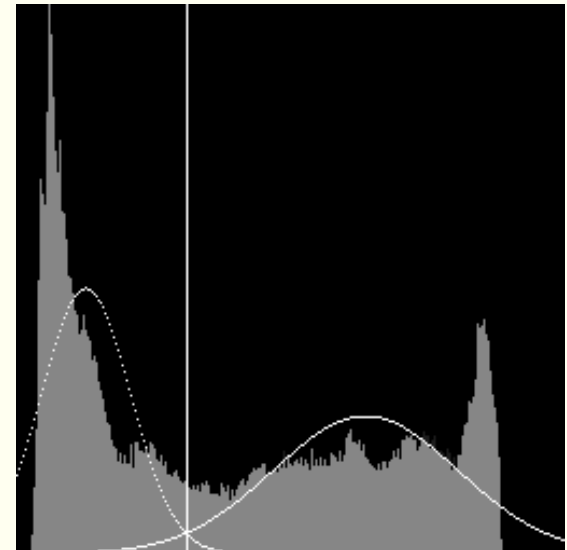


$T = 140$

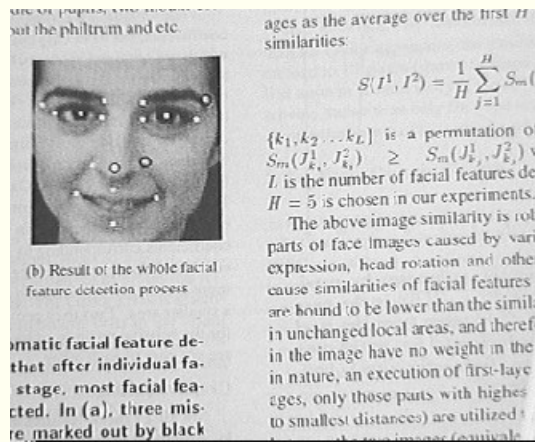


Metody znajdowania progu

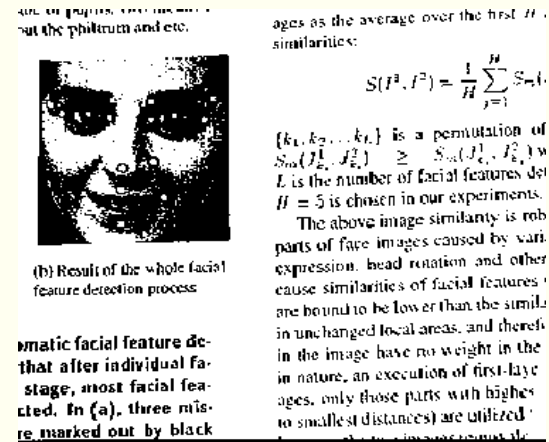
- **Iteracyjna** – w pierwszym przybliżeniu zakłada, że cztery narożne piksele należą do tła, pozostałe do obiektu; w kolejnych iteracjach wyznaczany jest nowy próg jako średnia arytmetyczna wartości średnich otrzymanych dla pikseli tła i obiektu; za ostateczny próg wybierana jest ta wartość, która nie zmieniła się w kolejnej iteracji
- **Metoda Otsu** – jako próg przyjmuje się wartość leżącą na przecięciu dwóch krzywych Gaussa przybliżających histogram



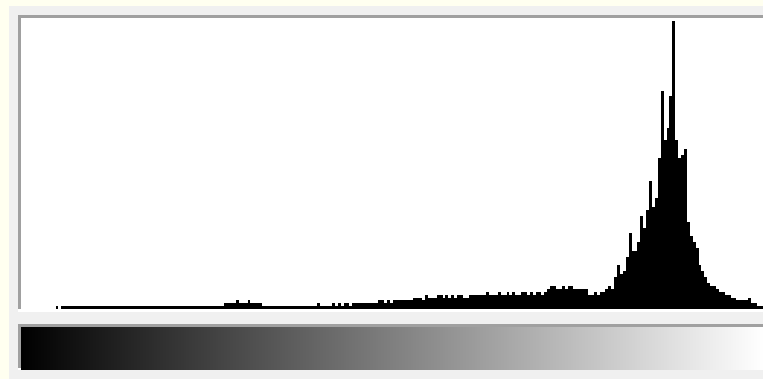
Przykład działania iteracyjnej metody wyznaczania progu



obraz oryginalny



obraz po progowaniu (próg=0.567)



histogram obrazu oryginalnego

Wydzielanie krawędzi

■ Operatory gradientowe

- ◆ Operatory kierunkowe
- ◆ Operator Laplace'a
- ◆ Filtr LOG (Laplacian Of Gaussian)

■ Górnoprzepustowe filtry częstotliwościowe

■ Gradient morfologiczny

Obraz krawędzi uzyskuje się przez odjęcie obrazu poddanego erozji od obrazu otrzymanego w wyniku dylatacji

Operatory gradientowe

Operator Roberta

$$h_1 = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} \quad h_2 = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}$$

Operator Laplace'a

$$h_1 = \begin{bmatrix} 0 & 1 & 0 \\ 1 & -4 & 1 \\ 0 & 1 & 0 \end{bmatrix} \quad h_2 = \begin{bmatrix} 1 & 1 & 1 \\ 1 & -8 & 1 \\ 1 & 1 & 1 \end{bmatrix}$$

Operator Prewitta

$$h_1 = \begin{bmatrix} 1 & 1 & 1 \\ 0 & 0 & 0 \\ -1 & -1 & -1 \end{bmatrix} \quad h_2 = \begin{bmatrix} 0 & 1 & 1 \\ -1 & 0 & 1 \\ -1 & -1 & 0 \end{bmatrix} \quad h_3 = \begin{bmatrix} -1 & 0 & 1 \\ -1 & 0 & 1 \\ -1 & 0 & 1 \end{bmatrix}$$

Uwaga

Maski h_1, h_2, \dots, h_8 odpowiadają kolejno orientacjom krawędzi $0^\circ, -45^\circ, -90^\circ, \dots$
Suma wag zawsze równa 0 (filtr górnoprzepustowy)

Operatory gradientowe c.d.

Operator Sobela

$$h_1 = \begin{bmatrix} 1 & 2 & 1 \\ 0 & 0 & 0 \\ -1 & -2 & -1 \end{bmatrix} \quad h_2 = \begin{bmatrix} 0 & 1 & 2 \\ -1 & 0 & 1 \\ -2 & -1 & 0 \end{bmatrix} \quad h_3 = \begin{bmatrix} -1 & 0 & 1 \\ -2 & 0 & 2 \\ -1 & 0 & 1 \end{bmatrix}$$

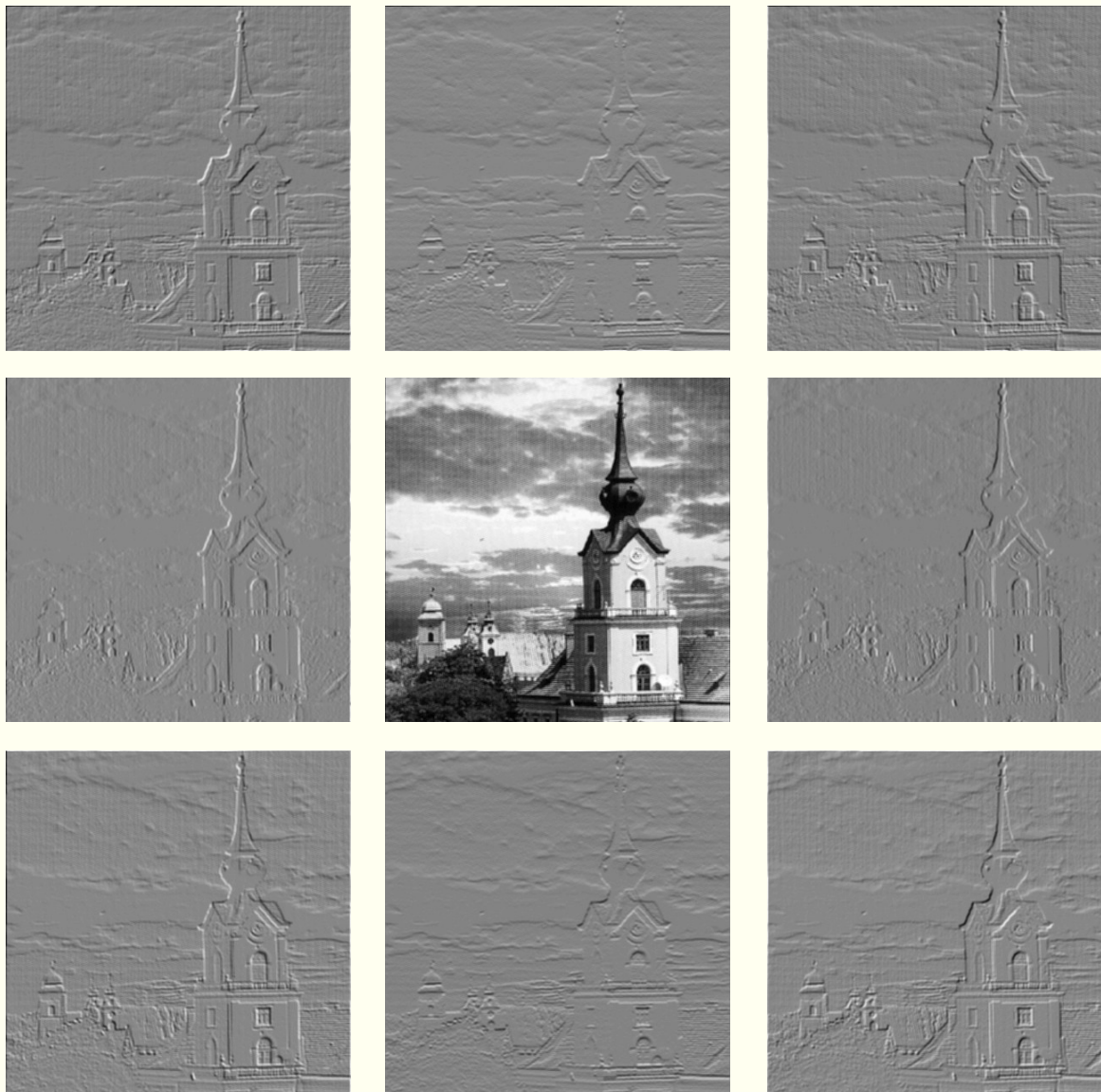
Operator Robinsona

$$h_1 = \begin{bmatrix} 1 & 1 & 1 \\ 1 & -2 & 1 \\ -1 & -1 & -1 \end{bmatrix} \quad h_2 = \begin{bmatrix} 1 & 1 & 1 \\ -1 & -2 & 1 \\ -1 & -1 & 1 \end{bmatrix} \quad h_3 = \begin{bmatrix} -1 & 1 & 1 \\ -1 & -2 & 1 \\ -1 & 1 & 1 \end{bmatrix}$$

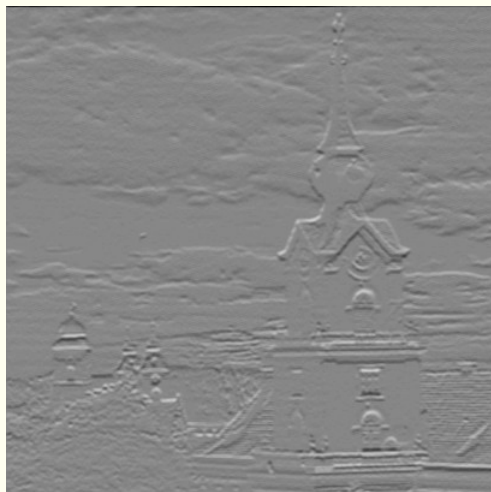
Operator Kirscha

$$h_1 = \begin{bmatrix} 3 & 3 & 3 \\ 3 & 0 & 3 \\ -5 & -5 & -5 \end{bmatrix} \quad h_2 = \begin{bmatrix} 3 & 3 & 3 \\ -5 & 0 & 3 \\ -5 & -5 & 3 \end{bmatrix} \quad h_3 = \begin{bmatrix} -5 & 3 & 3 \\ -5 & 0 & 3 \\ -5 & 3 & 3 \end{bmatrix}$$

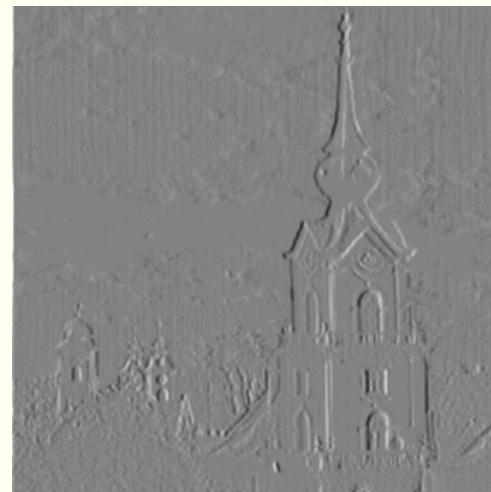
Przykład zastosowania operatora Sobela



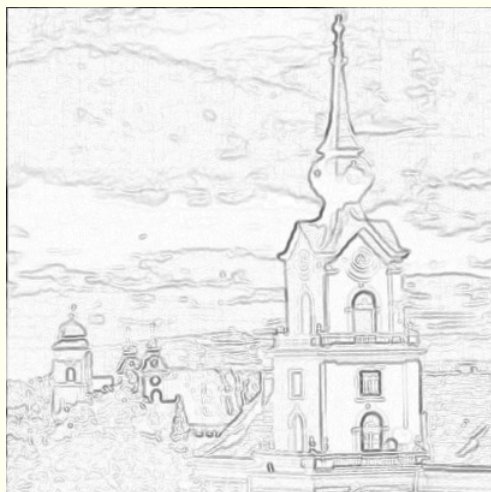
Wydobywanie konturów



Krawędzie poziome



Krawędzie pionowe



Kontury (Euklidesowa kombinacja obrazów krawędzi poziomych i pionowych)



Kontury (modułowa kombinacja obrazów krawędzi poziomych i pionowych)

Przykład zastosowania operatorów Laplace'a i LOG

Oryginał 3×3



Laplasjan



LOG: $n=5$, $\sigma=0.5$



LOG: $n=7$, $\sigma=1$

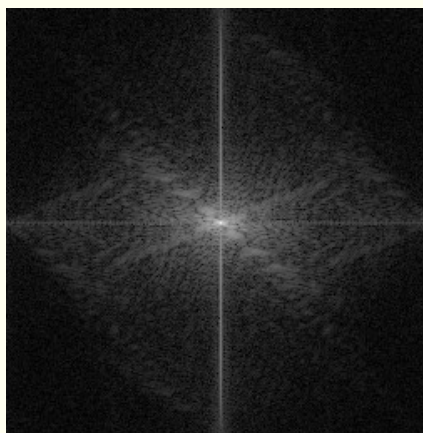


Górnoprzepustowe filtry częstotliwościowe – filtr idealny

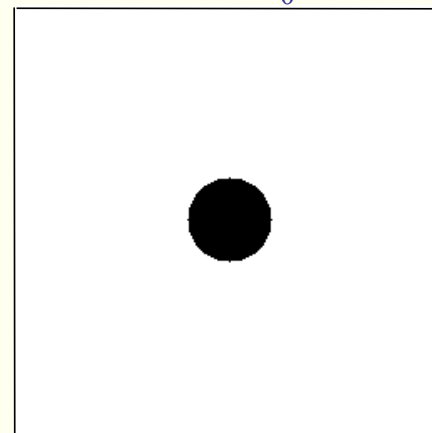
obraz wejściowy



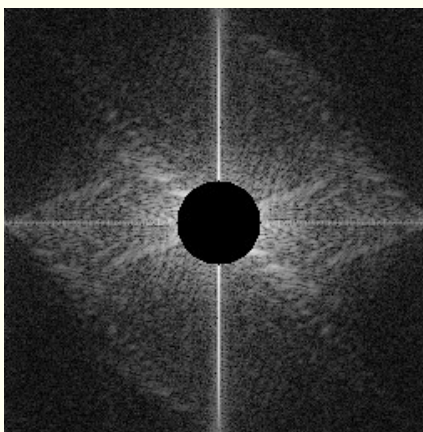
$|G(u, v)|$



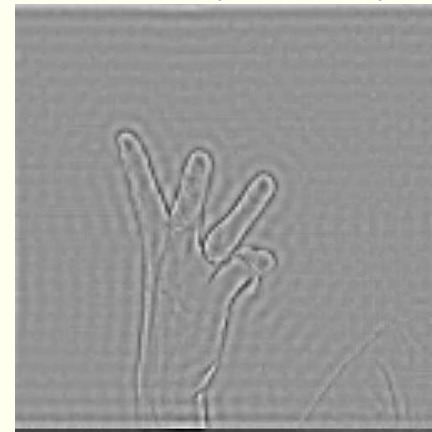
$|H(u, v)|, \omega_0 = 50$



$|G(u, v)H(u, v)|$



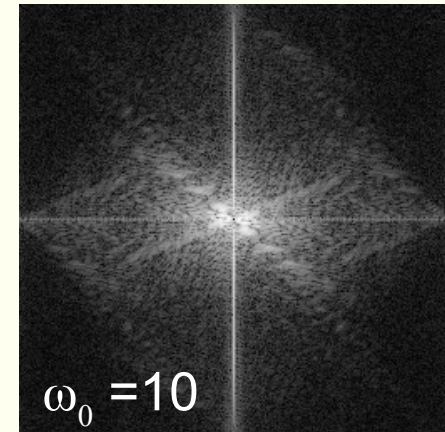
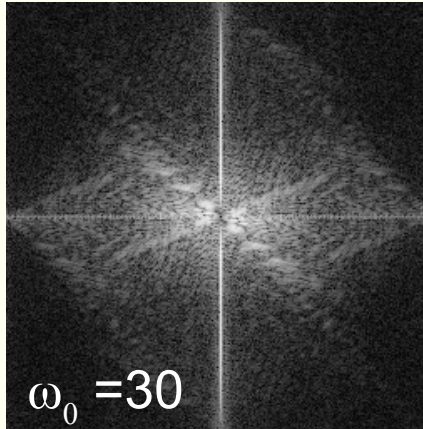
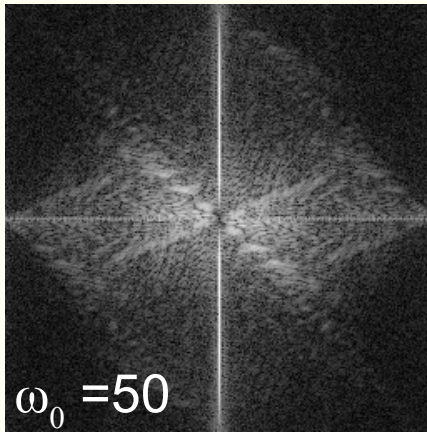
obraz wynikowy



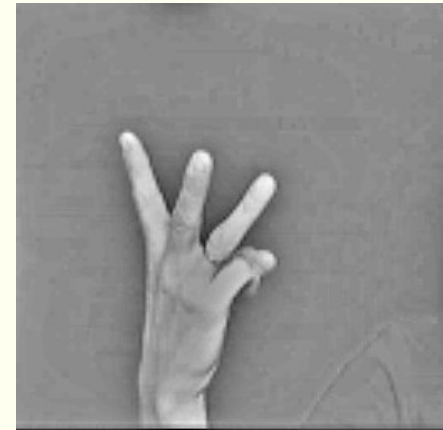
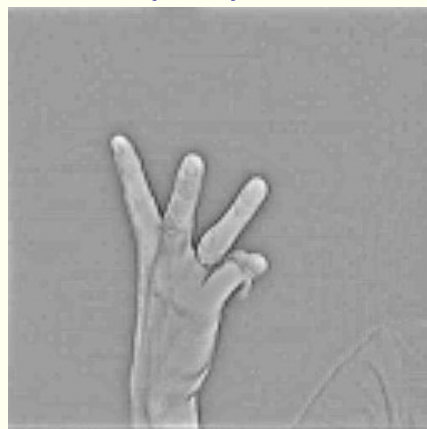
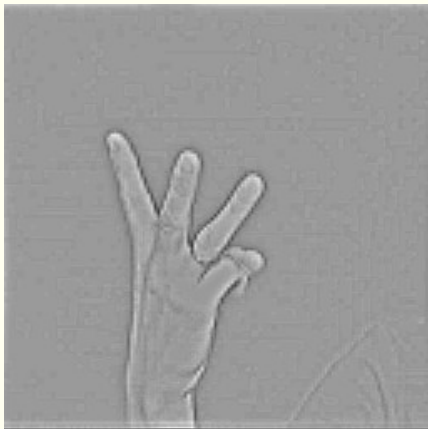
$G(u, v)$
transformata obrazu
wejściowego
 $H(u, v)$
transformata filtru

Górnoprzepustowy filtr Butterwortha

$$|G(u, v)H(u, v)|$$



obrazy wynikowe

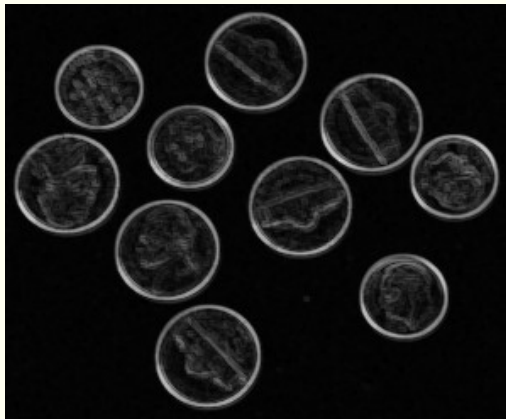


Gradient morfologiczny

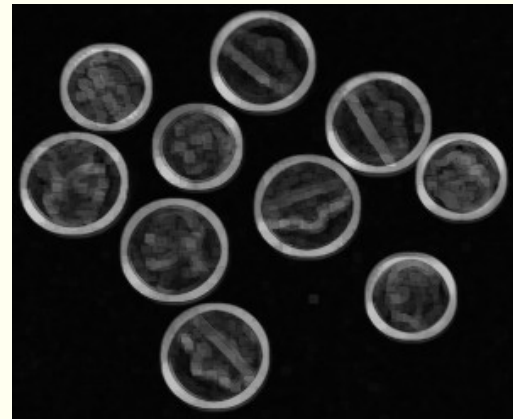
oryginał (300×246)



3×3



5×5

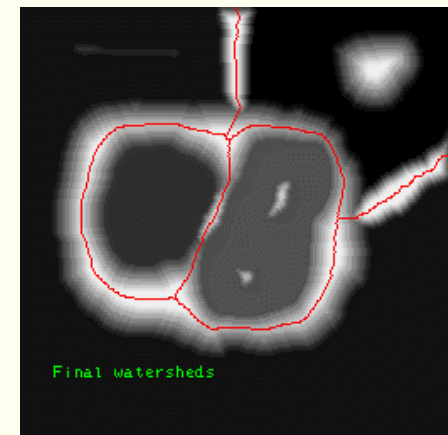
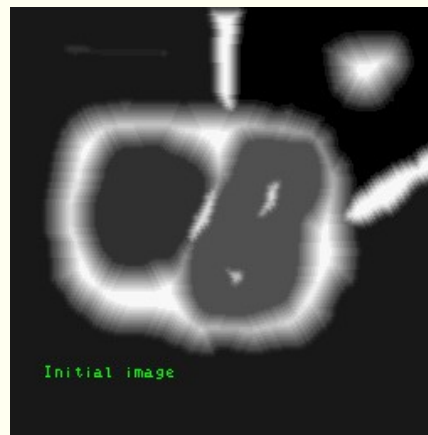
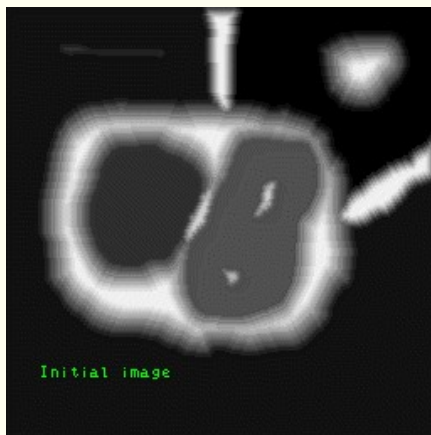


Wydzielanie obiektów

1. Segmentacja morfologiczna (watershed)
2. Wydzielanie obiektów na podstawie sekwencji obrazów
3. Metoda wykorzystująca model koloru obiektu
4. Obrazy historii ruchu i obrazy energii ruchu
5. Śledzenie obiektów

Segmentacja morfologiczna (watershed)

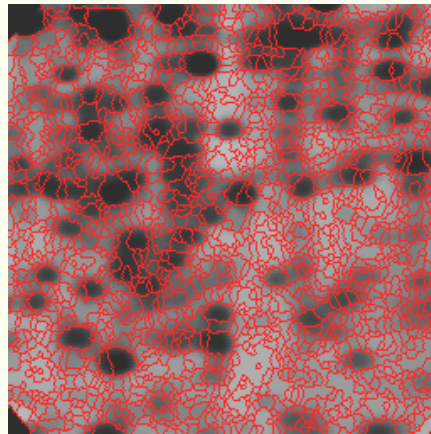
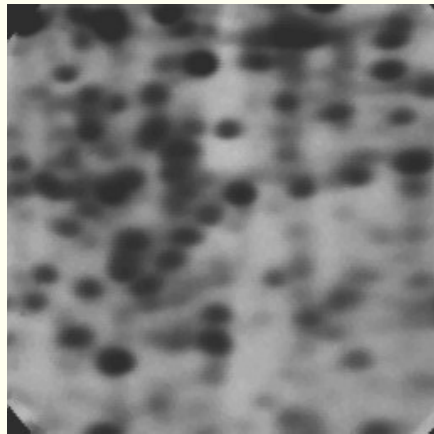
1. Obraz szary traktuje się jako powierzchnię topologiczną
2. Powierzchnię tę „zalewa się” rozpoczynając od najniżej położonego punktu, nie dopuszczając do połączenia się zalewisk; w efekcie obraz zostaje podzielony na 2 zbiory:
 1. obszary zalane (zalewiska) i
 2. linie działowe



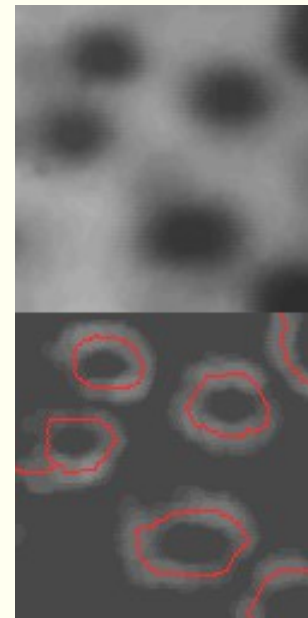
Segmentacja watershed - problem

Jeśli taką transformację zastosujemy do obrazu gradientu, to obszary zalane teoretycznie powinny odpowiadać obszarom o zbliżonym odcieniu szarości tego obrazu

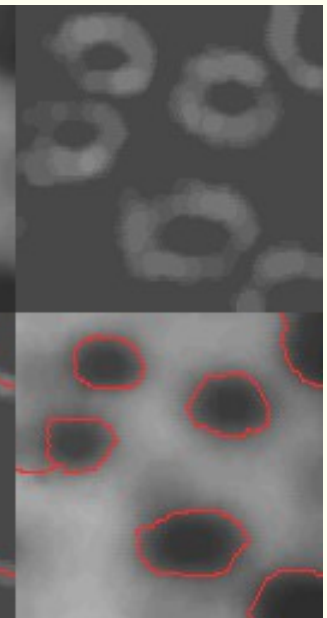
W praktyce w wyniku transformacji powstaje zbyt dużo fałszywych linii działowych ze względu na lokalne nieregularności w obrazie gradientu



Obraz wejściowy



Obraz gradientu

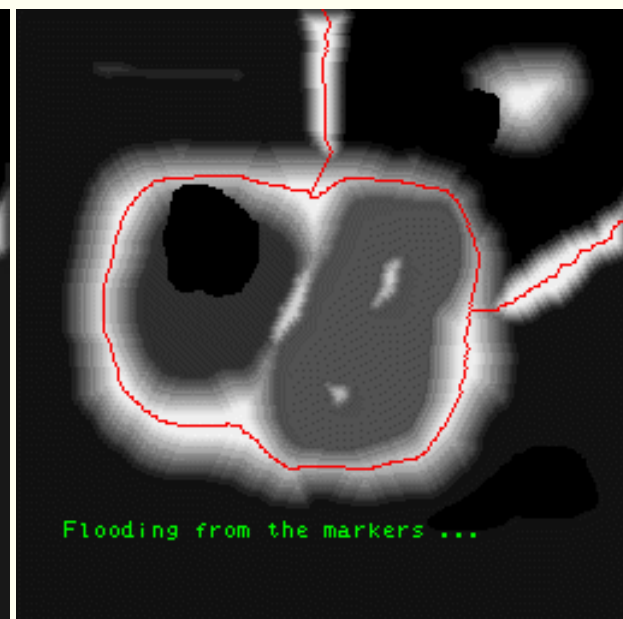
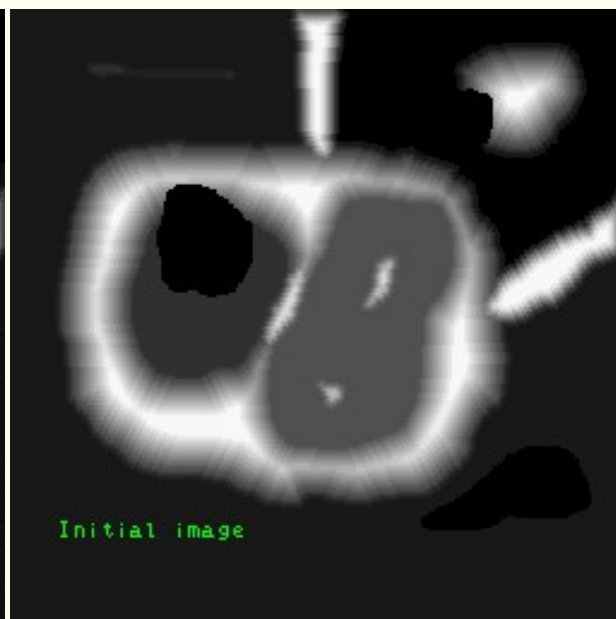
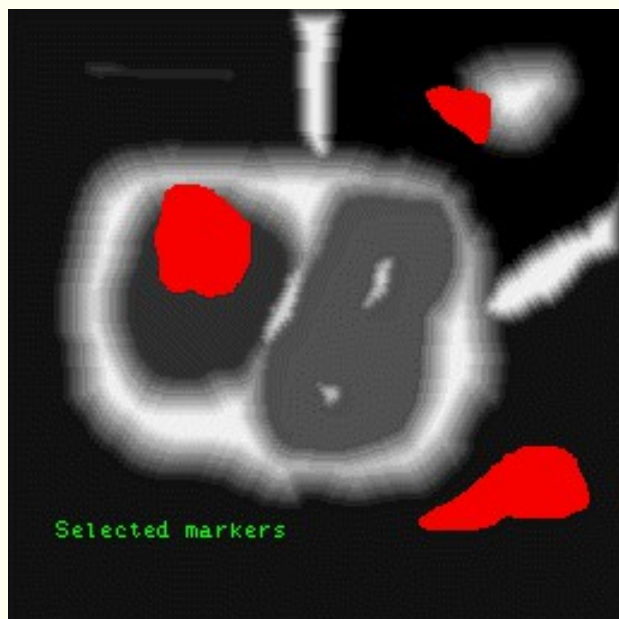


Obraz gradientu
z wynikiem
transformacji
watershed

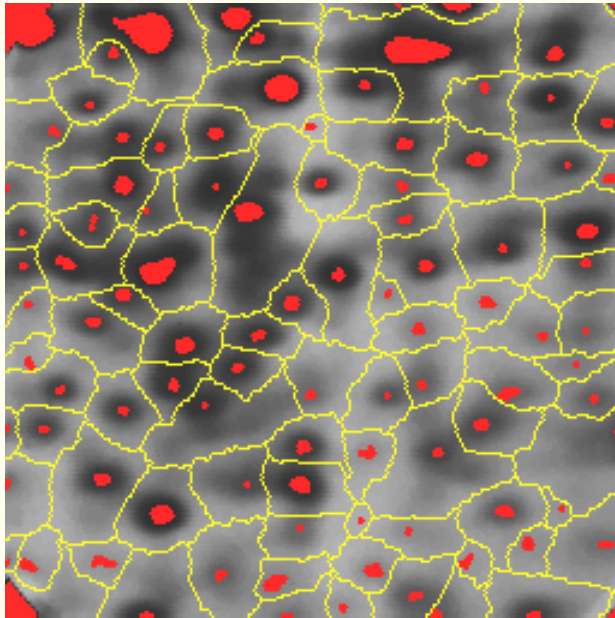
Kontury wynikowe

Rozwiązanie problemu – segmentacja watershed oparta na znacznikach

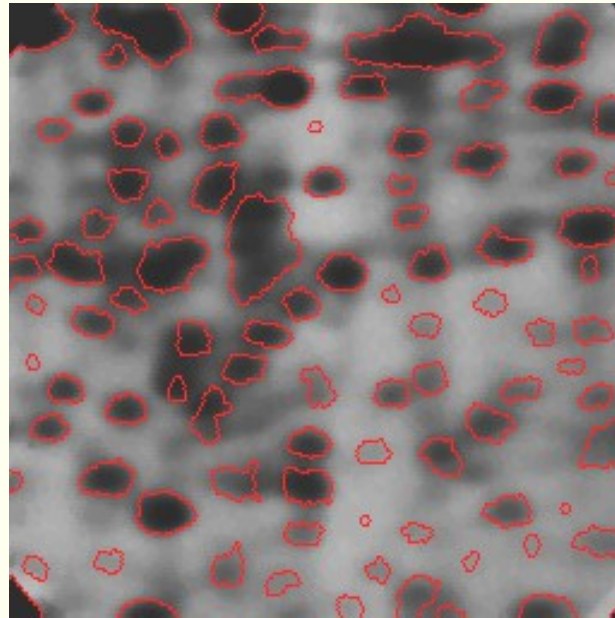
Problem fałszywych linii działowych eliminuje się pozwalając na „zalewanie” obrazu (powierzchni topologicznej) od wcześniej zdefiniowanych miejsc (znaczników)



Przykład segmentacji watershed opartej na znacznikach



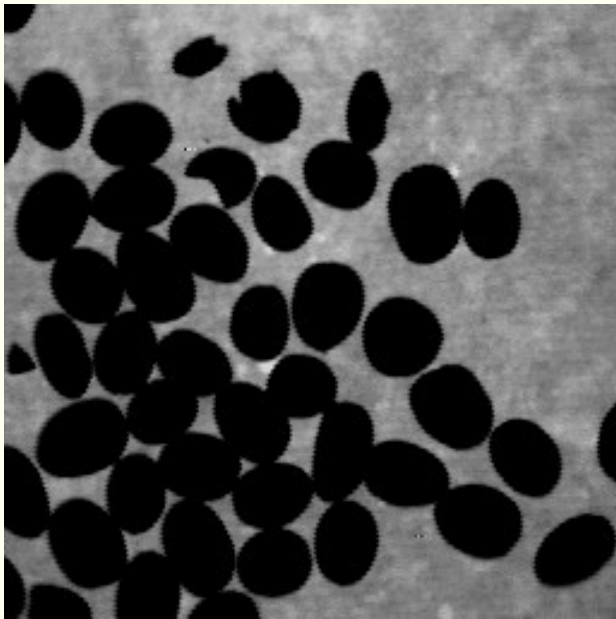
Obraz ze znacznikami



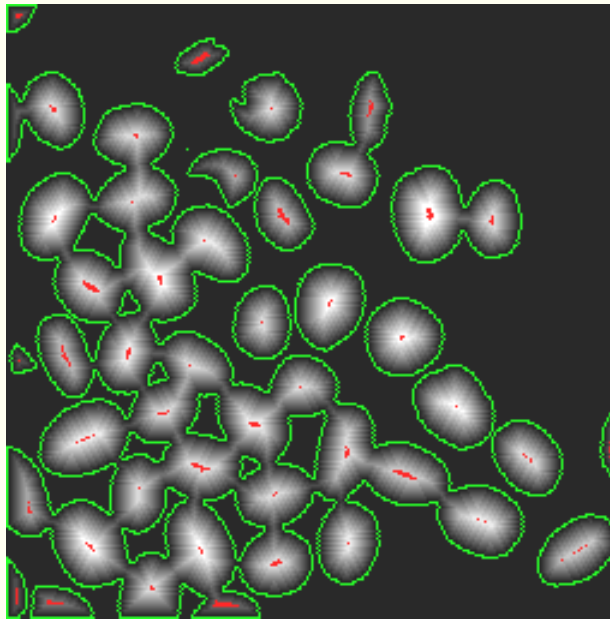
Obraz z wynikiem transformacji watershed przeprowadzonej na obrazie gradientu

Przykład

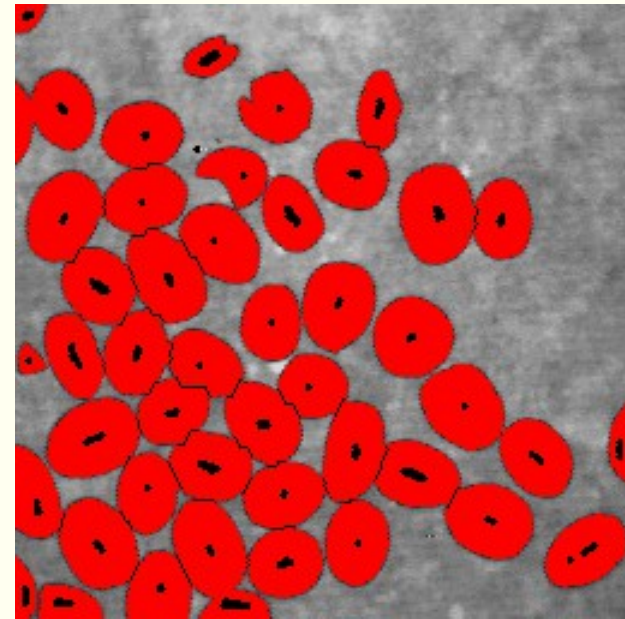
Oddzielanie ziaren kawy



Obraz szary
ziaren kawy



Obraz po wyznaczeniu
funkcji odległości



Wynik transformacji
watershed zastosowanej do
obrazu odległości
(znacznikami były maksima
funkcji odległości)

Funkcja odległości

Niech Y – zbiór w przestrzeni Z^2

Funkcja odległości definiowana jest następująco:

$$\forall y \in Y, d(y) = \text{dist}(y, Y^c)$$

Czyli każdy punkt y ze zbioru Y otrzymuje wartość równą minimalnej odległości od tła (dopełnienia Y^c zbioru Y). Pozostałe punkty otrzymują wartość 0.



Obraz binarny



Funkcja odległości

Wydzielanie obiektów na podstawie sekwencji obrazów

Metody:

1. Odejmnowanie tła od obrazów z sekwencji
2. Odejmnowanie obrazów (detekcja obszarów, w których wystąpił ruch)

Ad 1)



Ad 2)



Charakterystyka metod detekcji obiektów na podstawie sekwencji obrazów

1. Metoda odejmowania tła nie nadaje się do użycia, jeśli tło się zmienia, np. w ruchliwym otoczeniu lub przy zmieniających się warunkach oświetleniowych
2. Metoda odejmowania obrazów wskazuje jedynie miejsca, w których wystąpił ruch, nie daje możliwości określenia kształtu obiektu

Metoda wykorzystująca model koloru obiektu

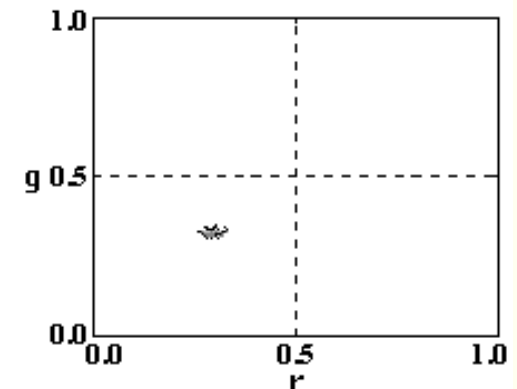
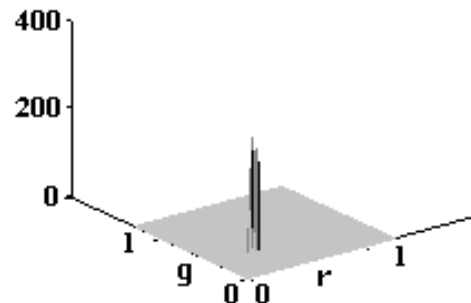
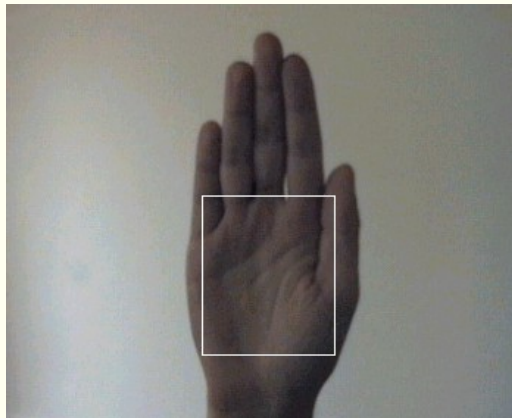
- W celu zmniejszenia zależności od oświetlenia wykorzystuje się normalizację $(R, G, B) \Rightarrow (r, g)$:

$$r = R / L$$

$$g = G / L$$

gdzie $L = R + G + B$.

- Model koloru stanowi 2-wymiarowy histogram wartości pikseli należących do próbki koloru w znormalizowanej przestrzeni (r, g)

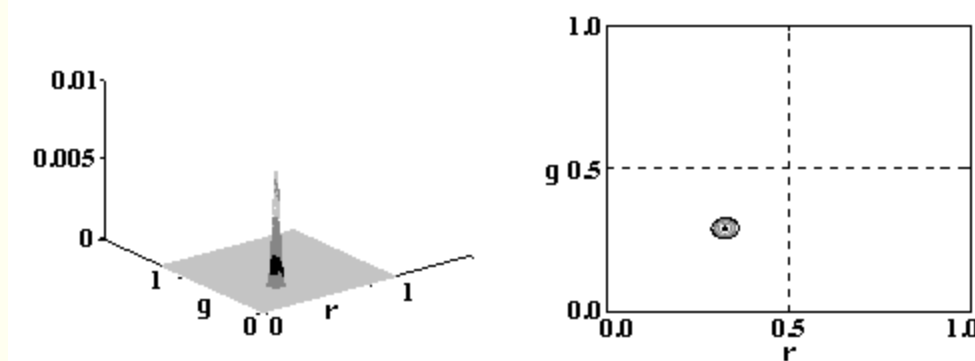


Metoda wykorzystująca model koloru obiektu c.d.

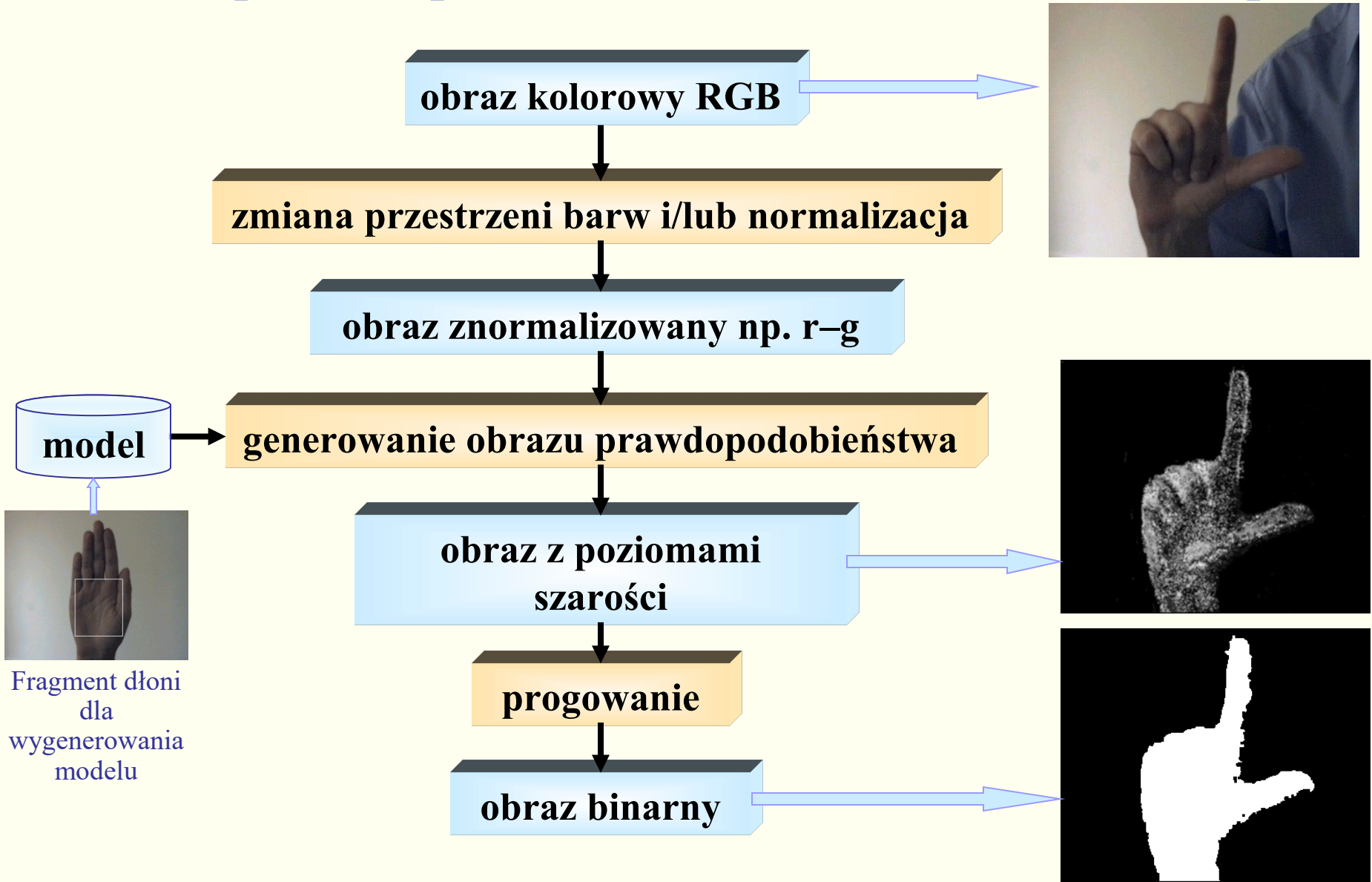
- Histogram ten można aproksymować dwuwymiarowym rozkładem Gaussa. Wtedy model koloru opisany jest za pomocą wartości oczekiwanej μ , i macierzy kowariancji C ,

$$\mu = \begin{bmatrix} \mu_r \\ \mu_g \end{bmatrix} \quad C = \begin{bmatrix} \delta_{rr} & \delta_{rg} \\ \delta_{gr} & \delta_{gg} \end{bmatrix}$$

które wyznacza się ze znanych ze statystyki wzorów.



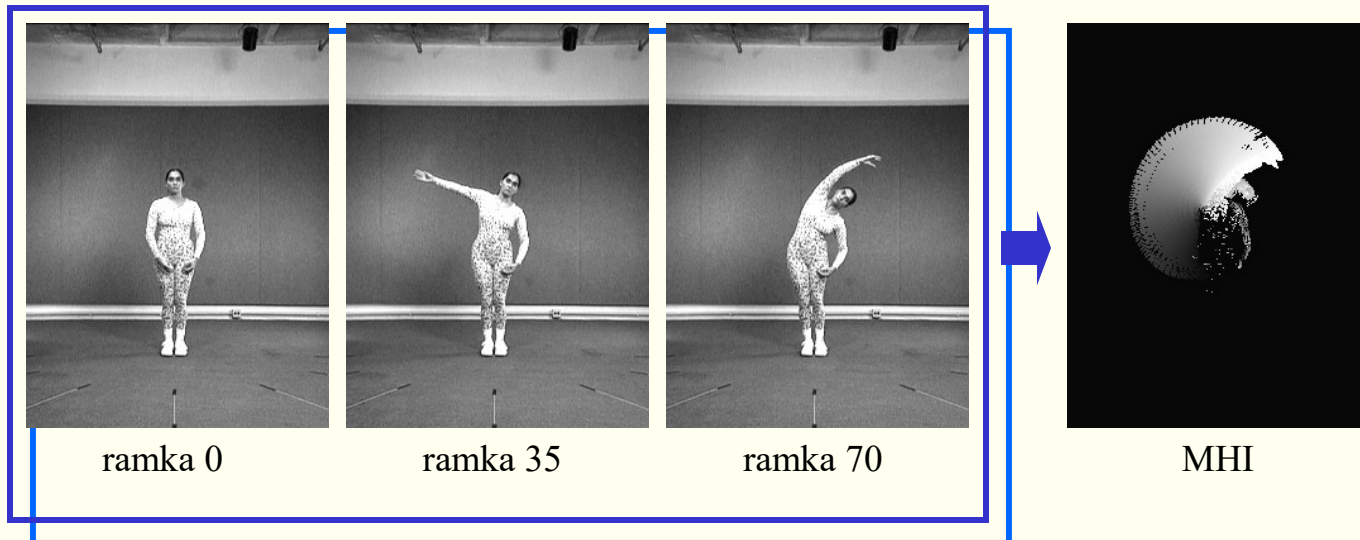
Wydzielanie obiektów z wykorzystaniem modelu barwy



Obraz historii ruchu

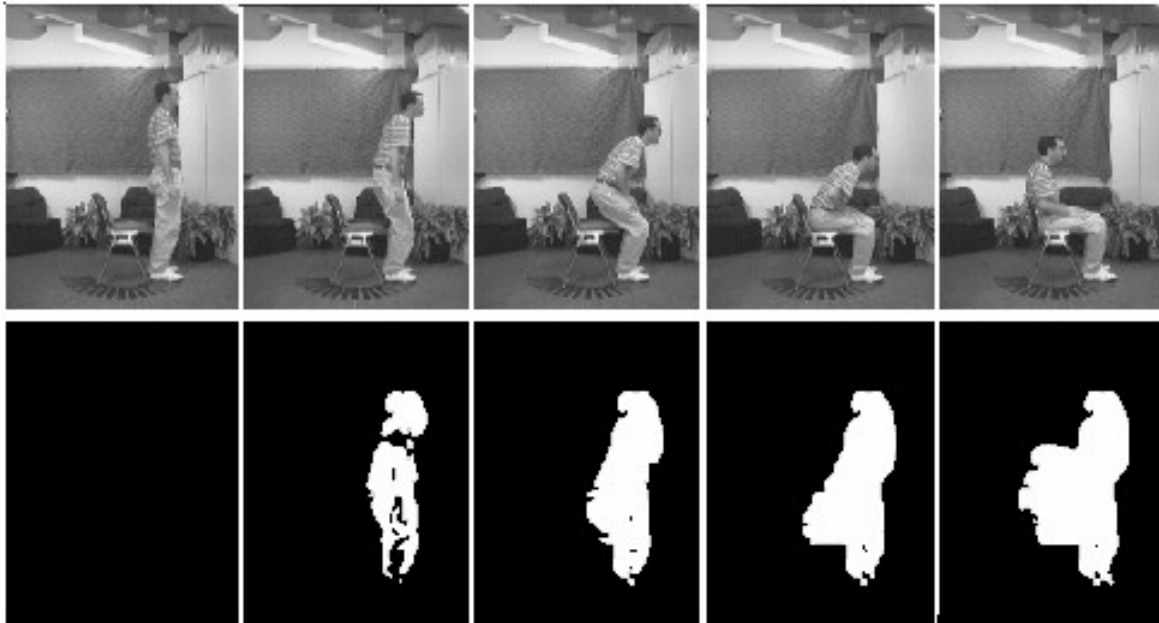
Obraz historii ruchu (MHI – Motion History Image) – obraz, na którym intensywność piksela jest funkcją historii ruchu w danym położeniu; jaśniejsze piksele odpowiadają ruchowi mniej odległemu w czasie

Wzorce te są generowane przez sumowanie różnic kolejnych obrazów poruszającego się obiektu



Obraz energii ruchu

Obraz energii ruchu (MEI – Motion Energy Image) - zbinaryzowany obraz MHI, pokazuje obszary, w których wystąpił ruch



Konieczne może być odfiltrowanie szumów z użyciem znanych metod przetwarzania wstępnego

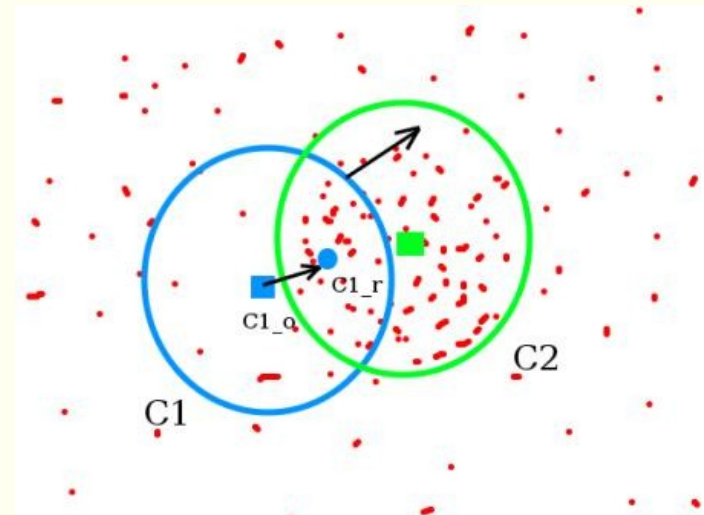
Śledzenie obiektów

Algorytmy:

- ◆ MeanShift
- ◆ CamShift (Continuous Adaptive Mean-SHIFT)

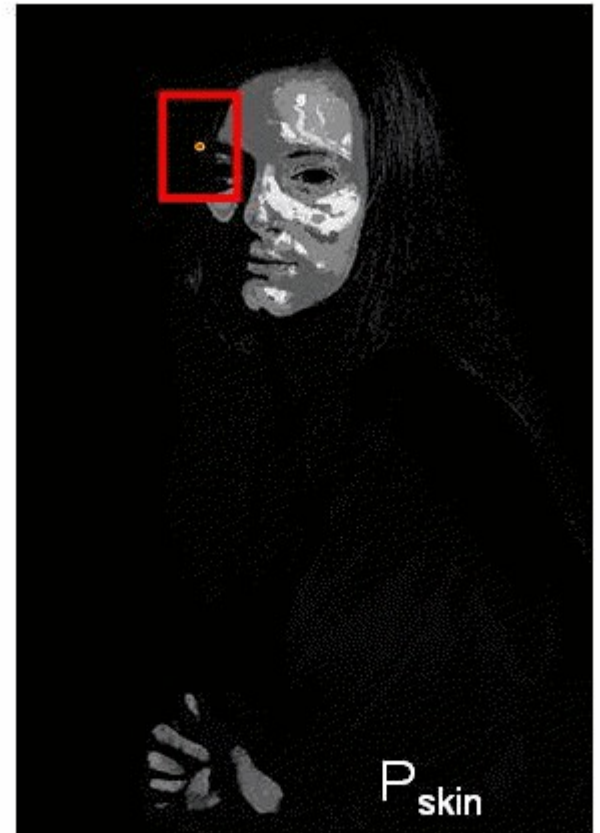
MeanShift - idea

- 1) Wskazać początkową lokalizację śledzonego obiektu (C1)
- 2) Przesunąć okno z poprzedniej lokalizacji do pozycji (C1_o), w której zagęszczenie punktów będzie największe (C1_r), tzn. tak, by jego środek pokrywał się ze środkiem ciężkości punktów zawartych w oknie
- 3) Jeżeli środek okna pokrywa się ze środkiem ciężkości punktów w oknie (lub odległość między tymi punktami jest mniejsza niż założony próg), to koniec, w przeciwnym razie wykonać ponownie punkt 2



CamShift - idea

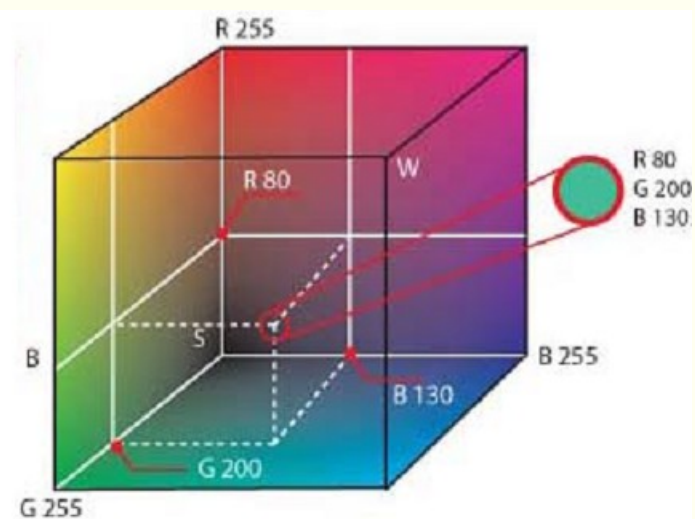
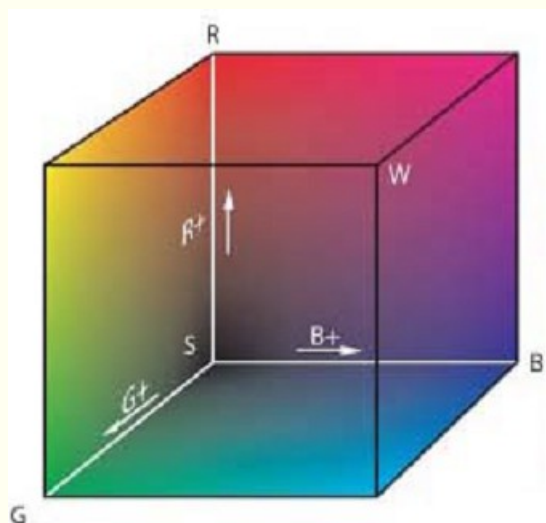
- 1) Wskazać początkową lokalizację śledzonego obiektu
- 2) Znaleźć nową lokalizację przy użyciu algorytmu MeanShift
- 3) Uaktualnić rozmiar okna wg wzoru: $s = 2 \times \sqrt{\frac{M_{00}}{256}}$ oraz obliczyć parametry odpowiadającej mu elipsy
- 4) Ponownie zastosować algorytm MeanShift w nowym oknie
- 5) Powtarzać punkty 3 i 4 aż do uzyskania założonej dokładności



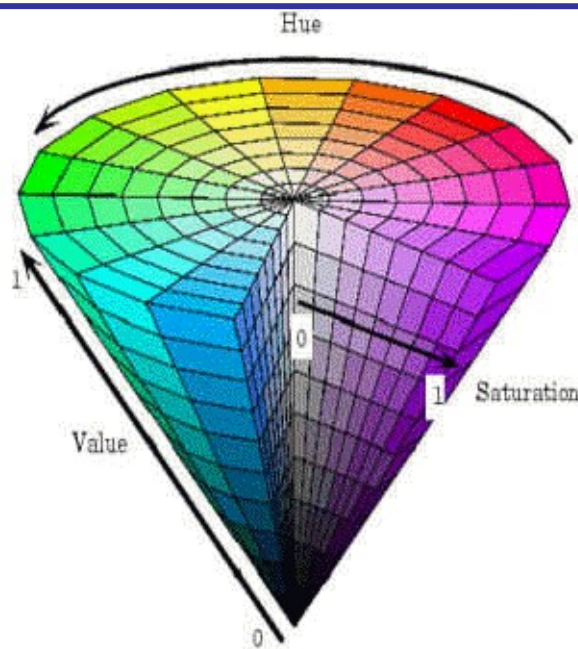
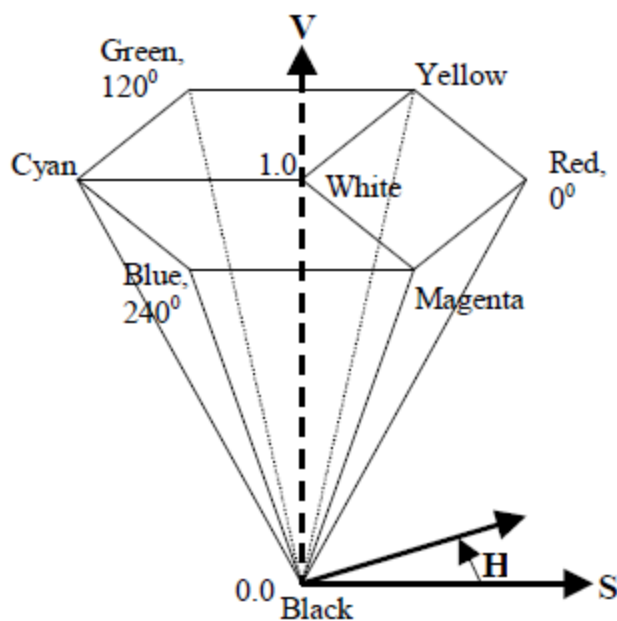
Mean shift window
initialization

Przestrzenie barw RGB i HSV

RGB



HSV



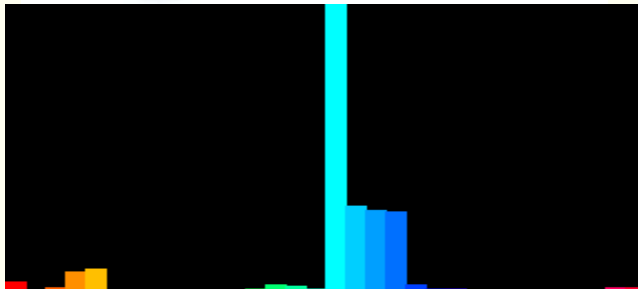
Konwersja z przestrzeni RGB do HSV

$$H = \cos^{-1} \{ [(R-G) + (R-B)] / [2 \cdot \sqrt{((R-G)^2 + (R-G)(G-B))}] \}$$

$$S = 1 - 3 \cdot \min(R, G, B) / (R+G+B)$$

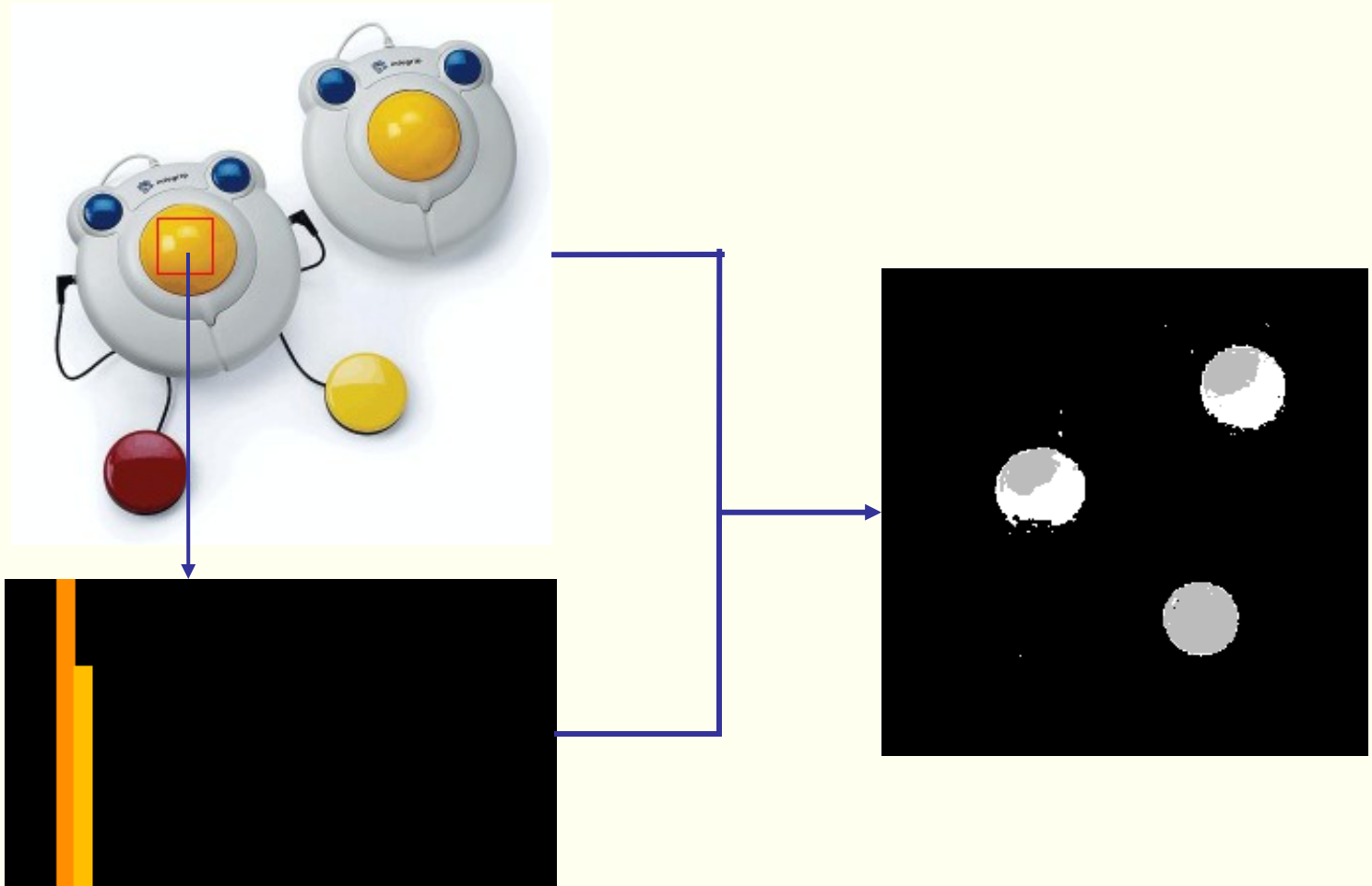
$$V = (R+G+B)/3$$

Histogram koloru

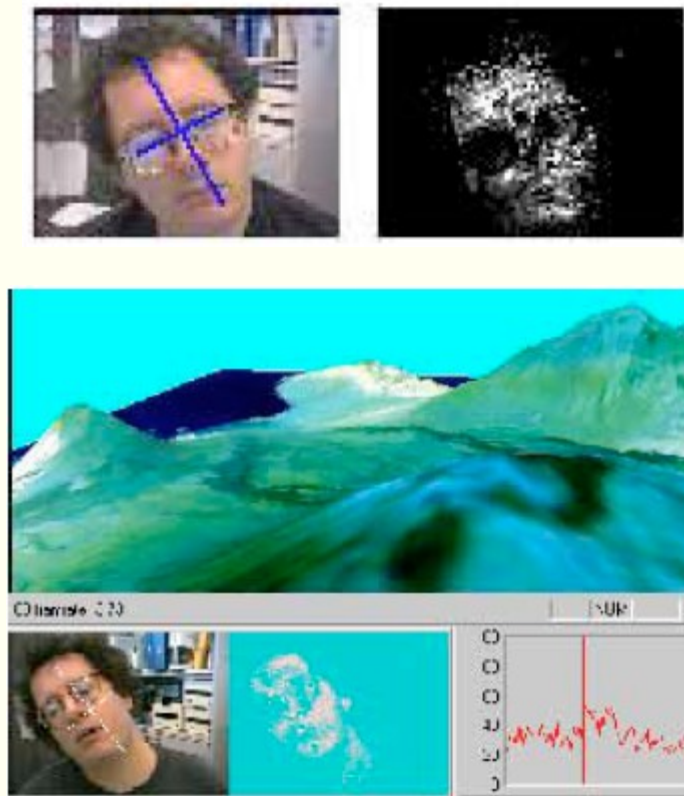


- Uwzględniana jest tylko składowa barwy (dla przestrzeni HSV jest to składowa H (*hue*))
- Zakres wartości dzielony jest na N podprzedziałów. Wartość odpowiadająca i -temu podprzedziałowi odpowiada liczbie tych pikseli obrazu, dla których wartość składowej barwy należy do podprzedziału i .
- Histogram jest to wykres słupkowy pokazujący wartości w kolejnych podprzedziałach

Histogram backprojection



Przykład: Wykorzystanie orientacji głowy do sterowania aplikacją



Źródło: Gary Bradski, Microcomputer Research Lab, Santa Clara, CA, Intel Corp.

Klasyfikacja - pojęcia

Klasyfikacja: przypisanie obiektu do określonej klasy

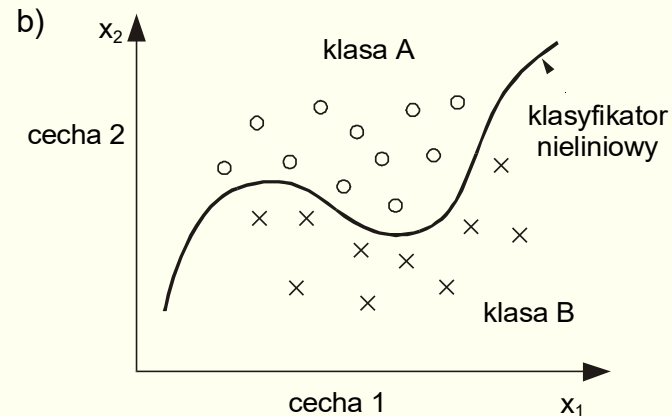
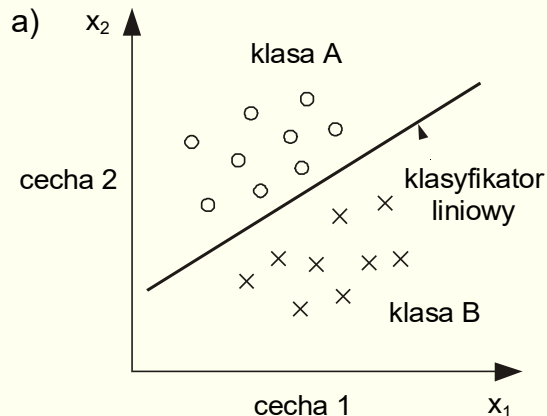
Przestrzeń cech: n -wymiarowa przestrzeń, w którą są odwzorowywane rozważane obiekty

Wektor cech: zbiór wartości cech dla obiektu o

$$x(o) = [x_1(o), x_2(o), \dots, x_n(o)]^T$$

Klasyfikator: dzieli przestrzeń cech na rozłączne obszary, każdy obszar odpowiada jednej klasie.

Przykład



Metoda KNN

(k – najbliższych sąsiadów)

Rozpoznawany obiekt zalicza się do klasy, która jest najliczniej reprezentowana wśród k najbliższych mu obiektów ze zbioru uczącego.

Nie ma ogólnych recept na wybór liczby k .

Metoda najbliższego sąsiada (NN)

jest to szczególny przypadek reguły k - najbliższych sąsiadów dla $k = 1$.

Niezmienne momentowe

- Dają zbliżone wartości niezależnie od rozmiaru i ułożenia obiektu na obrazie
- Wyznaczane są z wykorzystaniem wzoru na moment centralny rzędu pq m_{pq} :

$$m_{pq} = \sum_i \sum_j (i - \bar{i})^p (j - \bar{j})^q f(i, j),$$

$$\bar{i} = M_{10} / M_{00}, \quad \bar{j} = M_{01} / M_{00}$$

$$M_{pq} = \sum_i \sum_j i^p j^q f(i, j),$$

gdzie $f(i, j)$ – wartość piksela o współrzędnych (i, j) obrazu f

Niezmienne momentowe Hu

$$N_1 = \eta_{20} + \eta_{02}$$

$$N_2 = (\eta_{20} - \eta_{02})^2 + 4\eta_{11}^2$$

$$N_3 = (\eta_{30} - 3\eta_{12})^2 + (\eta_{03} - 3\eta_{21})^2$$

$$N_4 = (\eta_{30} + \eta_{12})^2 + (\eta_{03} - \eta_{21})^2,$$

$$N_5 = (\eta_{30} - 3\eta_{12})(\eta_{30} + \eta_{12})[(\eta_{30} + \eta_{12})^2 - 3(\eta_{21} + \eta_{03})^2] + \\ + (\eta_{03} - 3\eta_{21})(\eta_{03} + \eta_{21})[(\eta_{03} + \eta_{21})^2 - 3(\eta_{12} + \eta_{30})^2]$$

$$N_6 = (\eta_{20} - \eta_{02})[(\eta_{30} + \eta_{12})^2 - (\eta_{21} + \eta_{03})^2] + 4\eta_{11}(\eta_{30} + \eta_{12})(\eta_{03} + \eta_{21})$$

$$N_7 = (3\eta_{21} - \eta_{03})(\eta_{30} + \eta_{12})[(\eta_{30} + \eta_{12})^2 - 3(\eta_{21} + \eta_{03})^2] + \\ + (\eta_{30} - 3\eta_{21})(\eta_{21} + \eta_{03})[(\eta_{03} + \eta_{21})^2 - 3(\eta_{30} + \eta_{12})^2]$$

$$\text{gdzie } \eta_{pq} = m_{pq} / m_{00}^\gamma, \quad \text{dla } N_1 \text{ i } N_2: \quad \gamma = 2,$$

$$\text{dla } N_3, N_4, \dots, N_7: \gamma = 3$$

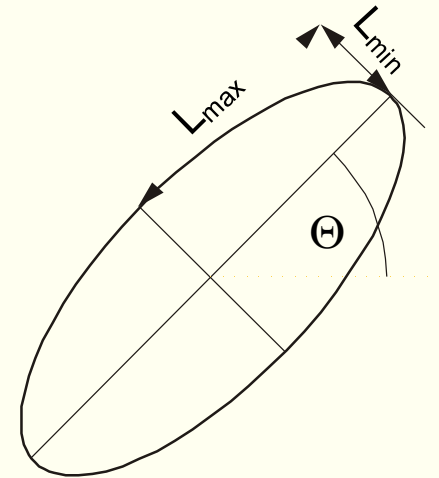
Cechy geometryczne

Orientacja

$$\Theta = \frac{1}{2} \tan^{-1} [2m_{11} / (m_{20} - m_{02})]$$

Ekscentryczność

$$\varepsilon = L_{\max} / L_{\min} = \left[(m_{20} - m_{02})^2 + 4m_{11}^2 \right]^{1/2} / A$$



Współczynnik zwartości

$$\gamma = \frac{P^2}{4\pi A}$$

A - powierzchnia, P - obwód

Omówienie przykładowych systemów wizyjnych

1. Rozpoznawanie języka migowego
2. Rozpoznawanie akcji człowieka w oparciu o obrazy historii ruchu

Rozpoznawanie Polskiego Języka Miganego

Rozpoznawszy gesty języka migowego na podstawie sekwencji obrazów można je tłumaczyć na język pisany lub mówiony (**demo**)

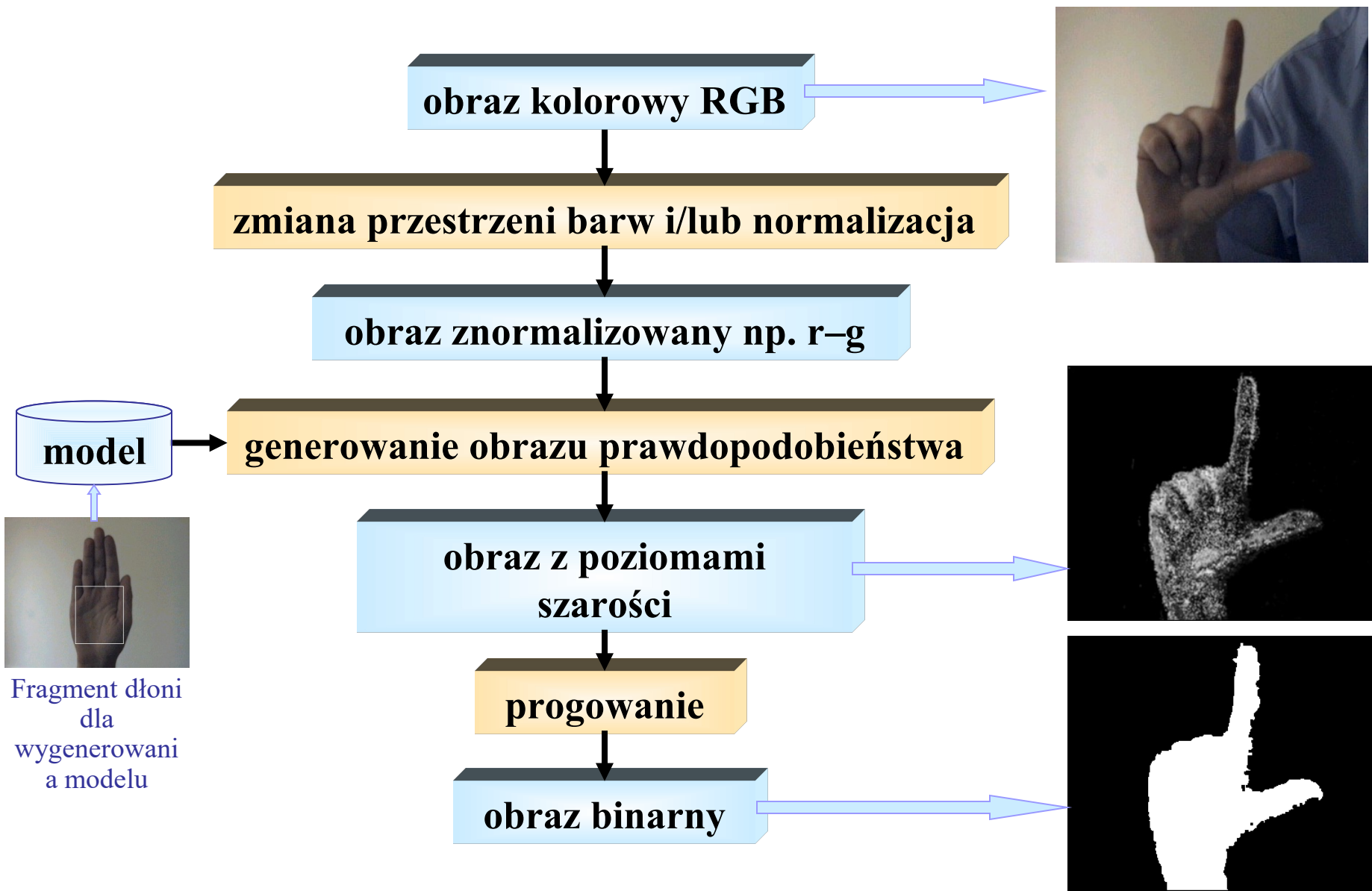


Stanowisko badawcze

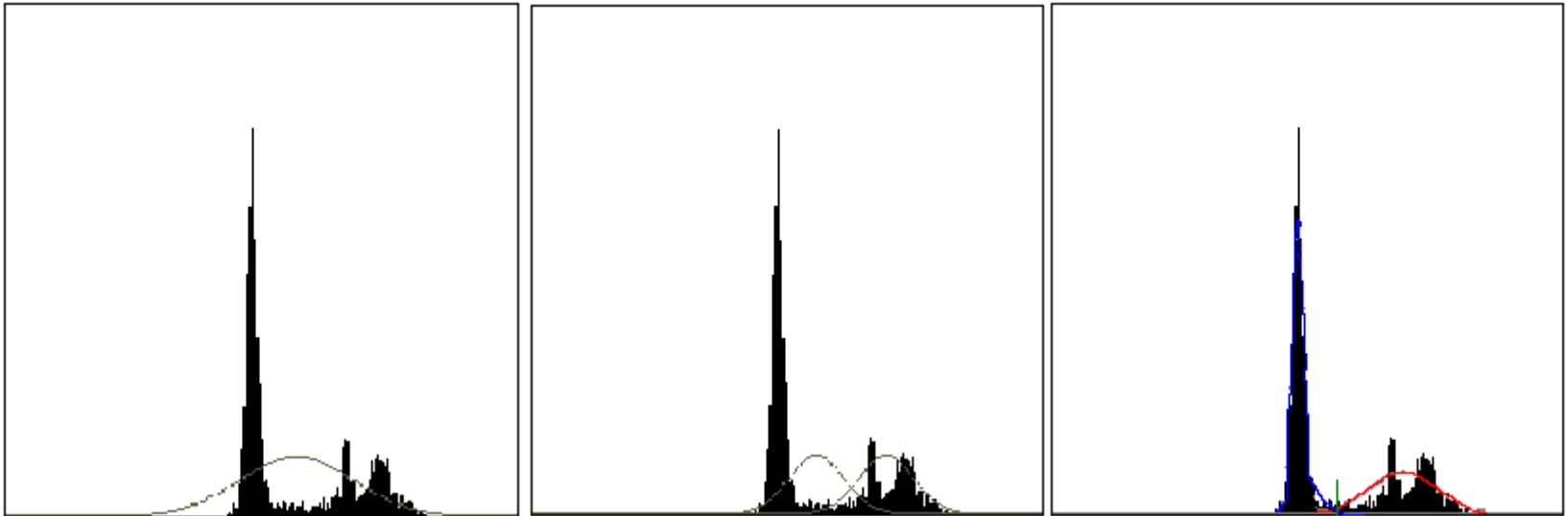
Problemy w rozpoznawaniu gestów

- Utrata informacji w wyniku projekcji 3D → 2D (np. niektóre palce mogą być niewidoczne przy pewnych konfiguracjach dłoni)
- Potrzeba rozróżnienia pomiędzy dłonią i twarzą, gdy dłoń zasłania twarz
- Konieczność wyodrębnienia poszczególnych gestów w sekwencji gestów
- Pozycja wykonującego gest względem kamery może być zmienna
- Niejednakowe wykonywanie gestów, nawet przez jedną osobę
- Duży nakład obliczeń

Wydzielanie dłoni i twarzy



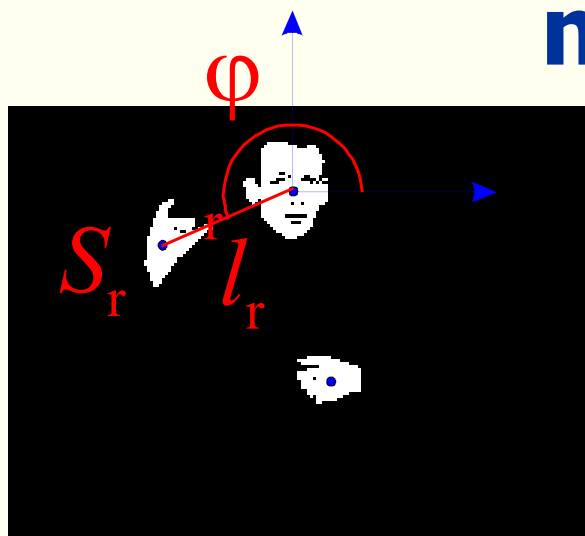
Progowanie obrazu prawdopodobieństwa



Na podstawie wartości r i g pikseli badanego obrazu otrzymuje się z modelu barwy skóry prawdopodobieństwa ich przynależności do skóry.

Otrzymany w ten sposób obraz prawdopodobieństwa poddawany jest binaryzacji z progiem obliczonym metodą Otsu, tj. interpretując jego histogram jako sumę dwóch rozkładów Gaussa. Wyznaczenie rozkładów aproksymujących odbywa się metodą iteracyjną. Próg binaryzacji wyznaczany jest jako punkt przecięcia się krzywych Gaussa.

Konstrukcja wektorów cech dla rozpoznawania polskiego języka migowego



Składowe wektora cech:

l_r – odległość środka ciężkości dłoni prawej od środka ciężkości twarzy

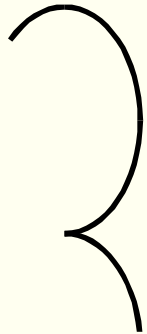
φ_r – orientacja odcinka łączącego środki ciężkości dłoni prawej i twarzy

S_r – pole powierzchni dłoni prawej

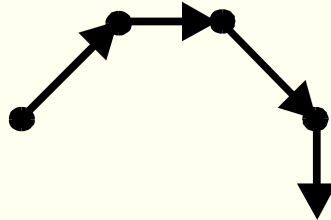
l_l , φ_l , S_l – analogicznie dla dłoni lewej

Sposób kodowania trajektorii ruchu

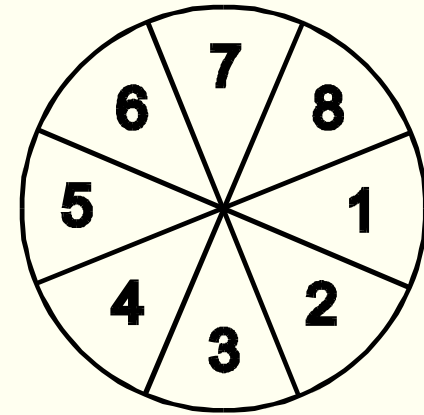
Trajektoria



Obserwacja



Model dyskretny

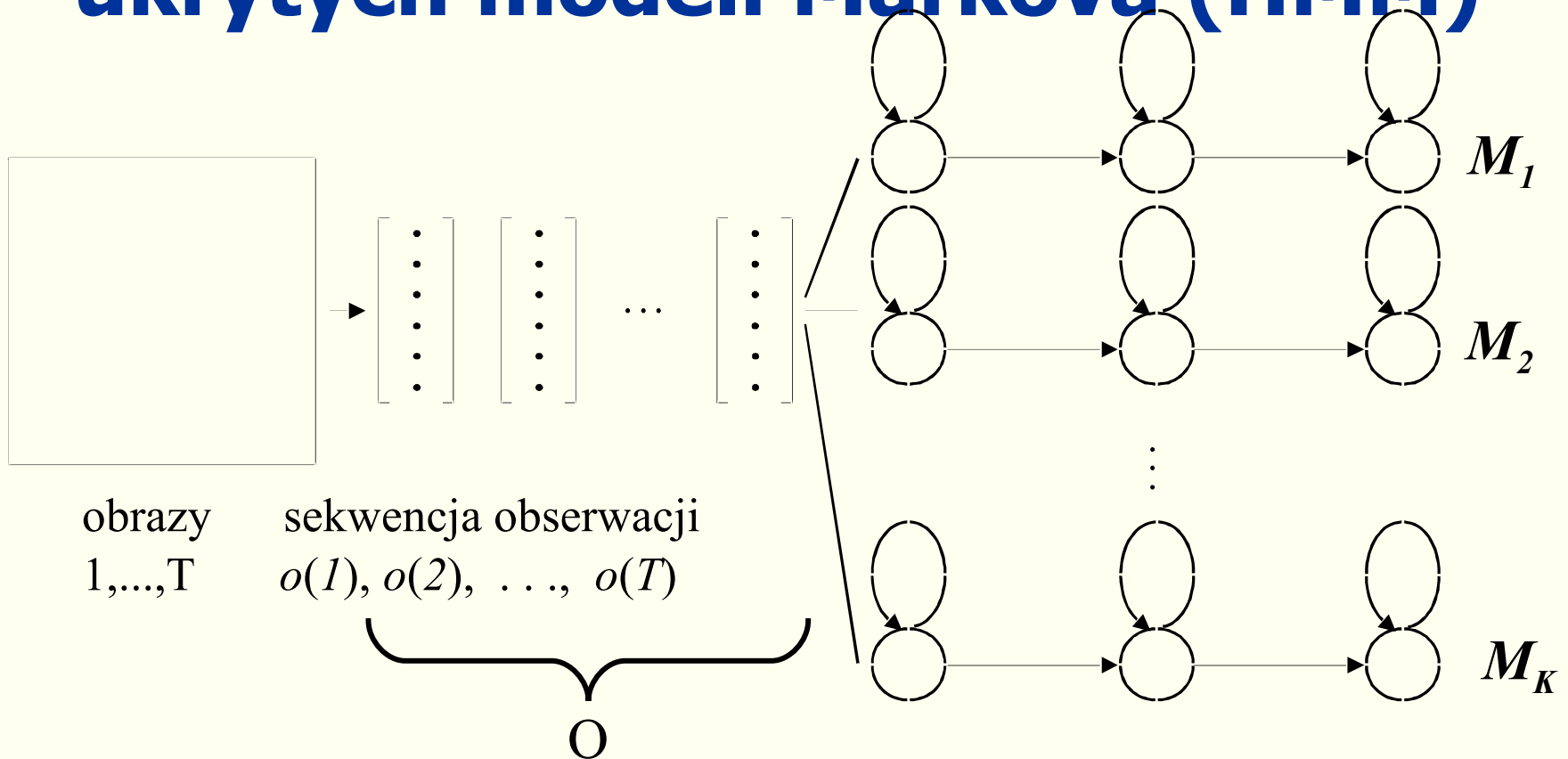


model ciągły

orientacja $\varphi(t)$

pozycja $l_x(t)$, $l_y(t)$

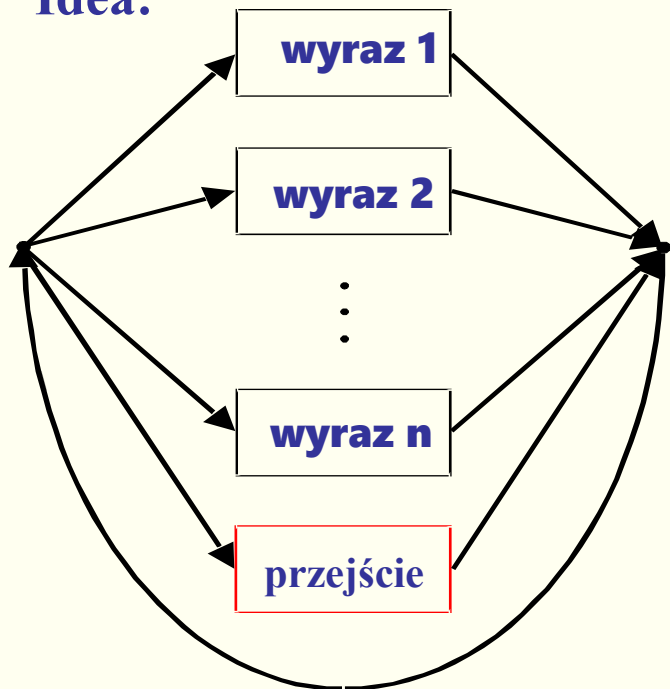
Rozpoznawanie metodą ukrytych modeli Markowa (HMM)



c^* - rozpoznana klasa, $c^* = \max_{1 \leq c \leq K} P(O|M_c)$

Rozpoznawanie zdań

Idea:

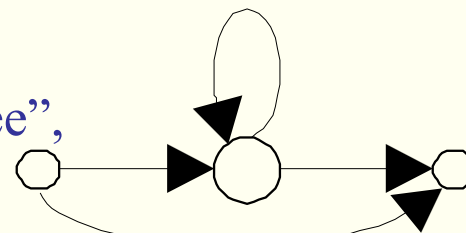


Modele wyrazów:

4 stanowe modele Bakisa, stany 1 i 4 nieemitujące,
funkcja gęstości prawdopodobieństwa obserwacji – suma 2 Gaussianów

Model „przejścia”:

3-stanowy model typu „Tee”,
stany 1 i 3 nieemitujące,
funkcja gęstości prawdopodobieństwa obserwacji – 1 Gaussian



Uczenie: wstępna reestymacja parametrów metodą Viterbiego, douczanie modeli metodą Bauma-Welcha

Model języka

Model bigram - prawdopodobieństwa następstwa wyrazów określane są poprzez analizę liczności występujących po sobie wyrazów w sekwencjach wykorzystanych do uczenia

$$P(i|j) = \frac{N(i,j)}{N(i)}$$

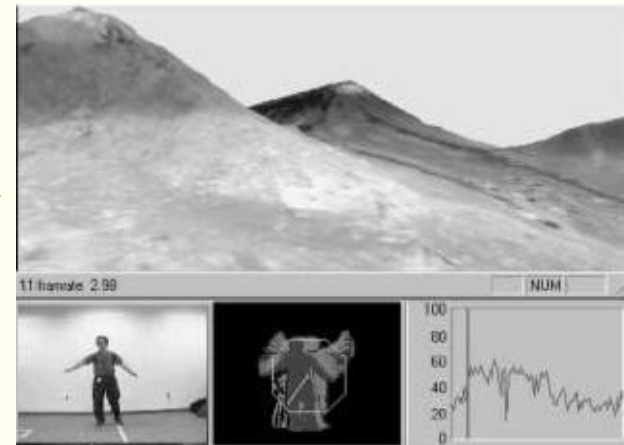
gdzie:

$N(i, j)$ – liczba wystąpień wyrazu i po wyrazie j

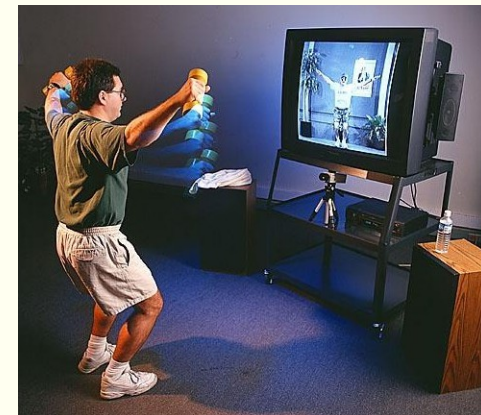
$N(i)$ – liczba wystąpień wyrazu i

Systemy sterowane za pomocą gestów ciała

- Sterowanie syntezaorem dźwięku / symulatorem lotu za pomocą gestów ciała

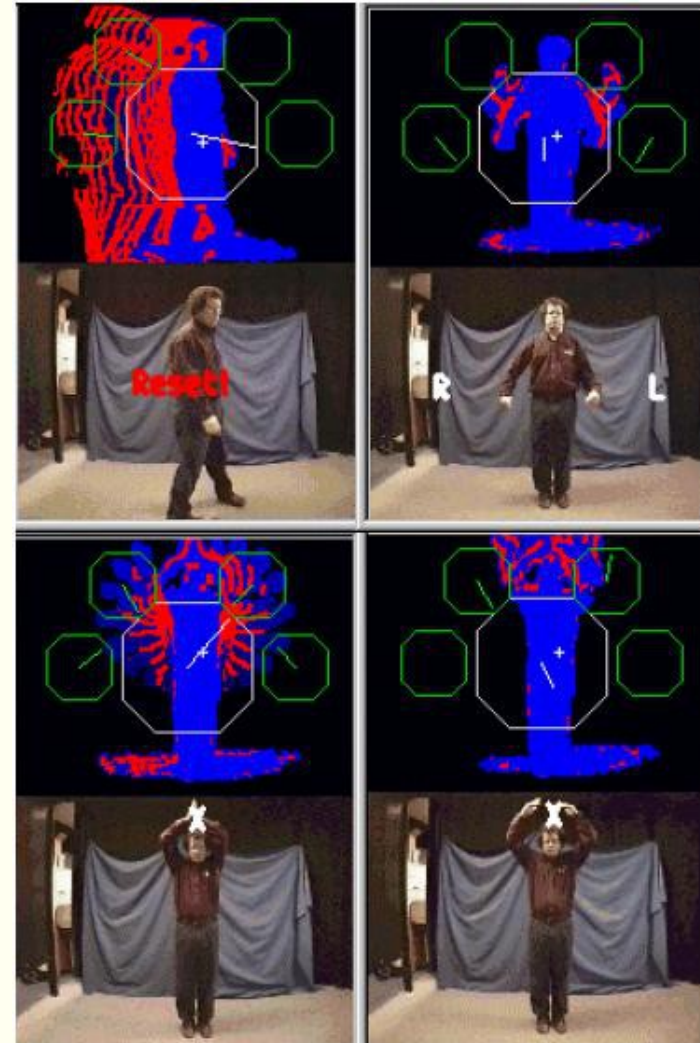


- Reprezentacja i rozpoznawanie akcji wykonywanej przez człowieka – wirtualny nauczyciel aerobiku



Systemy sterowane za pomocą gestów ciała c.d

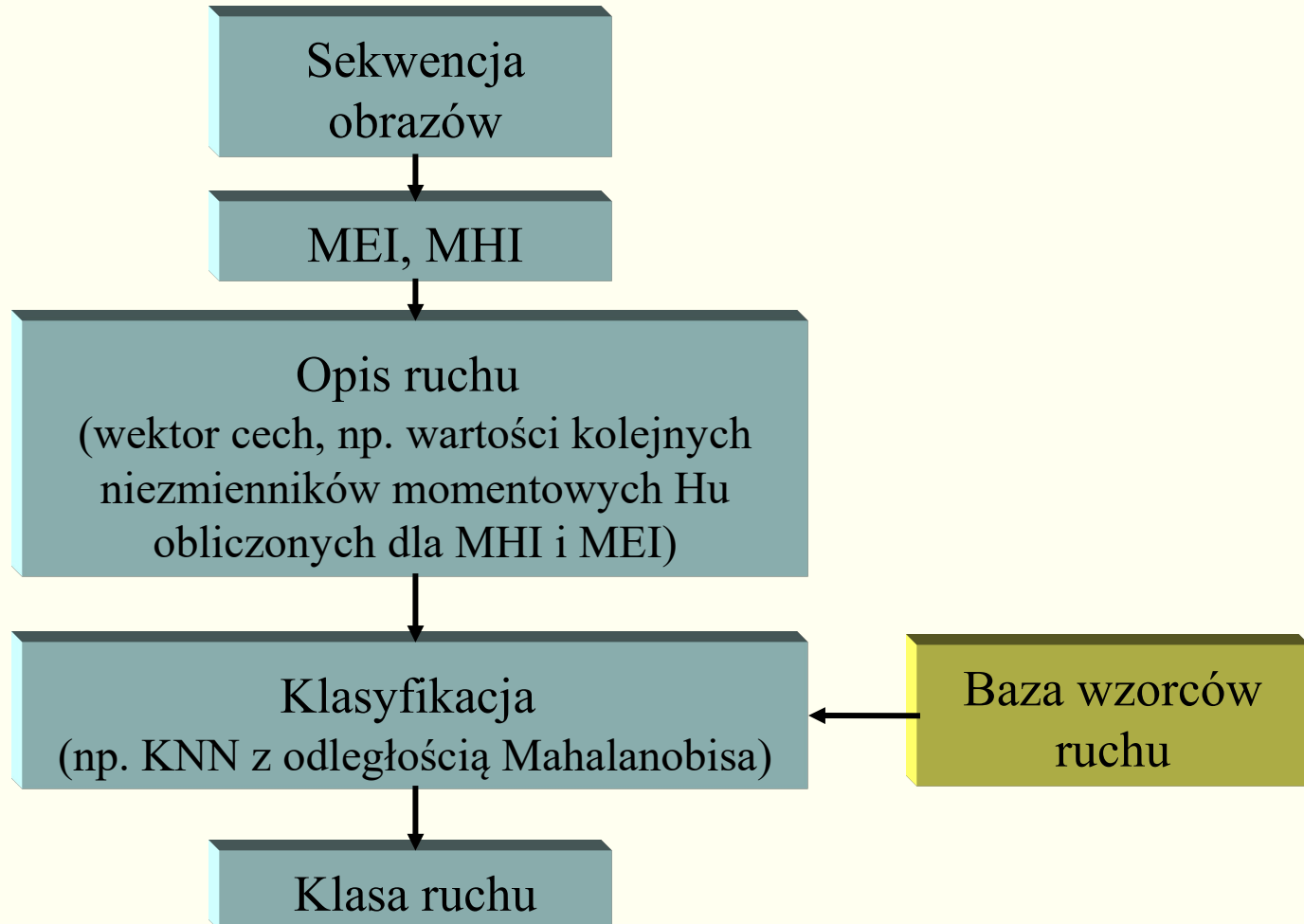
- Dyrygowanie muzyką z użyciem gradientu MHI
- KidsRoom – interaktywny pokój zabaw dla dzieci



Sposób rozwiązania

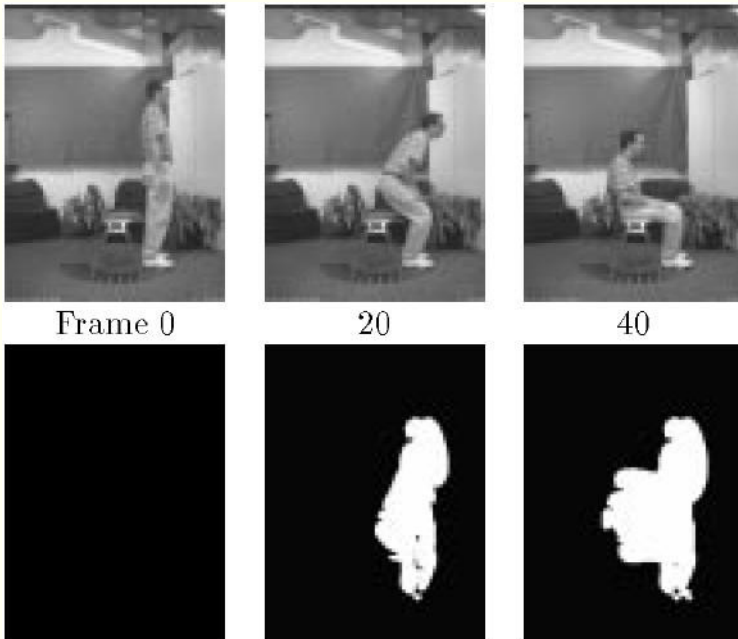
- Utworzenie wzorców ruchu w postaci obrazów historii ruchu (MHI) i energii ruchu (MEI)
- Obrazy te opisywane są z użyciem 14 niezmienników momentowych H_u (7 dla MEI i 7 dla MHI)
- Dla każdego kąta obserwacji każdej z akcji wyznaczany jest model statystyczny na podstawie wzorców uczących w postaci:
 - Wartości średniej dla MEI i MHI
 - Macierzy kowariancji dla MEI i MHI
- Miara podobieństwa badanej akcji do modelu wzorca akcji wyznaczana jest jako odległość Mahalanobisa między opisem tej akcji a każdym modelem statystycznym
- Akcja zaliczana jest do klasy, dla modelu której otrzymano najmniejszą wartość odległości

Schemat rozpoznawania akcji

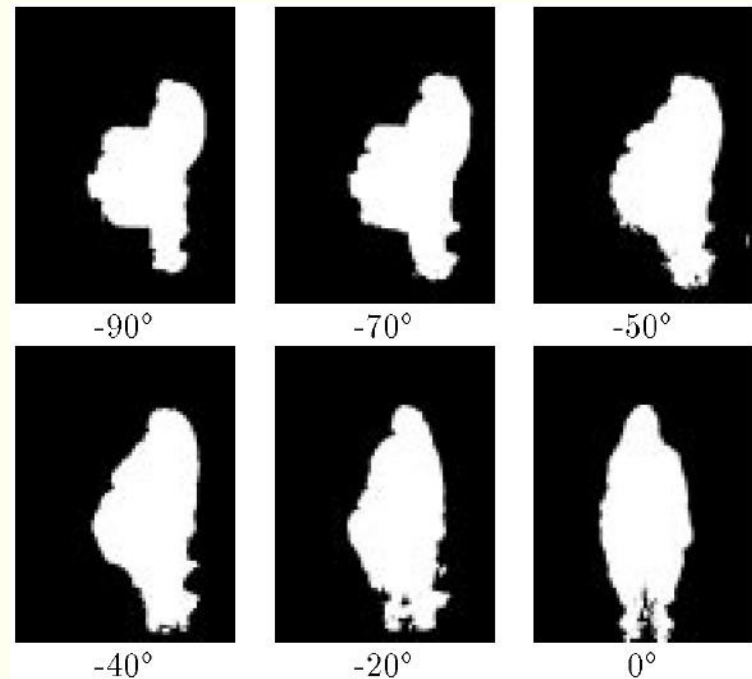


Przykładowe obrazy energii ruchu dla siadania

Wybrane klatki z sekwencji ruchu i odpowiadające im MEI

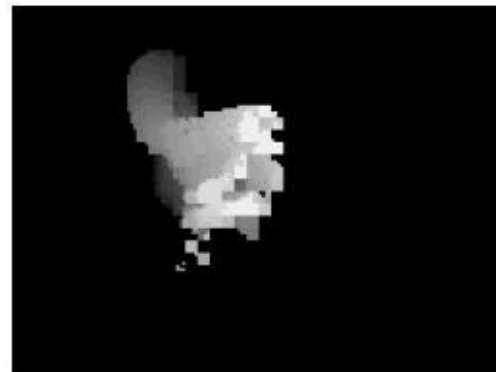


MEI dla akcji obserwowanej z różnych stron

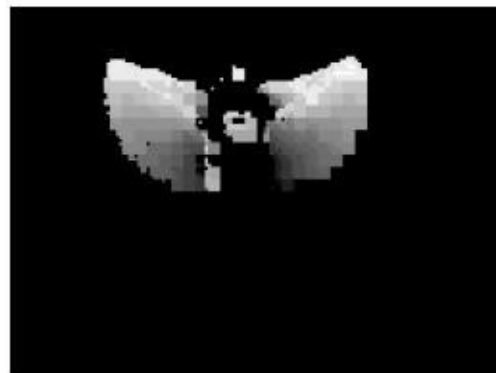


Przykładowe obrazy historii ruchu

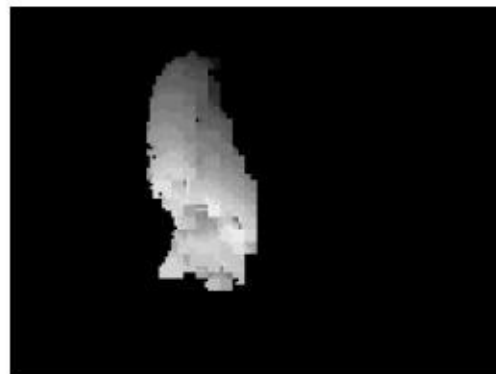
siadanie



machanie rękami

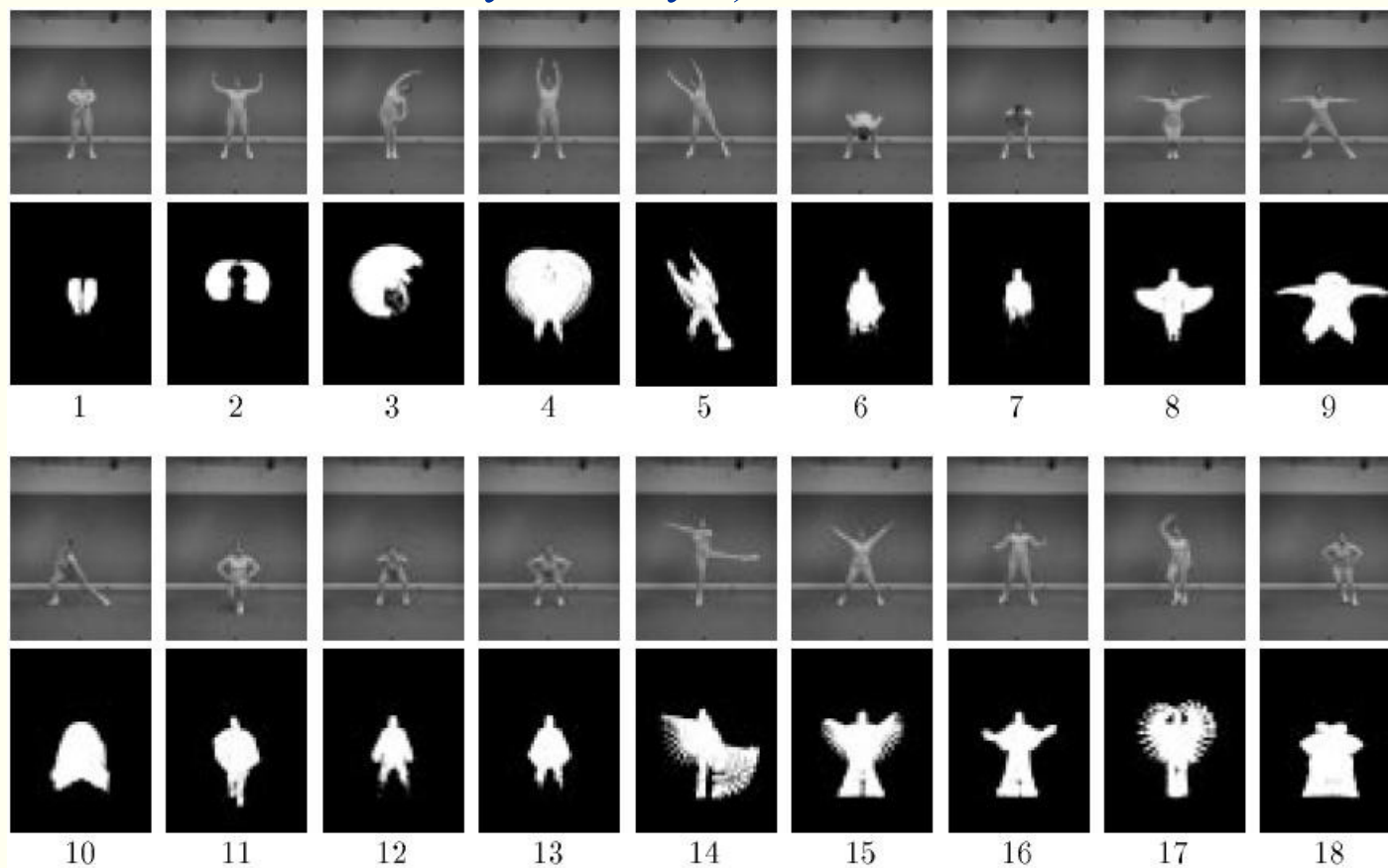


kucanie



Testy na przykładzie ćwiczeń aerobiku

- Zestaw 18 akcji, obserwowanych z 7 punktów (-90° do 90° z krokiem co 30° w kierunku horyzontalnym)



Wyniki testów dla przykładu z ćwiczeniami aerobiku

- Dla jednego punktu obserwacji (kąt -30°) wzorca czasowego badanej akcji:
 - ◆ 12 z 18 akcji zostało rozpoznanych poprawnie
 - ◆ Błędnie rozpoznawanym akcjom odpowiadały zbliżone pod względem wyglądu wzorce czasowe
- Dla dwóch punktów obserwacji (kąt -30° i 60°) wzorca czasowego badanej akcji:
 - ◆ Wartość podobieństwa do modelu wyznaczana była jako najmniejsza suma odległości Mahalanobisa otrzymanych dla 2 modeli akcji, którym odpowiadała różnica w kątach obserwacji równa 90°
 - ◆ 15 z 18 akcji zostało rozpoznanych poprawnie
- System pracował z częstotliwością 9 Hz na stacji Silicon Graphics 200MHz Indy z rozdzielczością obrazu 160×120

Cechy systemów wykorzystujących wzorce ruchu

- Proste akcje (typu siadanie, pompki) mogą być rozpoznane nawet na obrazach rozmytych i o małej rozdzielczości
- Nie ma potrzeby wykorzystywania informacji o trójwymiarowej strukturze sceny



Wyznaczanie kierunku ruchu

Oparte jest na czasowych obrazach MHI $tMHI$ zdefiniowanych następująco:

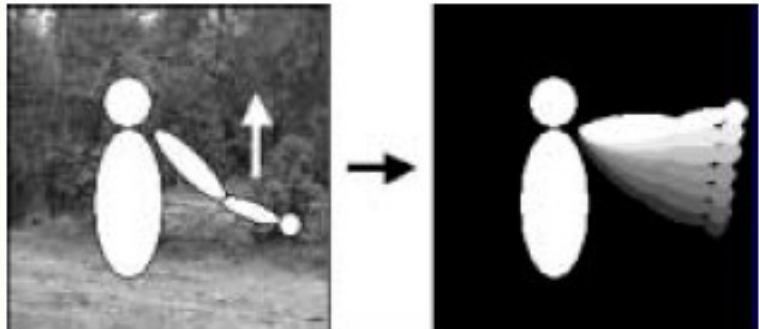
$$tMHI_{\delta}(x, y) = \begin{cases} \tau & \text{jesli piksel } (x, y) \text{ nalezy do sylwetki} \\ 0 & \text{jesli } tMHI_{\delta}(x, y) < (\tau - \delta) \end{cases}$$

gdzie τ - bieżący stempel czasowy (liczba w formacie sec.msec)

δ - maksymalny czas skojarzony z wzorcem (typowe wartości to kilka sekund)

Cechy $tMHI$:

- ✓ Reprezentacje wzorców są niezależne od szybkości systemu i częstotliwości akwizycji
→ ten sam ruch spowoduje utworzenie podobnych obszarów MHI nawet dla różnych częstotliwości akwizycji



Gradienty ruchu

Obliczanie gradientów:

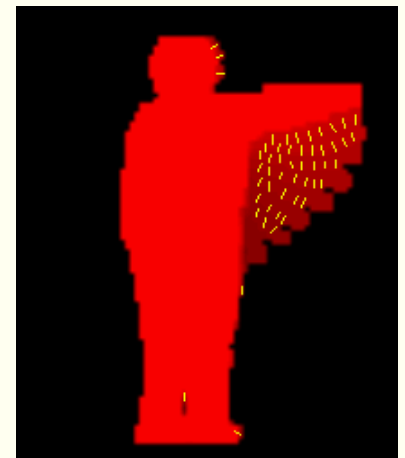
1. Wyznaczyć pochodne przestrzenne $F_x(x, y)$ i $F_y(x, y)$ w kierunkach X i Y , odpowiednio, wykorzystując operatory Sobela
2. Wyznaczyć orientację gradientu ze wzoru:
$$\phi(x, y) = \arctan (F_x(x, y) / F_y(x, y))$$

UWAGA:

Gradienty ruchu wyznaczane są tylko dla pikseli znajdujących się wewnątrz obiektu tMHI, dla których kontrast w lokalnym sąsiedztwie nie jest zbyt mały (tak jak we wnętrzu sylwetki) ani zbyt duży

Cechami ruchu mogą być:

- Histogram orientacji krawędzi,
- Globalny kierunek ruchu.



Wyznaczanie globalnego kierunku ruchu

Zależność pozwalająca obliczyć globalny kierunek ruchu:

$$\bar{\phi} = \phi_{ref} + \frac{\sum_{x,y} angDiff(\phi(x,y), \phi_{ref} \times norm(\tau, \delta, tMHI_{\delta}(x,y)))}{\sum_{x,y} norm(\tau, \delta, tMHI_{\delta}(x,y))}$$

gdzie:

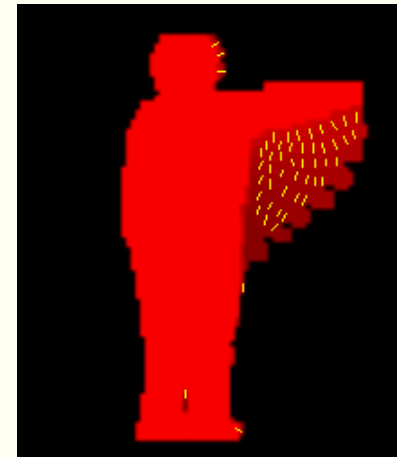
$\bar{\phi}$ – globalny kierunek ruchu

ϕ_{ref} – kierunek odniesienia (wartość odpowiadająca najwyższemu wierzchołkowi w histogramie orientacji gradientów)

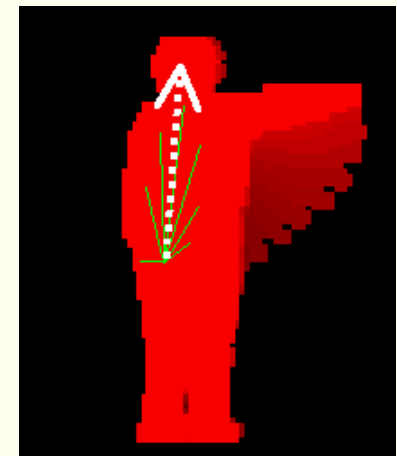
$\phi(x, y)$ – mapa gradientów ruchu

$norm(\tau, \delta, tMHI_{\delta}(x, y))$ – liniowo znormalizowany do zakresu [0, 1] obraz tMHI,

$angDiff(\phi(x, y), \phi_{ref})$ – odchylenie kierunku gradientu dla punktu (x, y) od kierunku odniesienia



$\phi(x, y)$



$\bar{\phi}$

Demonstracja działania metody wyznaczania globalnego kierunku ruchu

