

# Cyfrowe Przetwarzanie Obrazów

## Digital Image Processing

### Wykład 2



Aurelio A. Heckert

Copyright (c) 2008 Cezary Bołdak

Permission is granted to copy, distribute and/or modify this document under the terms of the GNU Free Documentation License, Version 1.2 or any later version published by the Free Software Foundation; with no Invariant Sections, no Front-Cover Texts, and no Back-Cover Texts. A copy of the license is included in the section entitled "GNU Free Documentation License"



## Operacje bezkontekstowe na obrazie

- przekształcają obraz/obrazy na inne obrazy

$$I \Rightarrow J \quad I \times I \times \dots \times I \Rightarrow J$$

- operacje na pojedynczych pikselach obrazu

$$I(x,y) \Rightarrow J(x,y)$$

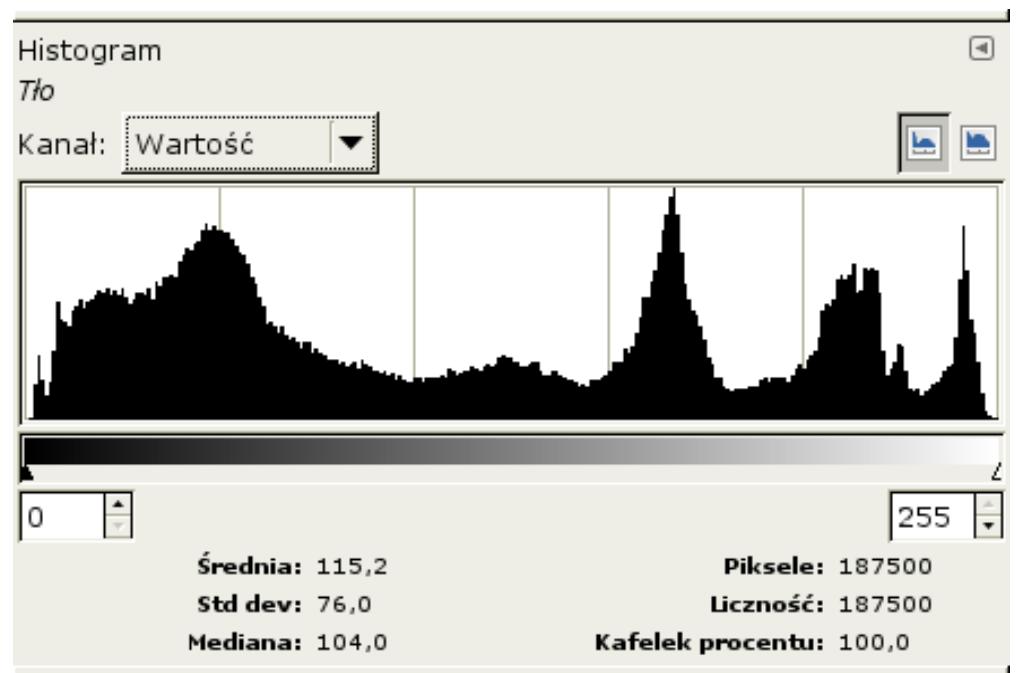
- otoczenie przetwarzanego piksela nie ma wpływu na operację
- wszystkie piksele o jednakowej intensywności są traktowane identycznie
- służą różnym celom:
  - poprawa jakości obrazu (kontrastu, jasności)
  - wydobycie, uwypuklenie pewnych cech
  - zmiana histogramu
  - zmiana kolorów
  - arytmetyka na obrazach (dodawanie, odejmowanie)
  - korekcja Gamma



## Histogram obrazu cyfrowego

- wykres obrazujący, ile pikseli o każdej intensywności jest w obrazie
- w osi OX odłożone są wszystkie możliwe intensywności w obrazie
- pozycja Y punktu (wysokość słupka) jest proporcjonalna do liczby pikseli obrazu o intensywności X

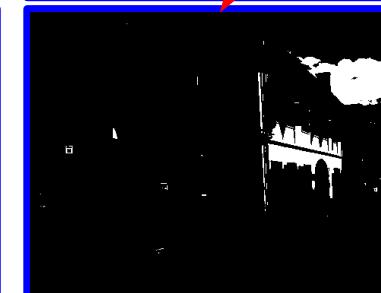
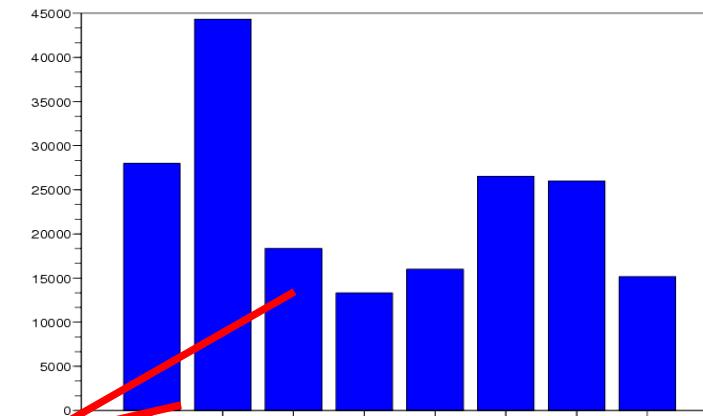
$$H(i) = \text{card}\{(x, y) : \text{Image}(x, y) = i\}$$





# Histogram obrazu cyfrowego

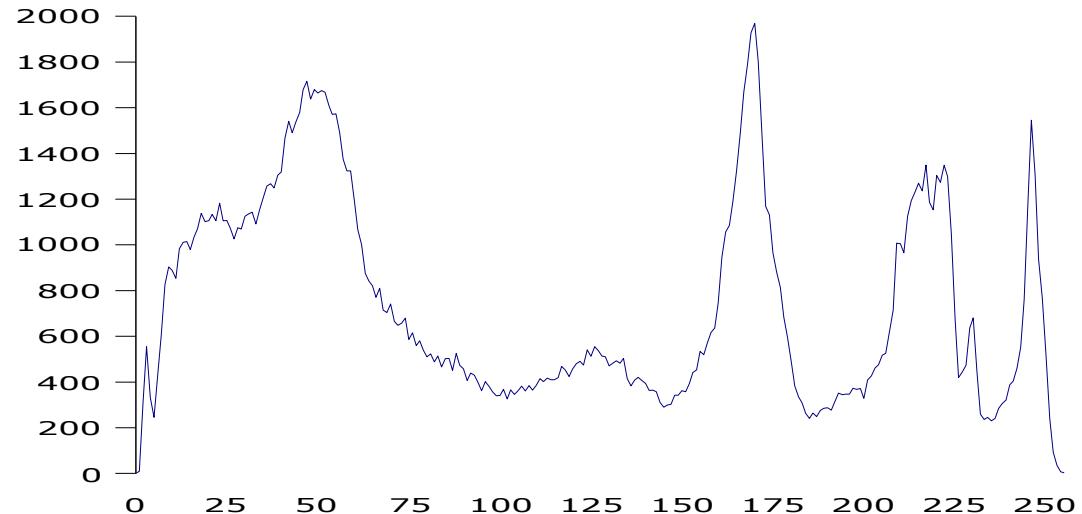
obraz z 8 poziomami intensywności





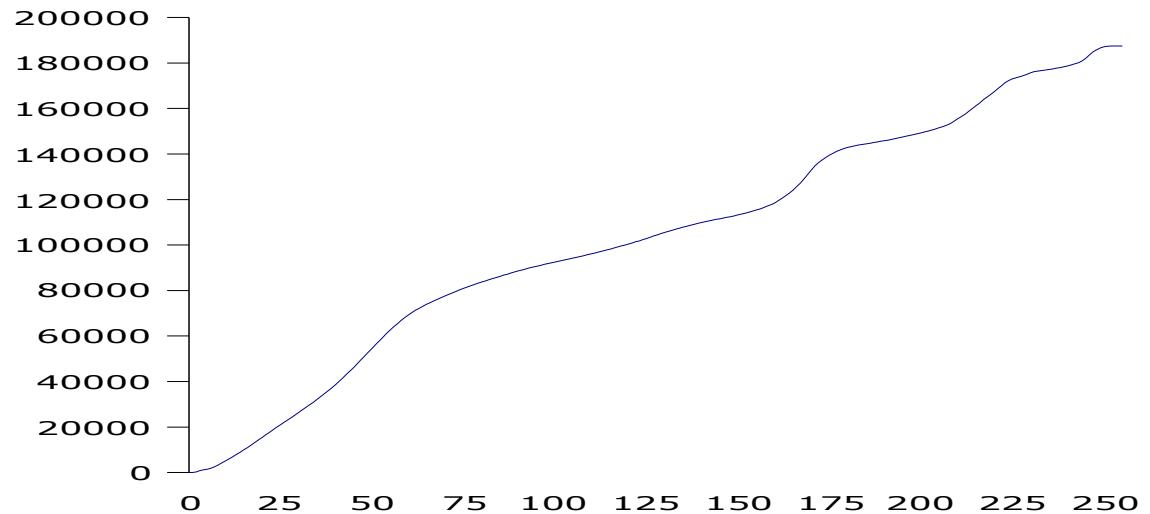
## Obliczanie histogramu (scilab)

```
image=gray_imread('W2.jpg');
[h, w]=size(image);
hist=zeros(256,1);
for y=1:h
    for x=1:w
        lev=int(image(y,x)*255+0.5);
        hist(lev+1)=hist(lev+1)+1;
    end
end
bar(hist);
xbasc;
plot(hist);
```



### histogram skumulowany

```
histcum=hist;
for i=2:size(histcum,1)
    histcum(i)=histcum(i) +
                histcum(i-1);
end
plot(histcum);
```

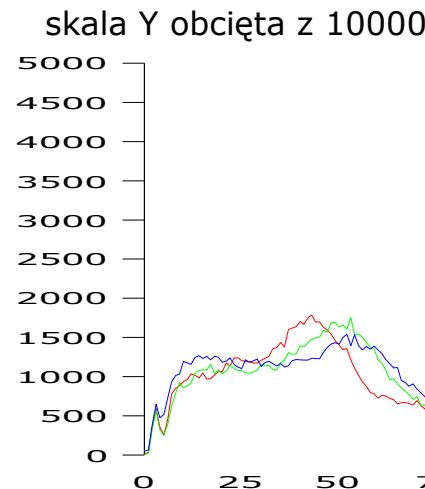




## Histogram obrazów kolorowych

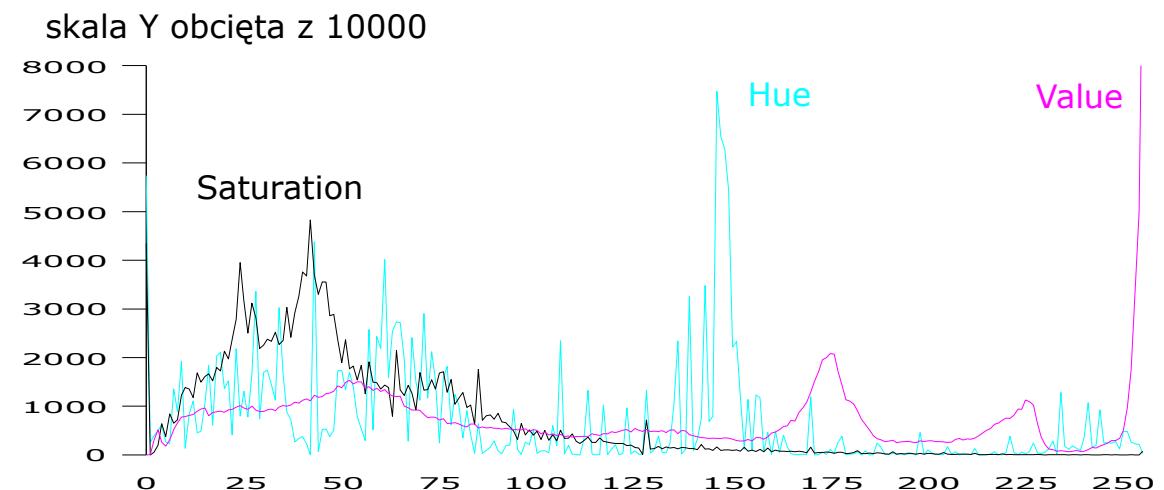
każdy z kanałów RGB oddzielnie

```
imRGB=imread('W2-RGB.jpg');
[h, w]=size(imRGB);
hRGB=zeros(256,3);
for y=1:h
    for x=1:w
        for k=1:3
            l=int(imRGB(y,x,k)*255+0.5);
            hRGB(l+1,k)=hRGB(l+1,k)+1;
        end
    end
end
plot(hRGB);
```



w innym modelu kolorów (HSV)

```
imHSV=rgb2HSV(imRGB);
hHSV=zeros(256,3);
for y=1:h
    for x=1:w
        for k=1:3
            l=int(imHSV(y,x,k)*255+0.5);
            if l<0 then; l=0; end
            hHSV(l+1,k)=hHSV(l+1,k)+1;
        end
    end
end
plot(hHSV);
```





## Zmiana wartości piksela Look-up Table (LUT)

- definiuje operacje bezkontekstowe
- tablica określająca wartość piksela po operacji

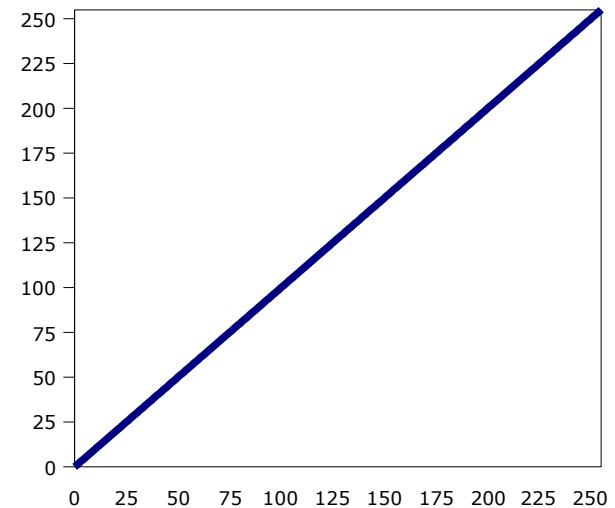
$$I(x,y) \Rightarrow \text{LUT}(I(x,y))$$

przykładowa LUT

piksel	f(piksel)
0	13
1	34
2	234
...	...
254	11
255	255

```
function res=LUTgray(image, LUT)
[h, w]=size(image);
res=zeros(h,w);
for y=1:h
    for x=1:w
        res(y,x)=LUT(int(image(y,x)*255+0.5)+1)/255;
    end
end
endfunction
```

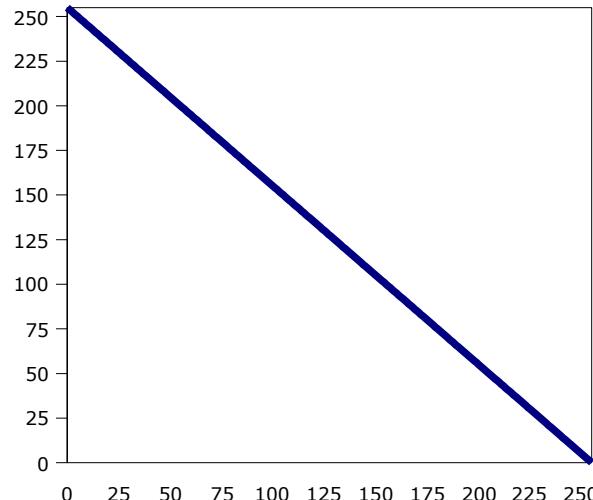
LUT operacji tożsamościowej



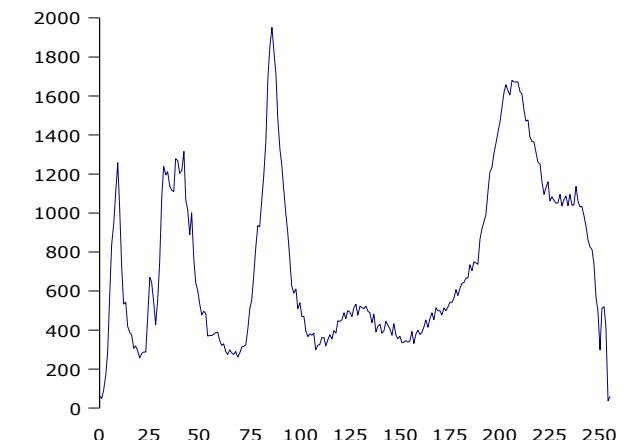
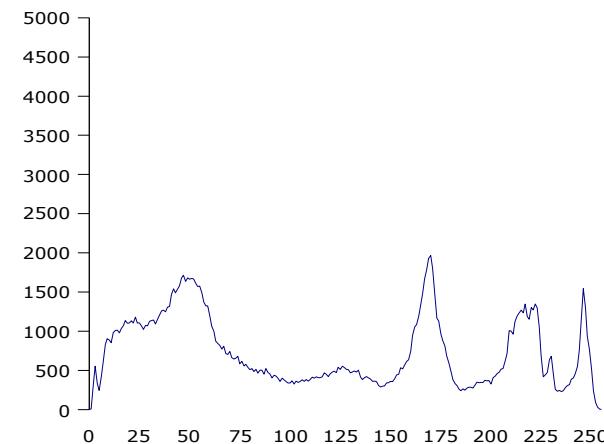


## Negatyw obrazu

```
for i=0:255  
    LUT(i+1)=255-i;  
end  
image2=LUTgray(image,LUT) ;  
imshow(image2) ;
```



skala Y obcięta

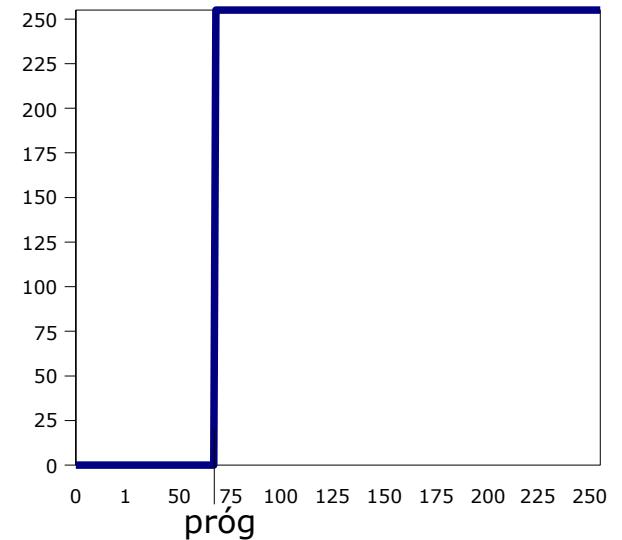




## Progowanie

$$Image(x, y) \Rightarrow \begin{cases} 0 & \text{jeżeli } Image(x, y) < próg \\ 255 & \text{jeżeli } Image(x, y) \geq próg \end{cases}$$

```
LUT=zeros(256,1);  
LUT(próg:256)=1;  
image2=LUTgray(image,LUT);  
imshow(image2);
```



próg=32



próg=127



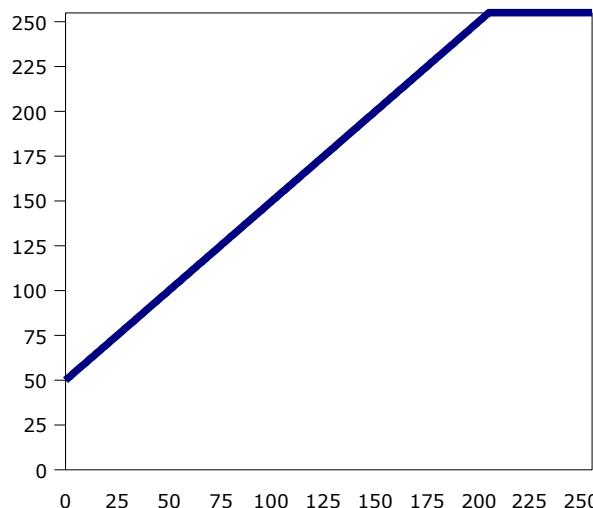
próg=184



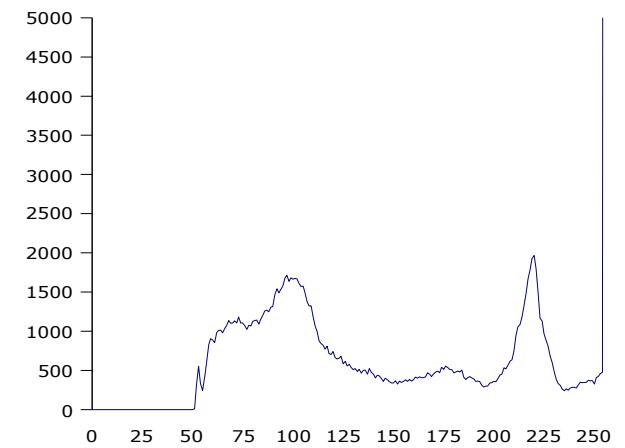
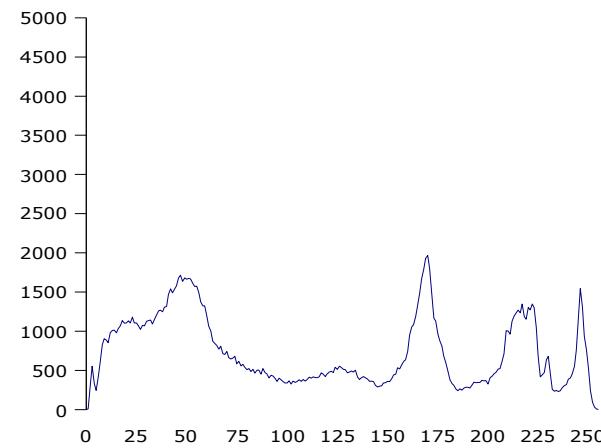


## Zwiększenie jasności

```
//zwiększenie jasności o 50
for i=0:255
    LUT(i+1)=min(255,i+50);
end
image2=LUTgray(image,LUT);
imshow(image2);
```



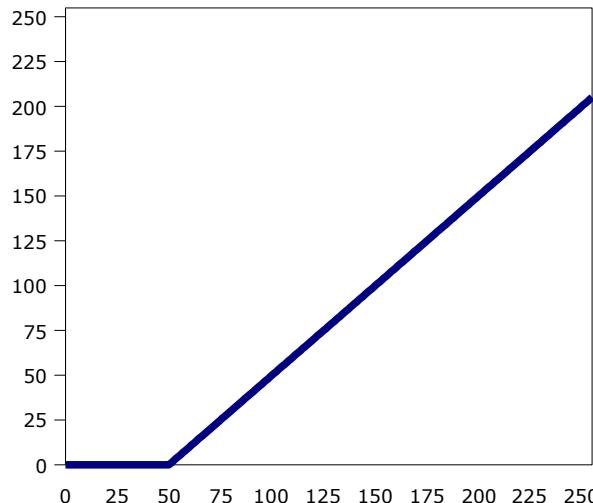
skala Y obcięta



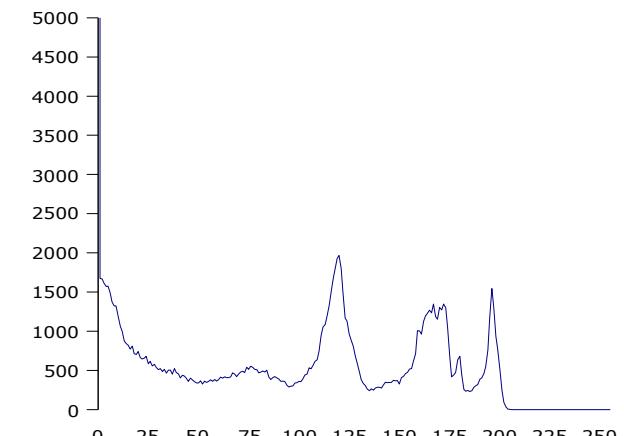
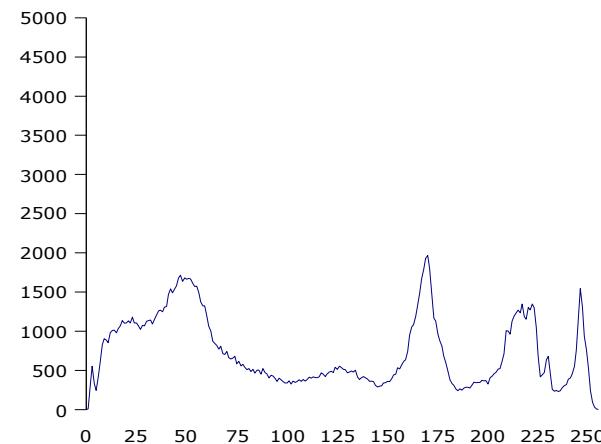


## Zmniejszenie jasności

```
//zwiększenie jasności o 50  
for i=0:255  
    LUT(i+1)=max(0,i-50);  
end  
image2=LUTgray(image,LUT);  
imshow(image2);
```



skala Y obcięta



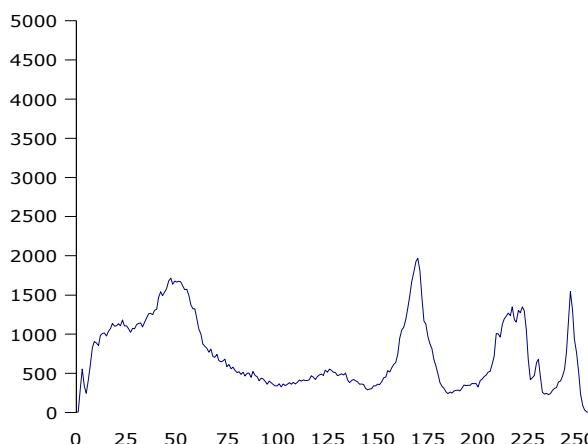


## Strata informacji przy zmianie jasności

jasność -= 100

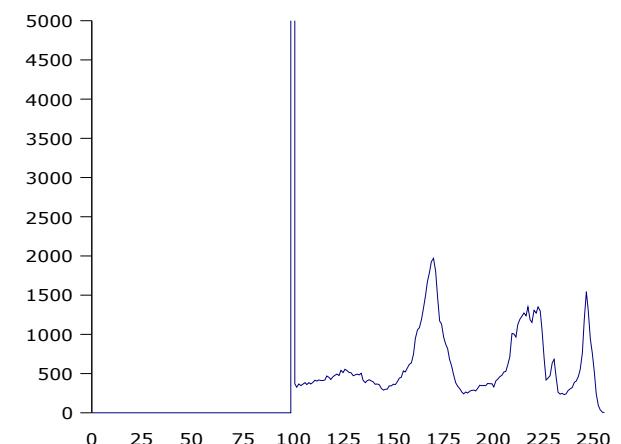


jasność += 100



- obcięte zostały wszystkie ciemne intensywności
- poziomy >100 nie uległy zmianie

skala Y obcięta



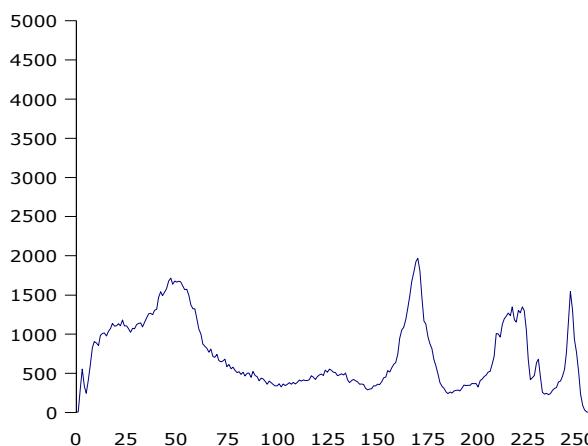


## Strata informacji przy zmianie jasności

jasność += 100

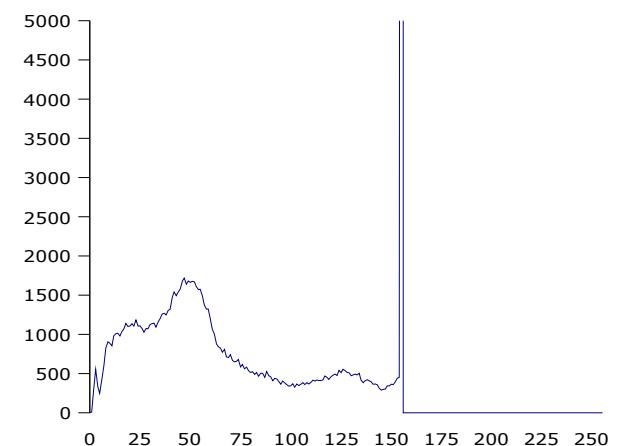


jasność -= 100



- obcięte zostały wszystkie jasne intensywności
- poziomy <255-100 nie uległy zmianie

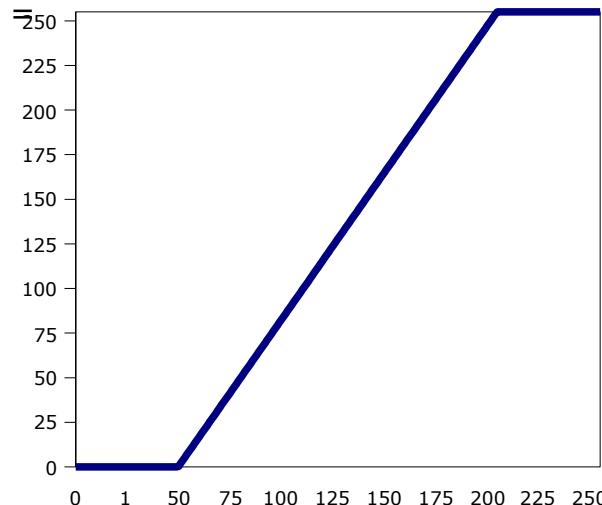
skala Y obcięta



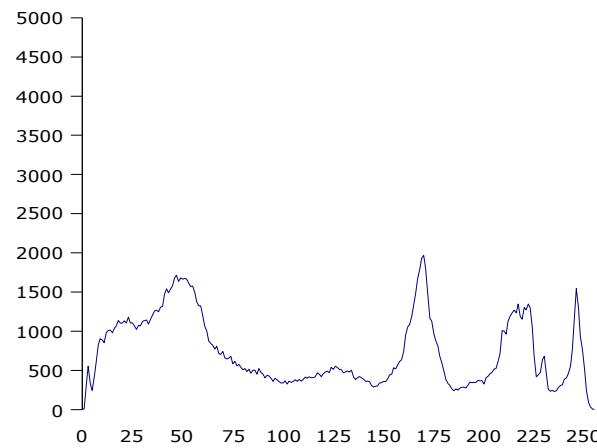


## Zwiększenie kontrastu

```
//zwiększenie kontrastu o 50
for i=0:255
    LUT(i+1)=int(127+(i-127)*255/155+0.5);
    LUT(i+1)=min(255,max(0,LUT(i+1)));
end
image2=LUTgray(image,LUT);
imshow(image2);
```



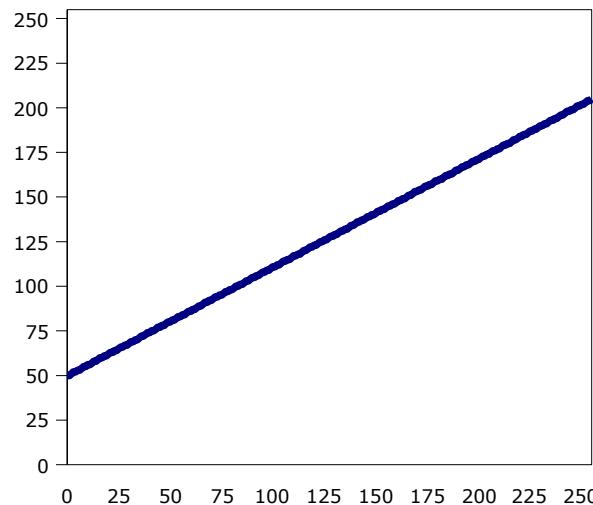
skala Y obcięta



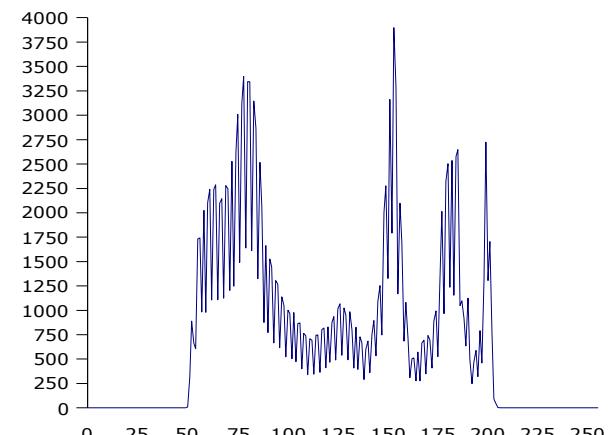
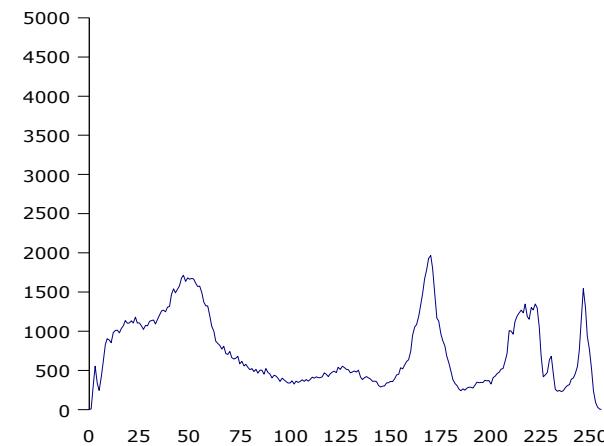


## Zmniejszenie kontrastu

```
//zwiększenie kontrasty o 50
for i=0:255
    LUT(i+1)=int(127+(i-127)*155/255+0.5);
    LUT(i+1)=min(255,max(0,LUT(i+1)));
end
image2=LUTgray(image,LUT);
imshow(image2);
```



skala Y obcięta



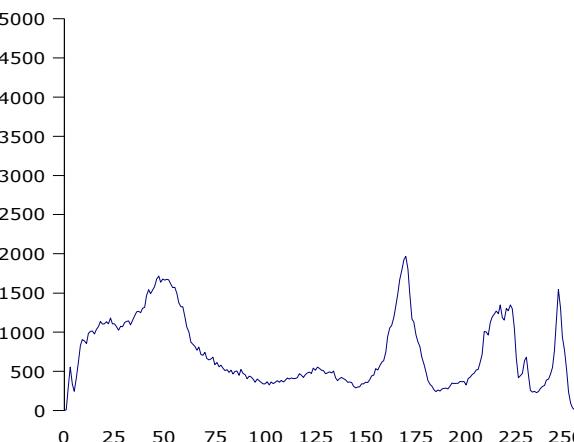


## Strata informacji przy zmianie kontrastu

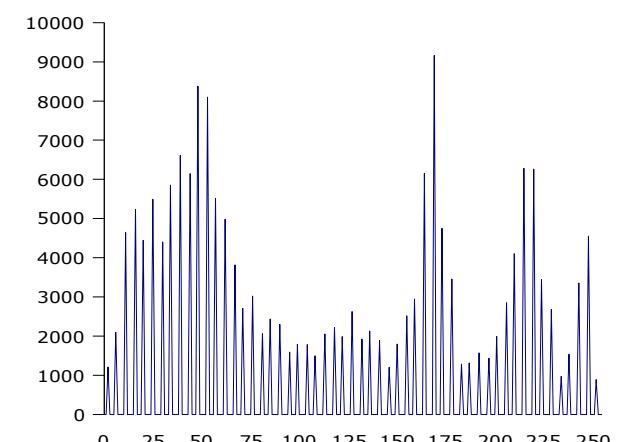
kontrast -= 100



kontrast += 100



- pozornie obraz wyjściowy przypomina wejściowy
- ale patrząc na histogramy można zauważyć, że została zmniejszona głębia kolorów



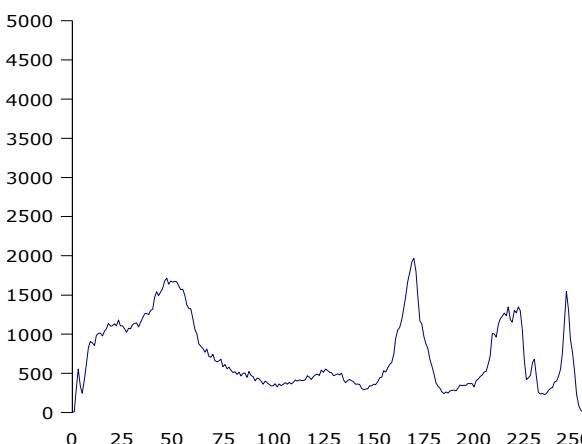


## Strata informacji przy zmianie kontrastu

kontrast  $\approx 75$

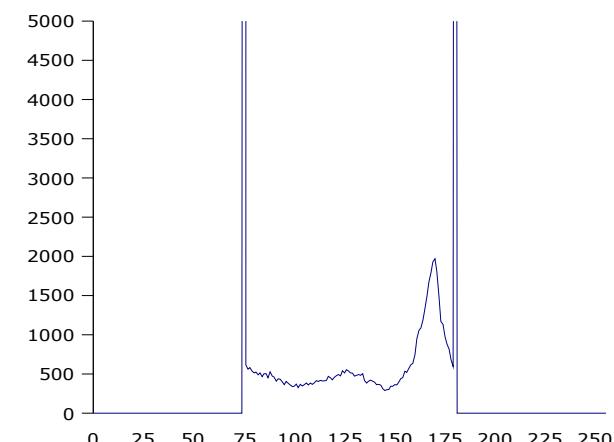


kontrast  $\approx -75$



- obraz wyjściowy ma „obcięte” jasne i ciemne poziomy
- patrząc na histogramy można zauważać, że średnie poziomy pozostały bez zmian

skala Y obcięta





## Korekcja gamma

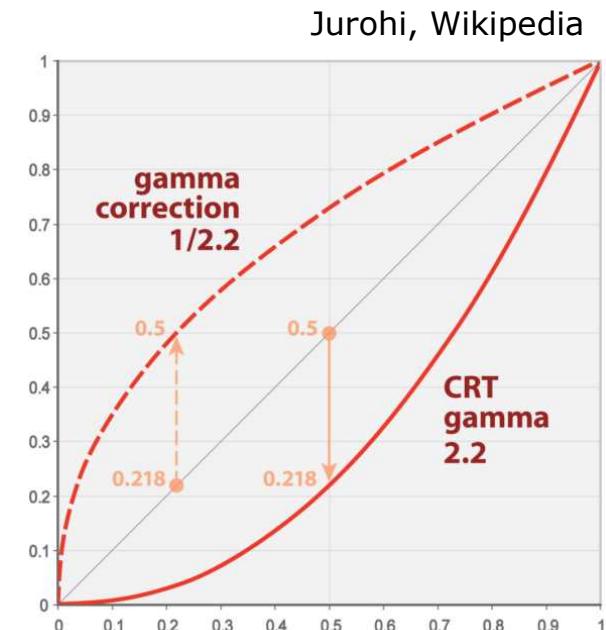
- zmiana intensywności wg nieliniowej funkcji

$$Image(x,y) \Rightarrow 255 \cdot \left( \frac{Image(x,y)}{255} \right)^{\gamma}$$

- operacja bezkontekstowa, realizowana jako LUT

```
for i=0:255
    LUT(i+1)=int(255*(i/255)^gamma+0.5);
end
image2=LUTgray(image,LUT);
imshow(image2);
```

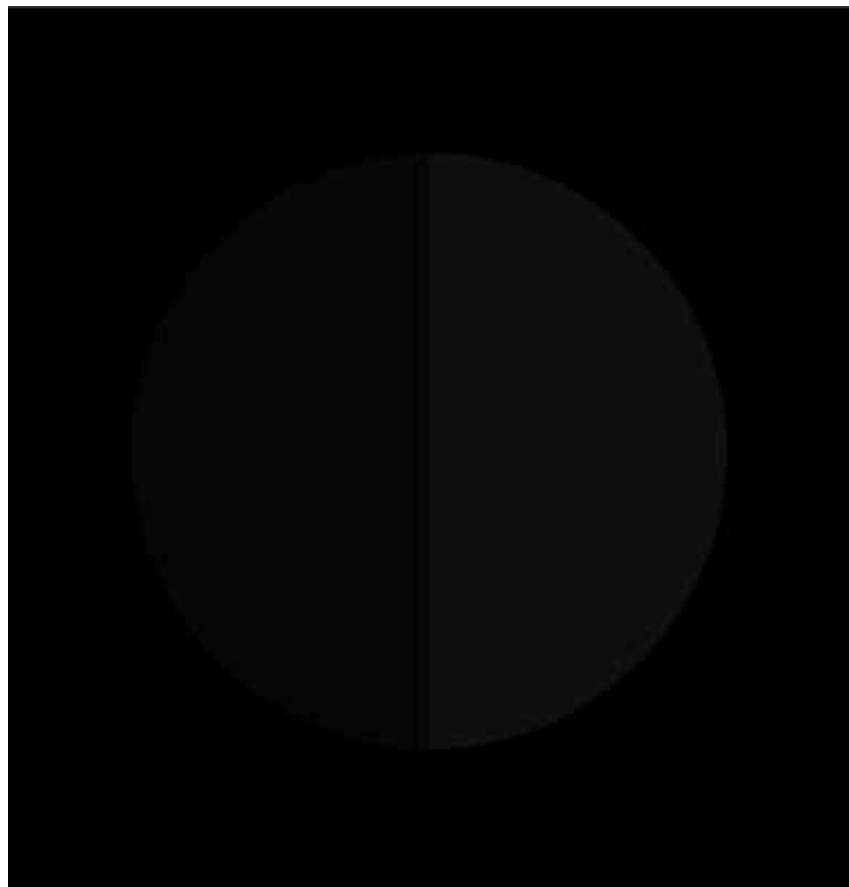
- bardzo częsta operacja dostosowywania kolorów do nieliniowej charakterystyki monitorów CRT
  - gamma monitorów CRT ok. 2.2
  - korekcja gamma 1/2.2



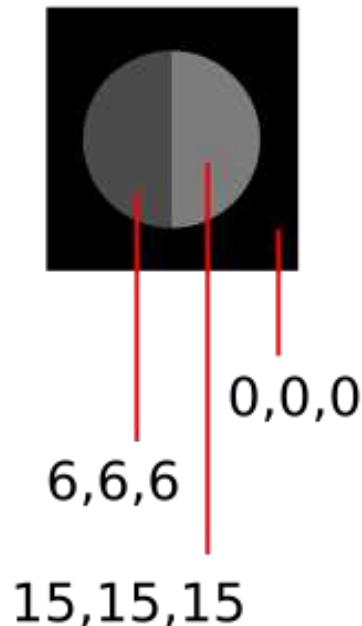


## Test ustawienia korekcji gamma monitora

Janke, Wikipedia



RGB values apply to  
large image; thumbnail  
is exaggerated for clarity

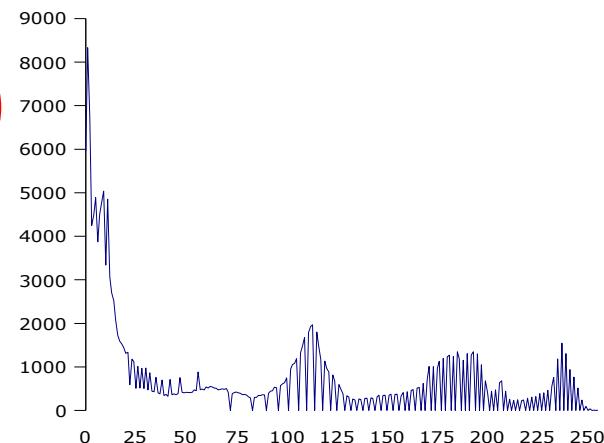
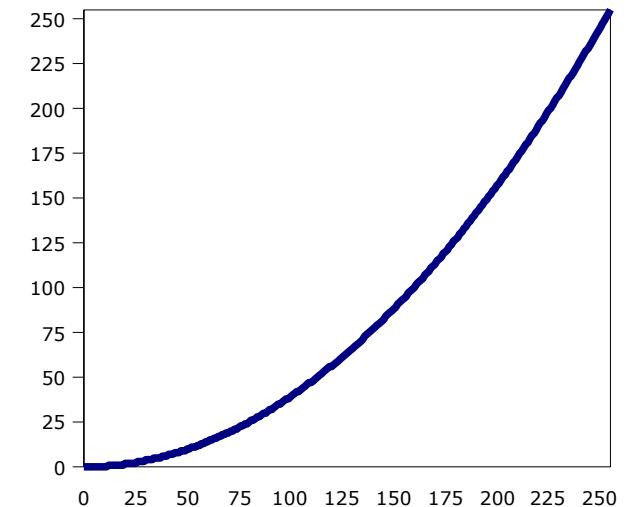
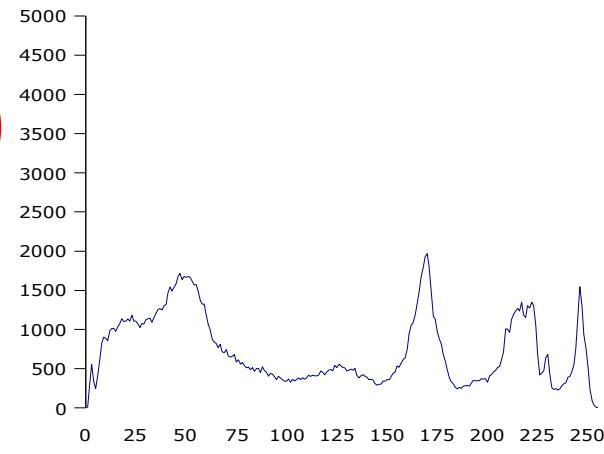


- lewe półkole powinno być praktycznie niewidoczne
- prawe półkole powinno być jak najbardziej widoczne



$\gamma > 1$

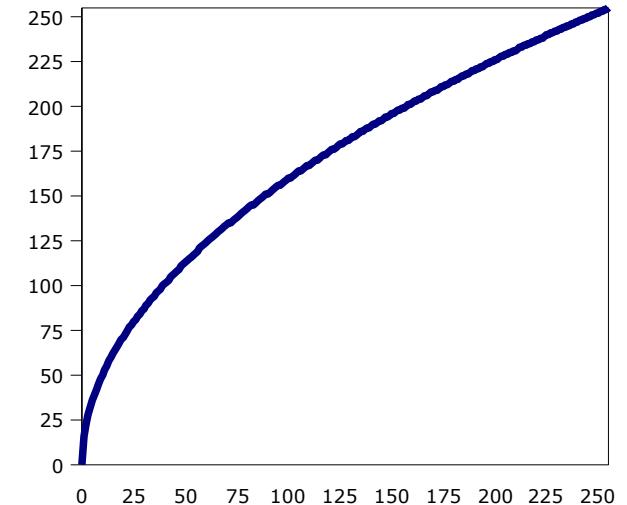
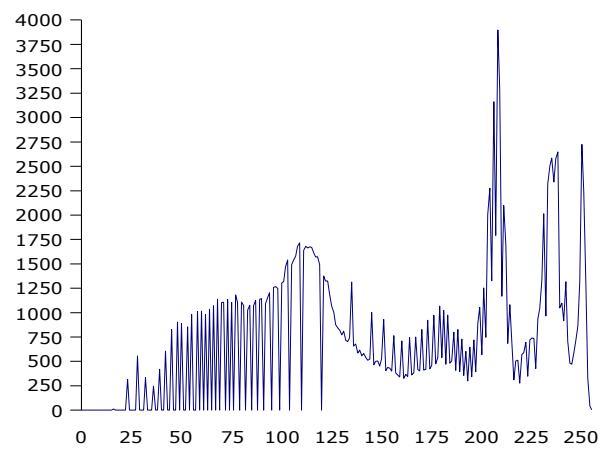
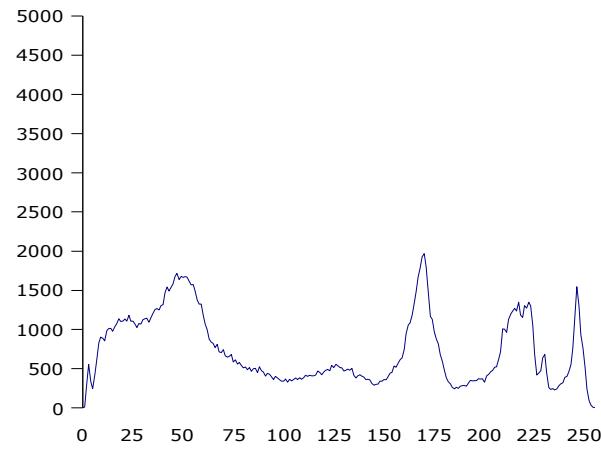
$\gamma = 2$  – ekspansja gamma, zwiększenie kontrastu jasnych obszarów





$\gamma < 1$

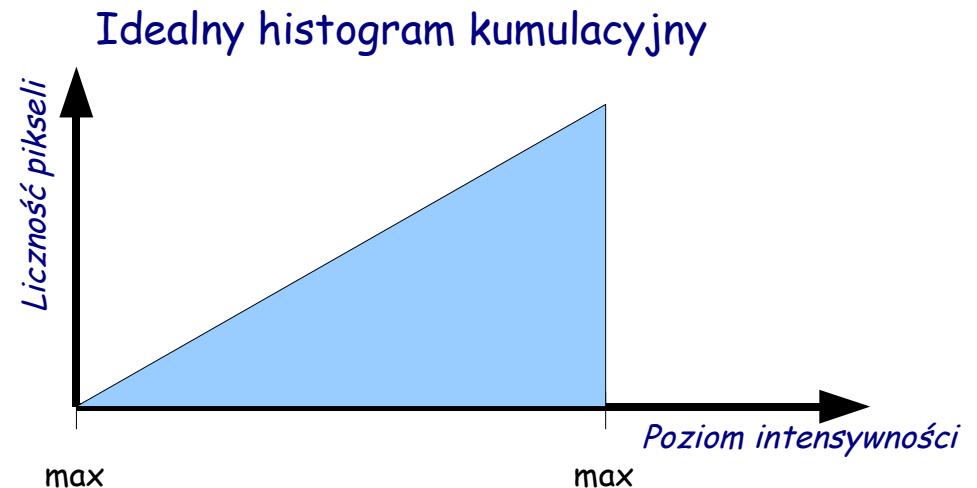
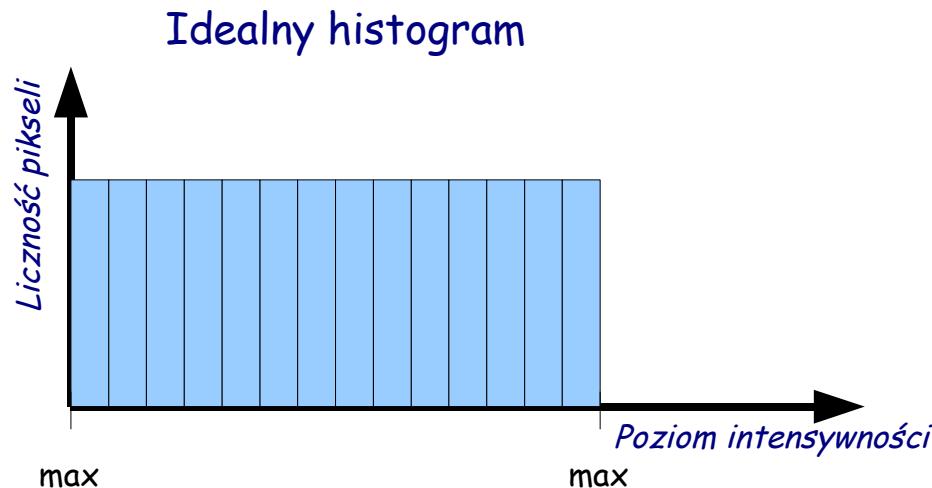
$\gamma = 0.5$  – kompresja gamma, zwiększenie kontrastu ciemnych obszarów





## Operacja poprawy histogramu

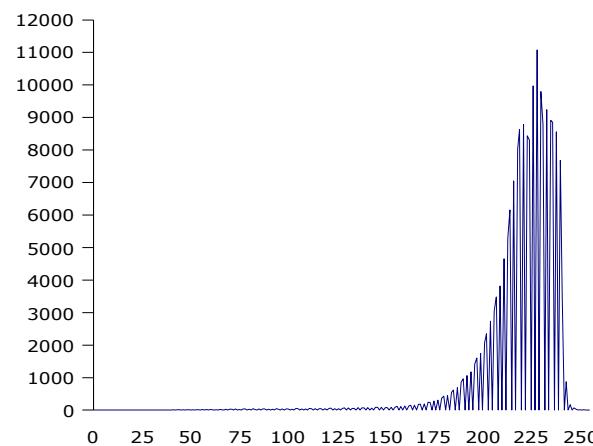
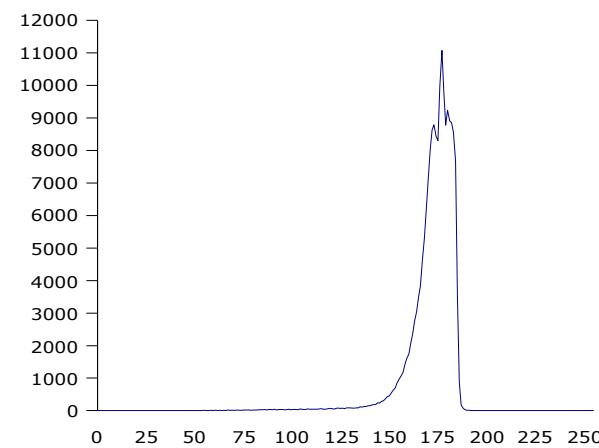
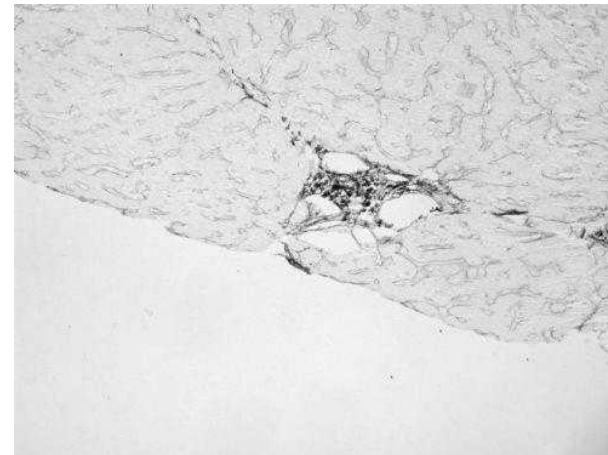
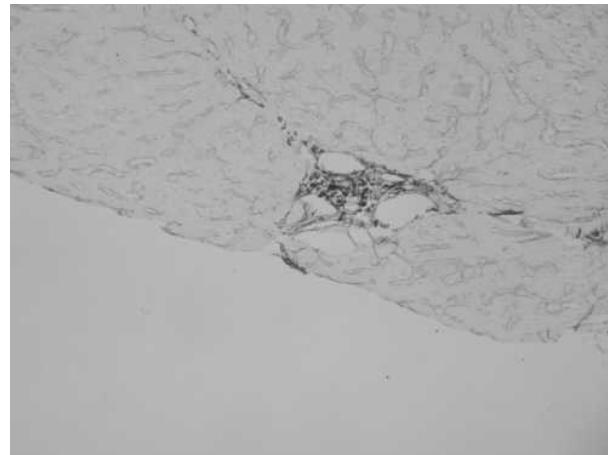
- mają za zadanie poprawić jakość kontrastu obrazu
- zmieniają histogram
- rozciągnięcie histogramu
  - piksele powinny używać wszystkich dostępnych poziomów intensywności
- wyrównanie (ang. *equalization*) histogramu
  - wszystkie poziomy powinny być w przybliżeniu równoliczne
  - czyli histogram powinien być płaski, bez górnego i dolnego skrótu





## Rozciagnięcie histogramu

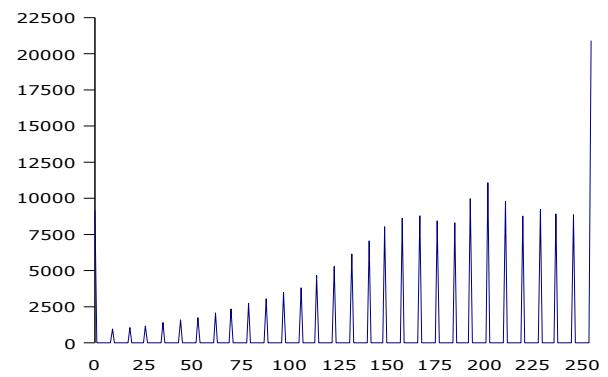
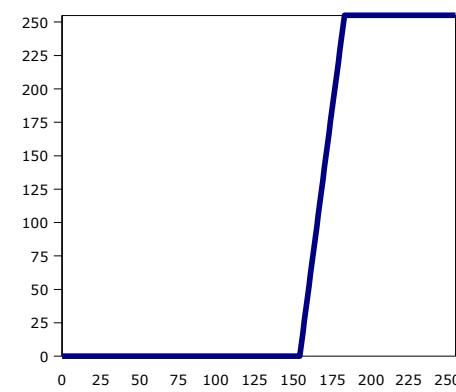
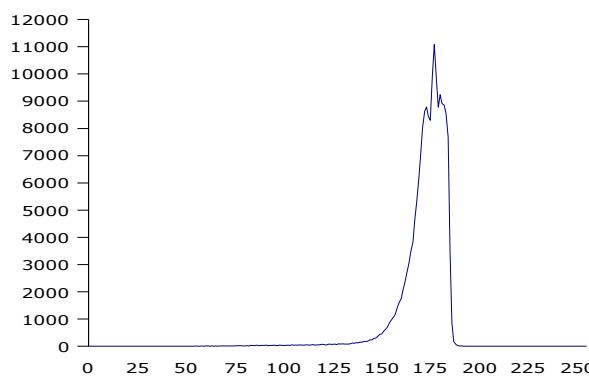
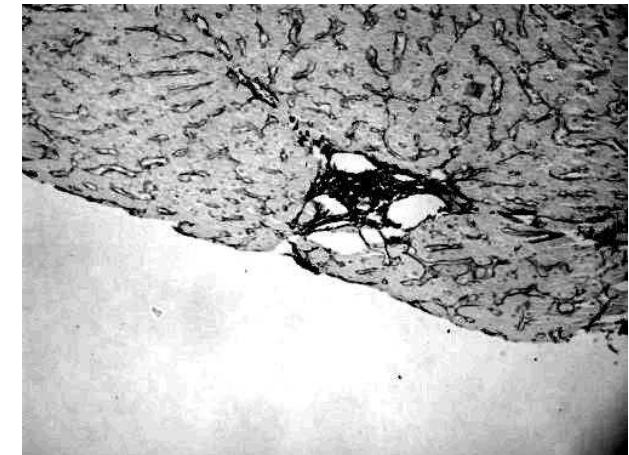
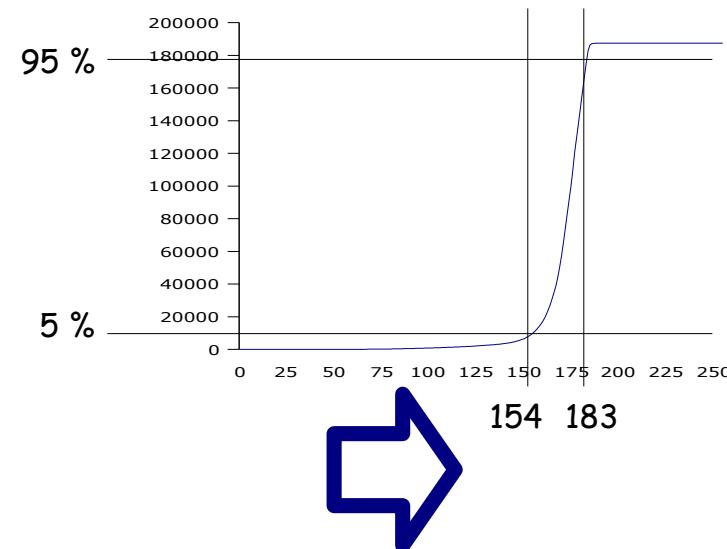
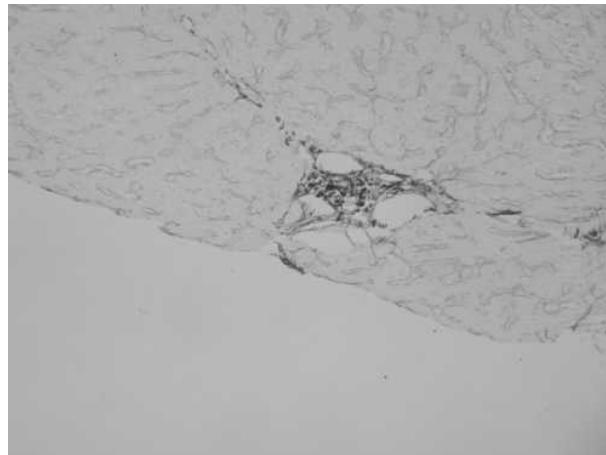
$$I(x, y) \Rightarrow (I(x, y) - \min) \cdot \frac{255}{\max - \min}$$





## Rozciagnięcie histogramu

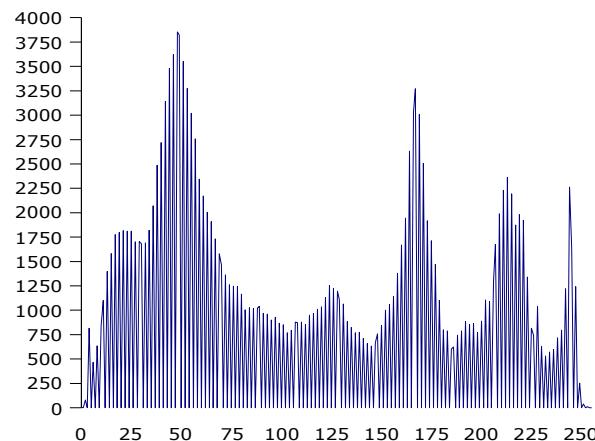
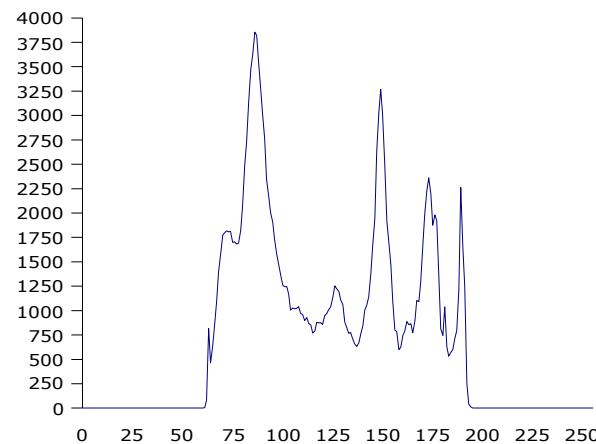
ponieważ „ogony” histogramu mogą być bardzo niskie i długie – zamiast *max* i *min* bierze się np. kwantyle 5% i 95% (percentyle) które można obliczyć z histogramu kumulacyjnego





## Rozciagnięcie histogramu

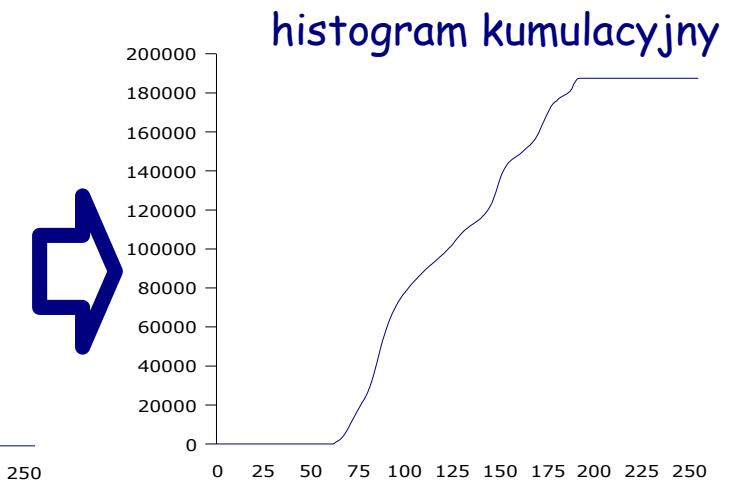
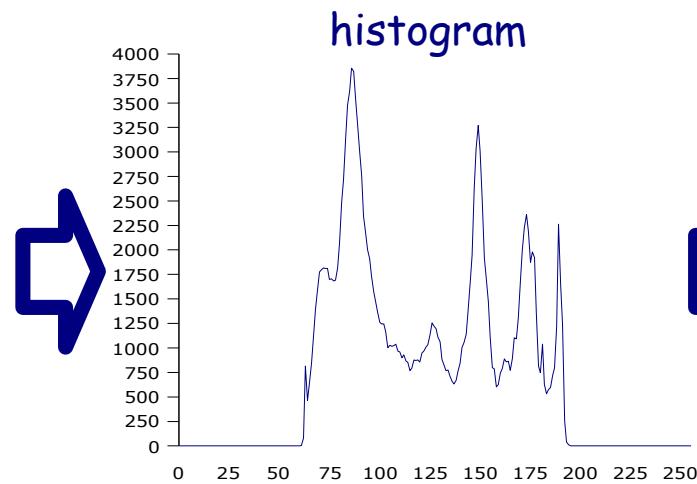
po rozciagnięciu nadal mogą pozostać górnki i doliny



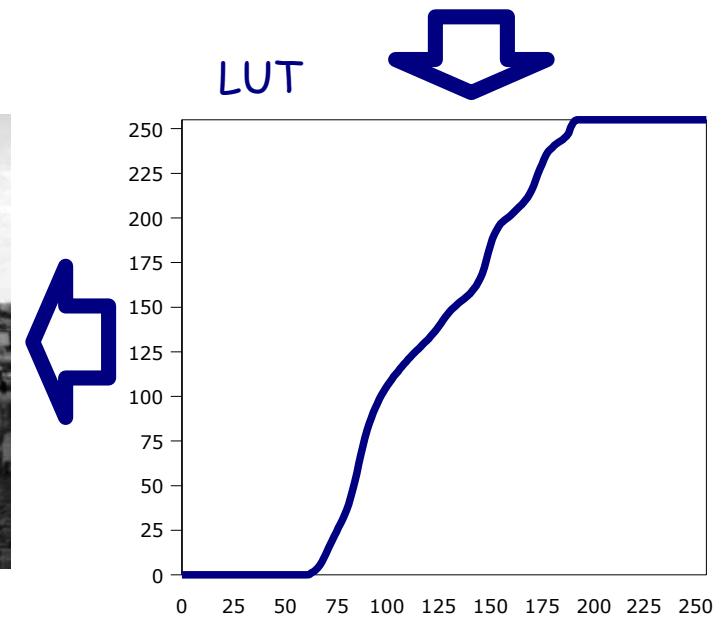
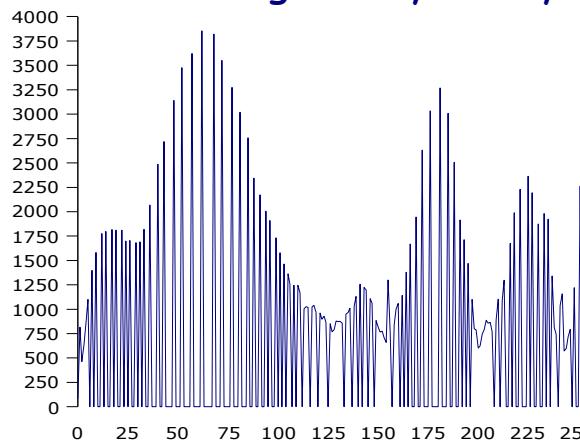


## Wyrównanie histogramu

jako tablicę LUT bierze się histogram kumulacyjny



histogram wynikowy



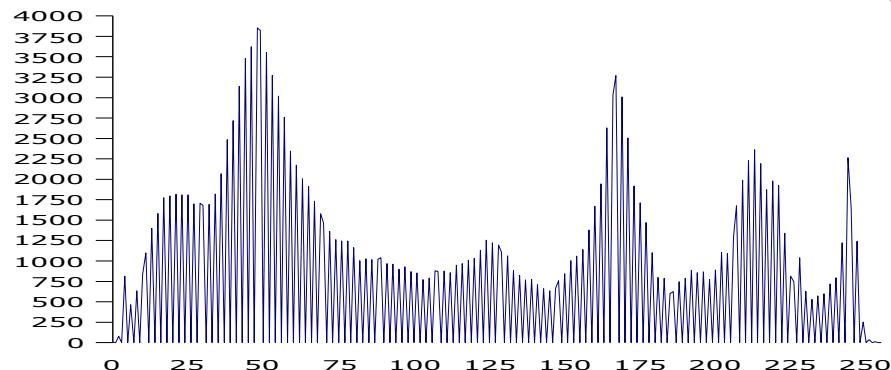


# Rozciągnięcie kontra wyrównanie histogramu

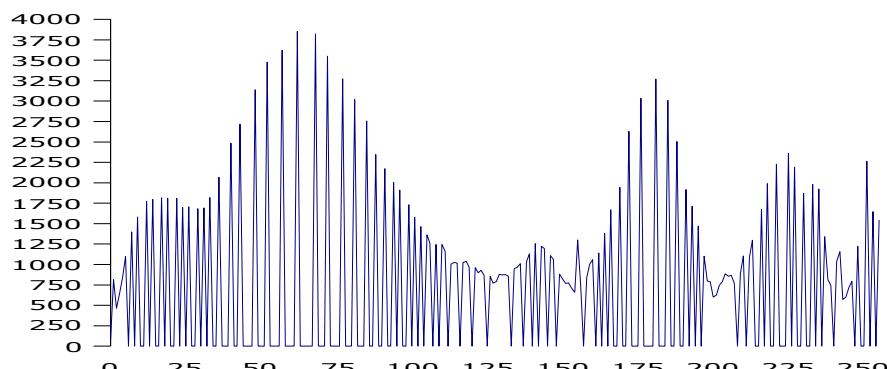
rozciągnięcie histogramu



wyrównanie histogramu



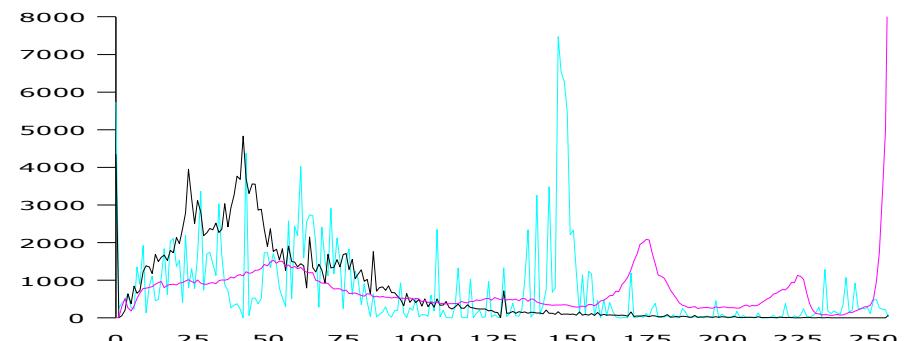
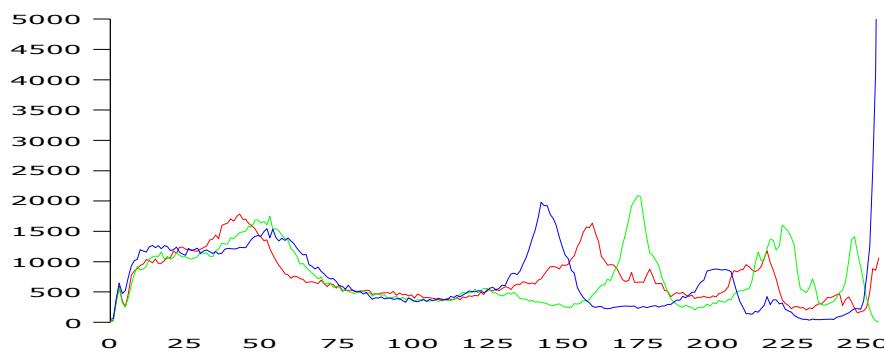
- wyrównanie daje lepsze wyniki (takie „inteligentne” rozciągnięcie), widać większe szczegółów
- histogram wyrównany nie jest idealny ale lepszy niż w rozciąganiu
- wyrównanie:
  - rozszerza „szczytach”
  - kompresuje „doliny”
- rozciągnięcie jest liniowe, wyrównanie – nieliniowe





# Operacja na histogramach obrazów kolorowych

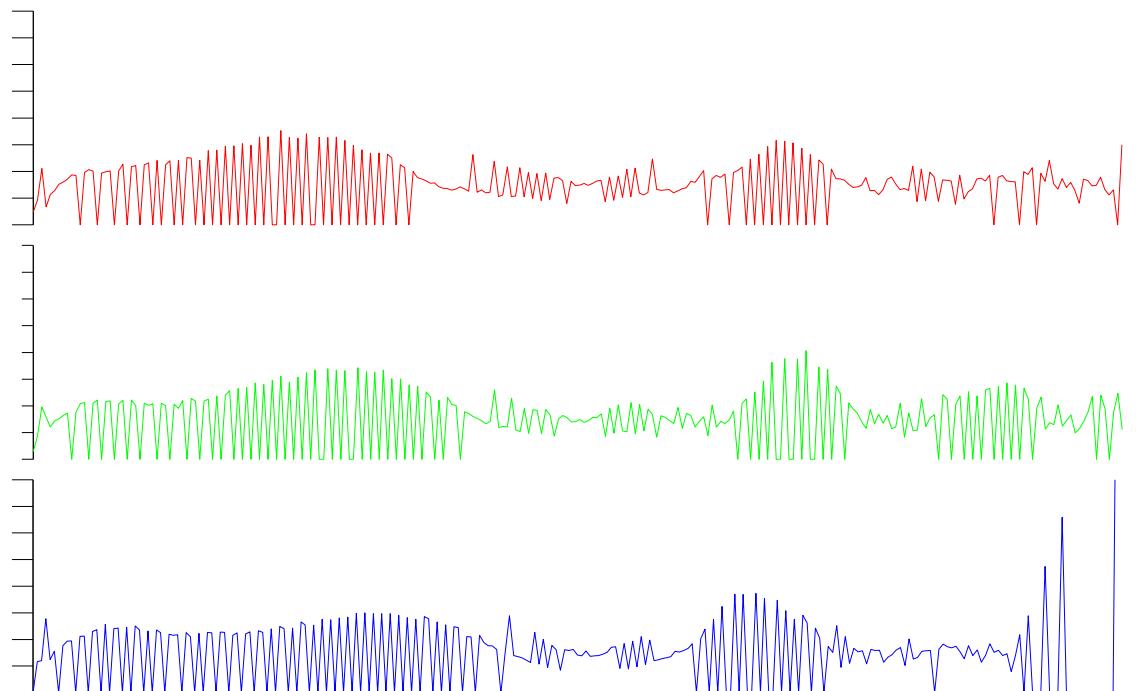
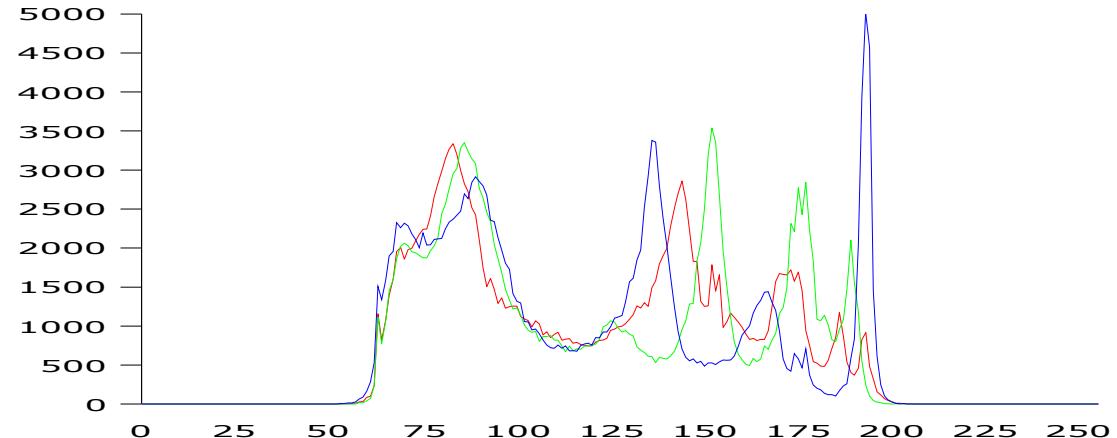
wykonywane na wszystkich kanałach (RGB, HSV) niezależnie lub z tymi samymi parametrami



# Wyrównanie histogramu obrazu kolorowego



każdy kanał RGB wyrównywany niezależnie



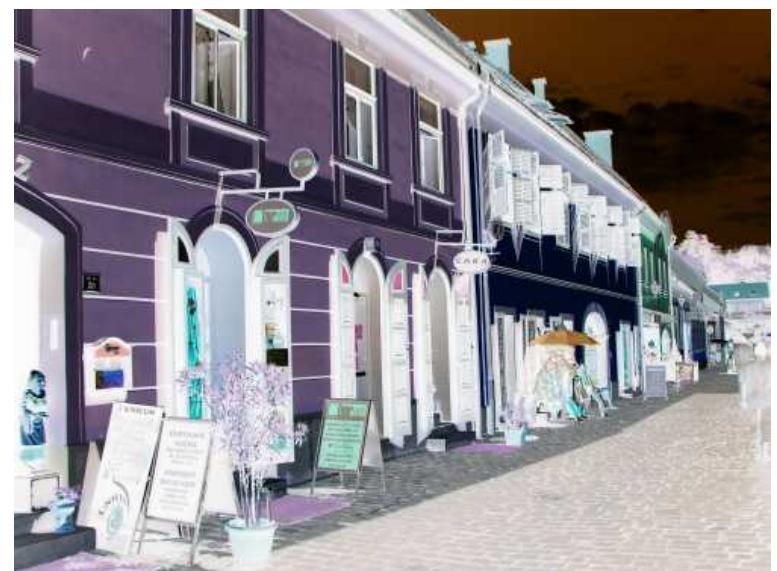


## Negatyw obrazu barwnego

- każdy kanał z RGB odwracany jest niezależnie

$$Image(x,y,kanał) = 255 - Image(x,y,kanał)$$

- operacja bezstratna – odwracalna





## Zwiększenie intensywności kanału R

- dodanie do każdego piksela wartości (50,0,0)
- „zaczerwienienie” obrazu





## Zmniejszenie intensywności kanału R

- odjęcie od każdego piksela wartości (50,0,0)
- dodanie składowej cyjan do obrazu





## Zwiększenie intensywności kanału G

- dodanie do każdego piksela wartości (0,50,0)
- „zazielenienie” obrazu





## Zmniejszenie intensywności kanału G

- odjęcie od każdego piksela wartości (0,50,0)
- dodanie składowej magenta (purpura, róż) do obrazu





## Zwiększenie intensywności kanału B

- dodanie do każdego piksela wartości (0,0,50)
- „zaniebesczenie” obrazu





## Zmniejszenie intensywności kanału B

- odjęcie od każdego piksela wartości (0,0,50)
- „zażółcenie” obrazu





## Balans kolorów

- balans kolorów (skala każdego z kanałów RGB) decyduje o naturalności kolorów (zgodnością z percepcją człowieka w świetle naturalnym)
- różne oświetlenie zmienia balans kolorów, konieczna jest korekcja balansu
- w prostych urządzeniach odbywa się to poprzez „korekcję balansu bieli”
  - mierzy się kolor wzorcowej bieli w danym oświetleniu ( $R_0, G_0, B_0$ )
  - wszystkie kolory modyfikuje się wg równania macierzowego

$$\begin{bmatrix} R \\ G \\ B \end{bmatrix} \Rightarrow \begin{bmatrix} 255/R_0 & 0 & 0 \\ 0 & 255/G_0 & 0 \\ 0 & 0 & 255/B_0 \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix}$$

- w bardziej zaawansowanych urządzeniach wykonuje się kalibrację kolorów wg tablicy kolorów wzorcowych – zależności są bardziej złożone niż powyżej



## Zwiększenie jasności

- dodanie do kanału *Value* w modelu kolorów HSV stałej wartości
- barwy i nasycenie nie ulegają zmianie





## Zmniejszenie jasności

- odjęcie od kanału *Value* w modelu kolorów HSV stałej wartości
- barwy i nasycenie nie ulegają zmianie





## Zwiększenie nasycenie kolorów

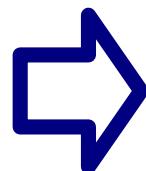
- dodanie do kanału *Saturation* w modelu kolorów HSV stałej wartości
- barwy i jasność nie ulegają zmianie





## Zmniejszenie nasycenie kolorów

- odjęcie od kanału *Saturation* w modelu kolorów HSV stałej wartości
- barwy i jasność nie ulegają zmianie



- ustawienie kanału *Saturation* na 0 tworzy obraz czarno-biały



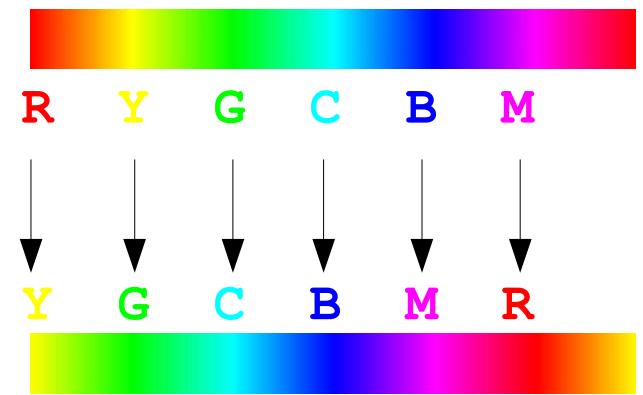
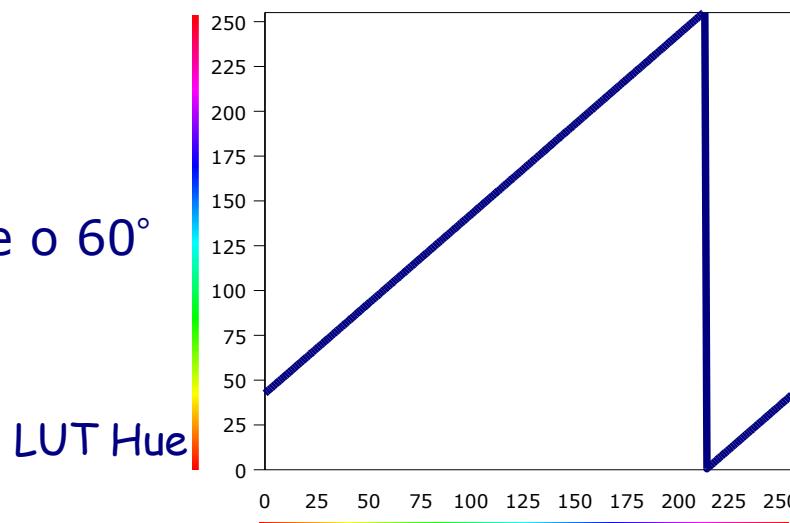


## Przesunięcie kolorów

- cykliczna przesunięcie kanału barwy *Hue* w modelu kolorów HSV



- przesunięcie o  $60^\circ$



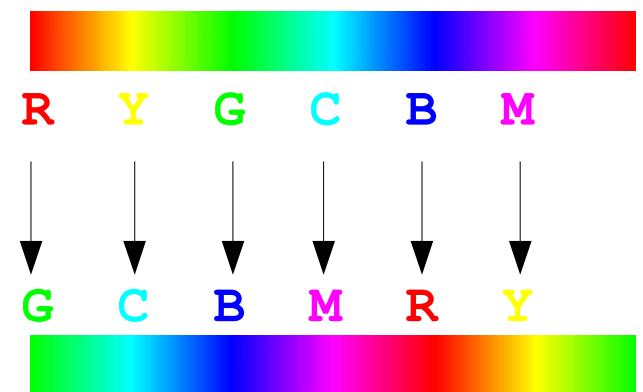
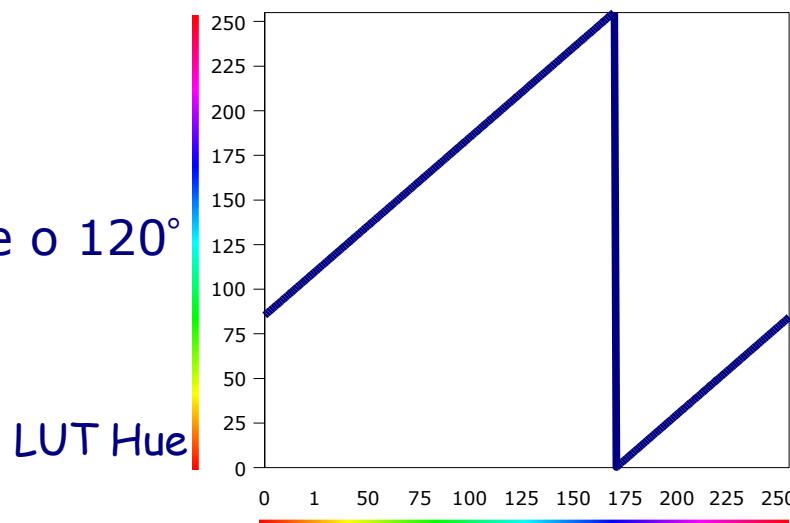


# Przesunięcie kolorów

- cykliczna przesunięcie kanału barwy *Hue* w modelu kolorów HSV



- przesunięcie o  $120^\circ$



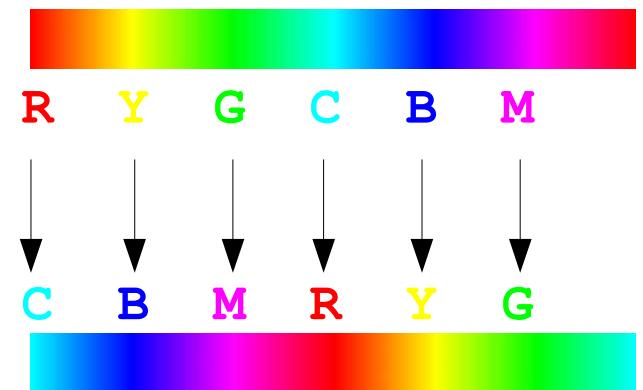
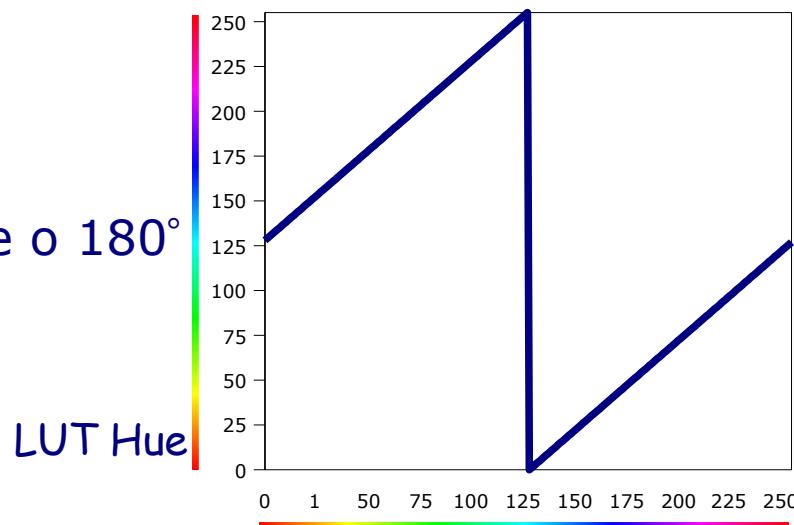


# Przesunięcie kolorów

- cykliczna przesunięcie kanału barwy *Hue* w modelu kolorów HSV



- przesunięcie o  $180^\circ$



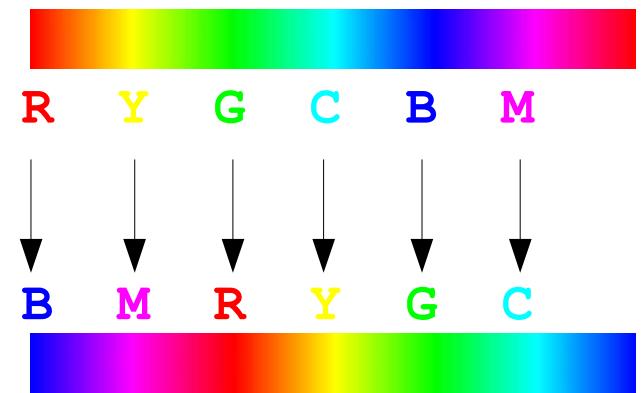
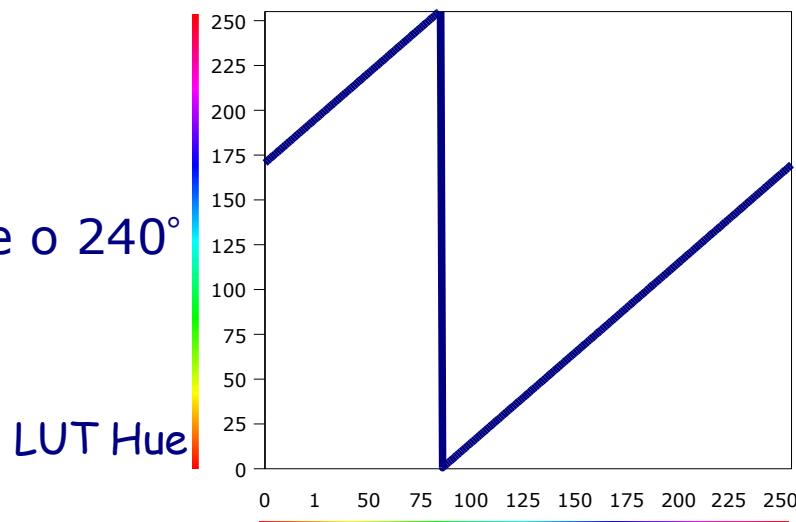


## Przesunięcie kolorów

- cykliczna przesunięcie kanału barwy *Hue* w modelu kolorów HSV



- przesunięcie o  $240^\circ$



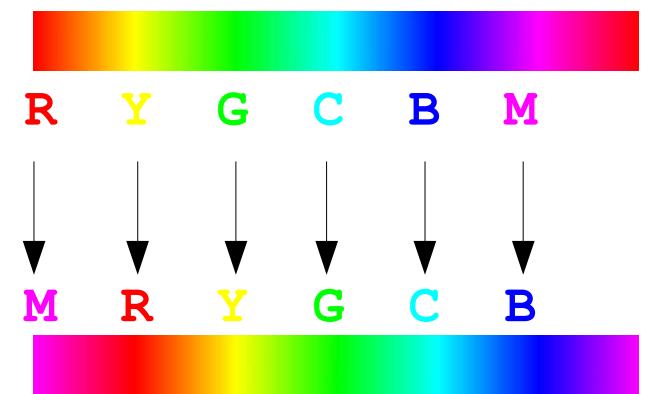
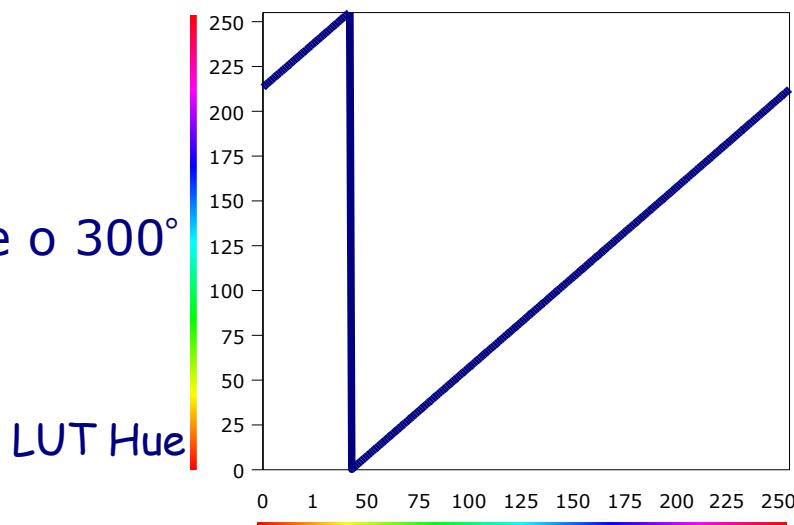


## Przesunięcie kolorów

- cykliczna przesunięcie kanału barwy *Hue* w modelu kolorów HSV



- przesunięcie o  $300^\circ$





## Arytmetyka obrazów

- z dwóch lub więcej obrazów powstaje jeden wynikowy
- zwykle stosowana do obrazów w skali szarości lub binarnych (ale nie tylko)

$Image1(x,y) \times Image2(x,y) \Rightarrow Image1(x,y) \text{ operation } Image2(x,y)$

- dodawanie obrazów
- odejmowanie obrazów
- mnożenie obrazów
- dzielenie obrazów
- operacje bitowe (OR, AND, XOR, NOT)



## Dodawanie obrazów

- efekt podwójnej ekspozycji

$$Image(x,y) = Image1(x,y) + Image2(x,y)$$

lub

$$Image(x,y,kanał) = Image1(x,y,kanał) + Image2(x,y,kanał)$$

- może pojawić się problem z wartościami powyżej maksymalnej wartości
- różne sposoby poradzenia sobie z tym zjawiskiem:
  - przeskalowanie obrazu wynikowego do oryginalnej skali  
$$Image(x,y) = ( Image1(x,y) + Image2(x,y) ) / 2$$
  - ustawienie wartości przekraczających *maks* na *maks*  
$$Image(x,y) = \min( maks, Image1(x,y) + Image2(x,y) )$$
  - efekt cykliczności – wartość wynikowa modulo *maks*  
$$Image(x,y) = ( Image1(x,y) + Image2(x,y) ) \% 256$$
- zastosowania:
  - efekt artystyczny
  - nakładanie tekstur
  - usuwanie szumu poprzez uśrednianie (dodanie proporcjonalne) kolejnych obserwacji tej samej sceny
  - uwypuklenie pewnych cech po dodaniu ich obrazu (np. wyostrzanie przez dodanie krawędzi)
  - dodanie stałej wartości rozjaśnia obraz



# Proporcjonalne dodawanie obrazów

$$Image(x,y) = ( Image1(x,y) + Image2(x,y) ) / 2$$



+



||





## Ważone dodawanie obrazów

$$Image(x,y) = W * Image1(x,y) + (1-W) * Image2(x,y), \quad 0 \leq W \leq 1$$

W=0,3



W=0,7



- automatyczne wyskalowanie obrazu wynikowego
- efekt przejścia między 2 obrazami płynnie zmieniając W:  $0 \rightarrow 1$



## Kanał alfa obrazów

- czwarty kanał, obok RGB, HSV, ...  
(lub drugi dla obrazów monochromatycznych)
- określa, w którym miejscu obraz jest przezroczysty (transparentny, najczęściej reprezentowany przez kolor czarny, waga  $W=0$ ) a w którym nieprzezroczysty (ang. *solid, opaque*, reprezentowany zwykle przez kolor biały, waga  $W=1$ )
- obraz przezroczysty „przepuszcza” część obrazu pod nim w stopniu określonym przez alfę (przezroczystość)
- ma to znaczenie przy nakładaniu się kilku obrazów



## Kanał alfa obrazów

obraz + jego kanał alfa



daje następujący wynik



nałożony na ten obraz+jego kanał alfa





## Dodawanie obrazów z saturacją

$$Image(x,y) = \min (255, Image1(x,y) + (1-W) * Image2(x,y) )$$



- mało czytelne obrazy
- regiony dominujące
- dużo obszarów z efektem saturacji - białych



## Dodawanie obrazów z „zawinięciem” wartości

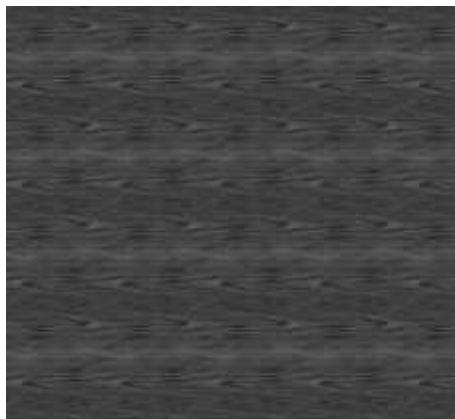
$$Image(x,y) = \text{modulo} (Image1(x,y) + (1-W) * Image2(x,y), 256 )$$



- można odnaleźć więcej szczegółów z obu obrazów
- nadal mało naturalny obraz



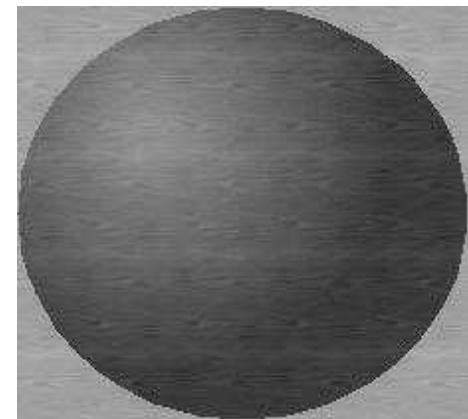
## Dodanie krawędzi do obrazu



+



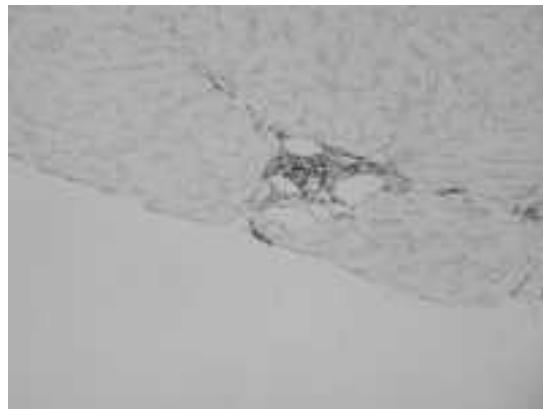
=



- dodawanie proporcjonalne
- obraz w dole po operacji dodania z saturacją tła oryginalnego obrazu kuli  
(tło uzyskane przez progowanie)



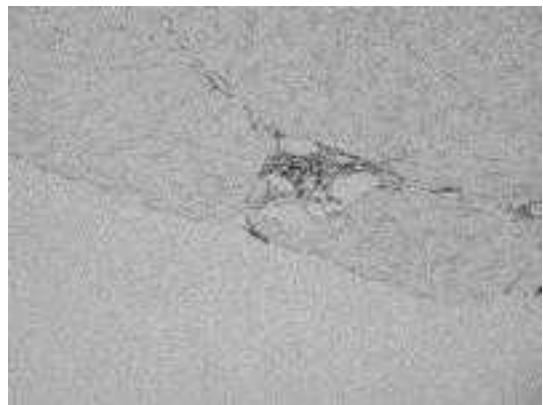
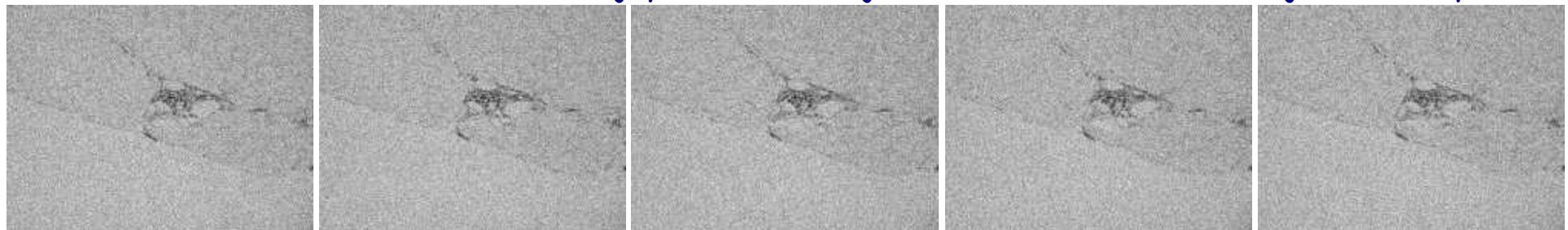
## Usuwanie szumu przez uśrednianie



Obraz oryginalny

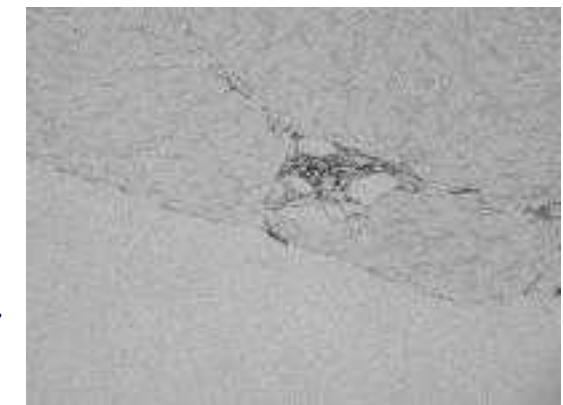
- nazywane uśrednianiem czasowym
- szum musi mieć średnią zero i stałą wariancję

5 kolejnych obserwacji z urządzenia które dodaje znaczący szum



Dodanie proporcjonalne  
5 powyższych obserwacji  
(szum zmalał)

Dodanie proporcjonalne  
10 obserwacji





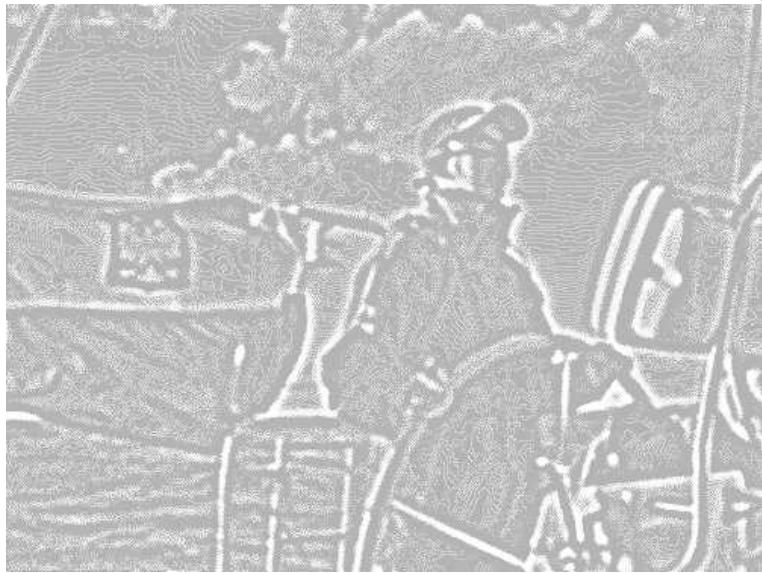
## Wystrzanie obrazu przez dodanie krawędzi



Obraz oryginalny (rozmyty)

+

||



Obraz krawędzi (Laplasjan)  
po wyrównaniu histogramu  
dla lepszej wizualizacji



Obraz wynikowy po dodaniu  
obrazu oryginalnego  
i krawędzi  
(tak naprawdę  
po 2 takich operacjach)



## Odejmowanie obrazów

$$Image(x,y) = Image1(x,y) - Image2(x,y)$$

lub

$$Image(x,y,kanał) = Image1(x,y,kanał) - Image2(x,y,kanał)$$

- może pojawić się problem z wartościami ujemnymi
- różne sposoby poradzenia sobie z tym zjawiskiem:
  - wartość bezwzględna różnicy  
$$Image(x,y) = | Image1(x,y) - Image2(x,y) |$$
  - ustawienie wartości ujemnych na 0  
$$Image(x,y) = \max( 0, Image1(x,y) - Image2(x,y) )$$
  - efekt cykliczności – wartość wynikowa modulo maks  
$$Image(x,y) = ( Image1(x,y) - Image2(x,y) + 256 ) \% 256$$
- operacja w zasadzie asymetryczna
- zastosowania:
  - obrazy różnicowe – wyodrębnianie różnic między obrazami podobnych obiektów lub tego samego obiektu w różnych obrazach
  - odejmowanie od obrazu niejednorodnego tła – w sytuacji gdy np. oświetlenie przy akwizycji było nierówne
  - odjęcie stałej wartości zmniejsza jasność obrazu



## Odejmowanie – desaturacja

$$Image(x,y) = \max(0, Image1(x,y) - Image2(x,y))$$



- dużo regionów czarnych - nieczytelnych



## Odejmowanie – desaturacja

$$Image(x,y) = \max(0, Image2(x,y) - Image1(x,y))$$



- operacja komplementarna – inny obiekt jest pierwszoplanowy, a inny jest tłem



## Odejmowanie cykliczne

$$Image(x,y) = \text{modulo} (Image1(x,y) - Image2(x,y), 256)$$



- więcej widocznych szczegółów
- efekt nieczytelny – co widać na tym obrazie ?



## Odejmowanie cykliczne

$$Image(x,y) = \text{modulo} (Image2(x,y) - Image1(x,y), 256)$$



- operacja komplementarna – negatyw poprzedniej



## Odejmowanie bezwzględne

$$Image(x,y) = | Image2(x,y) - Image1(x,y) |$$



- operacja symetryczna
- widać 2 obrazy, trochę jak w dodawaniu



## Obrazy różnicowe

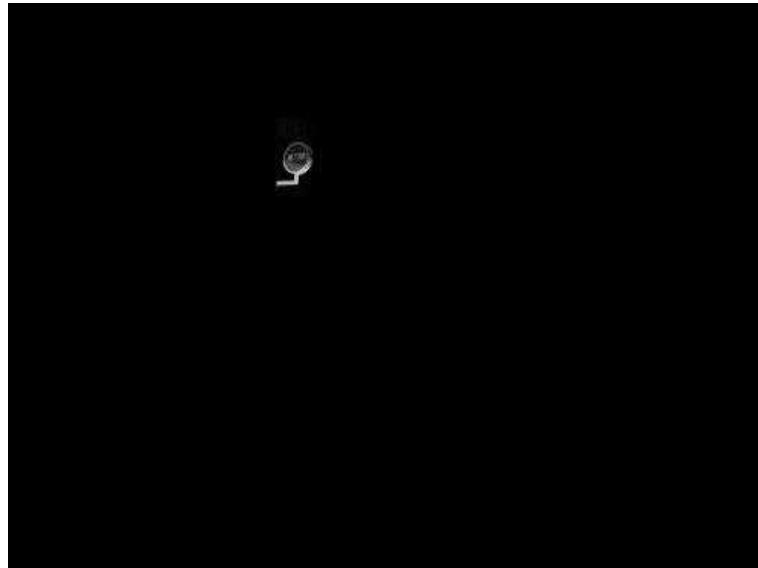


- zagadka – czym różnią się te dwa obrazy ?

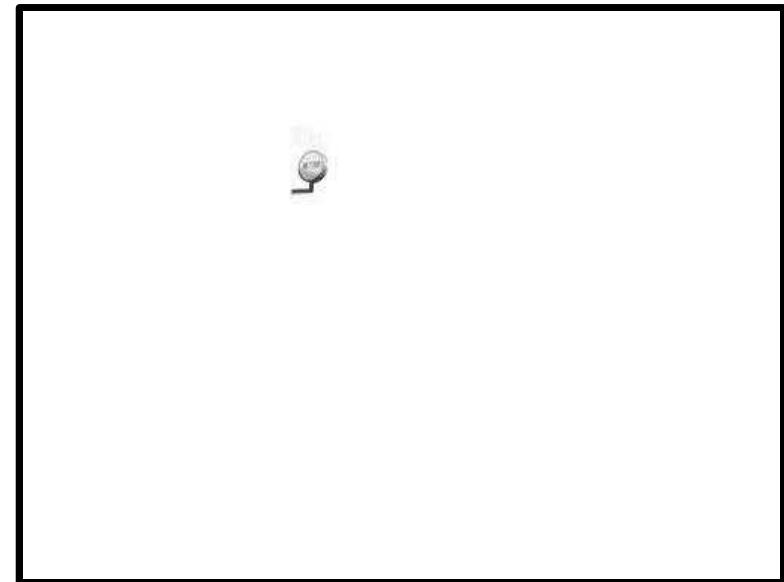


## Obrazy różnicowe

Obraz różnicowy (różnica bezwzględna)



Negatyw różnicy

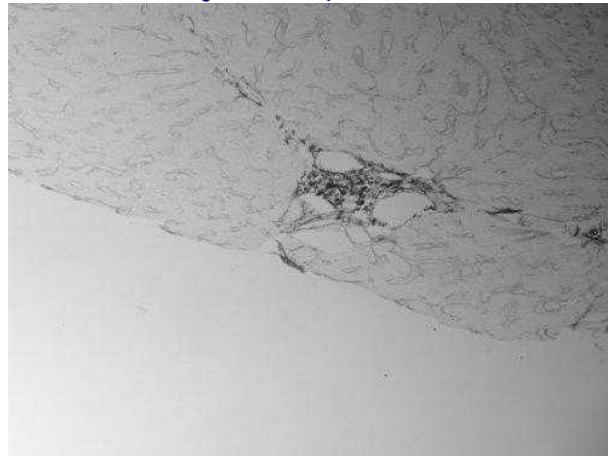


- obrazy różnicowe wykazują różnice
- często wykorzystywana technika w ocenie jakości segmentacji obrazów



## Odejmowanie tła

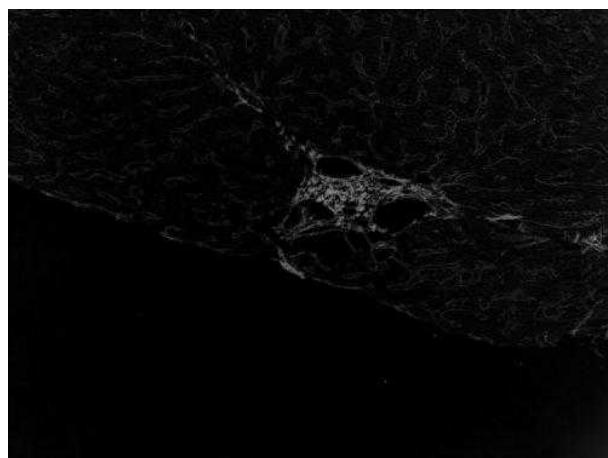
Obraz wejściowy



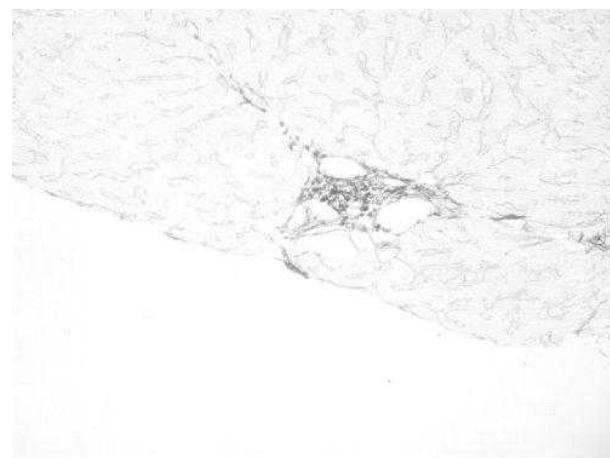
Kamera filmuje jednolite tło



Odjęcie tła od obrazu z obcięciem wartości



Negatyw odejmowania





## Mnożenie obrazów

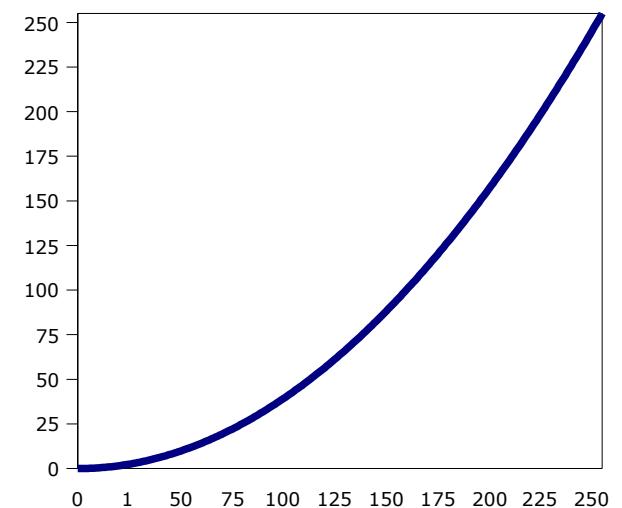
$$Image(x,y) = Image1(x,y) * Image2(x,y)$$

lub

$$Image(x,y,kanał) = Image1(x,y,kanał) * Image2(x,y,kanał)$$

- jak w dodawaniu może się pojawić problem z przekraczaniem zakresu
- rzadko stosuje się mnożenie 2 rzeczywistych, różnych obrazów
- wyjątek – mnożenie przez obraz binarny – nakładanie maski, ale lepiej nadają się tu bitowe operacje (AND)
- częściej mnożenie obrazu przez liczbę – rozjaśnia to obraz lepiej niż dodanie do obrazu stałej
- jeśli składowe obrazu reprezentowane są przez wartości z przedziału  $[0,1]$  wynik mnożenia nie wykracza z zakresu
- mnożenie takiego obrazu przez samego siebie powoduje pociemnienie obrazu

LUT potęgowania obrazu  
z wartościami z zakresu  $[0,1]$   
(wykres  $y=x^2$  w przedziale  $(0,1)$ )





## Potęgowanie obrazów

- potęgowanie obrazu którego wartości należą do zakresu  $[0,1]$  – obraz ciemnieje, wartości nie wychodzą z zakresu



- potęgowanie obrazu którego wartości należą do zakresu  $[0,255]$  – wartości prawie wszędzie wychodzą z zakresu (tu: obcięcie wartości poza zakresem – saturacja)

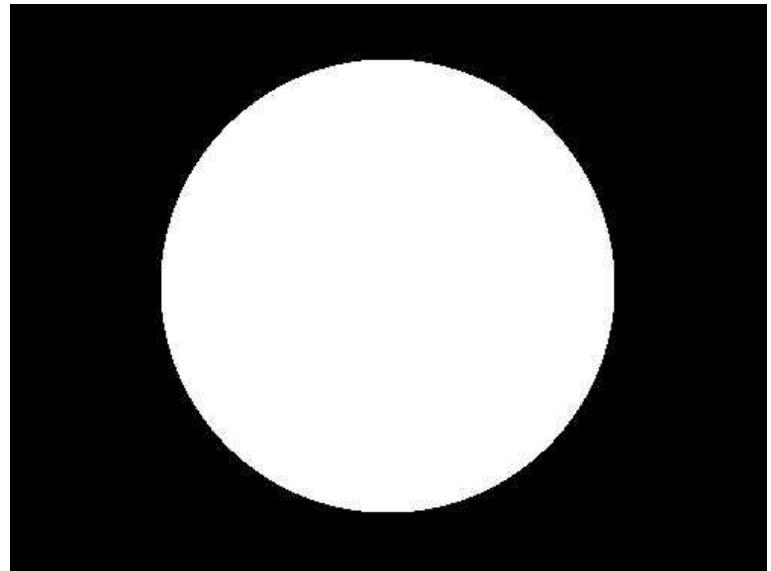




## Maskowanie binarne obrazów



\*



||



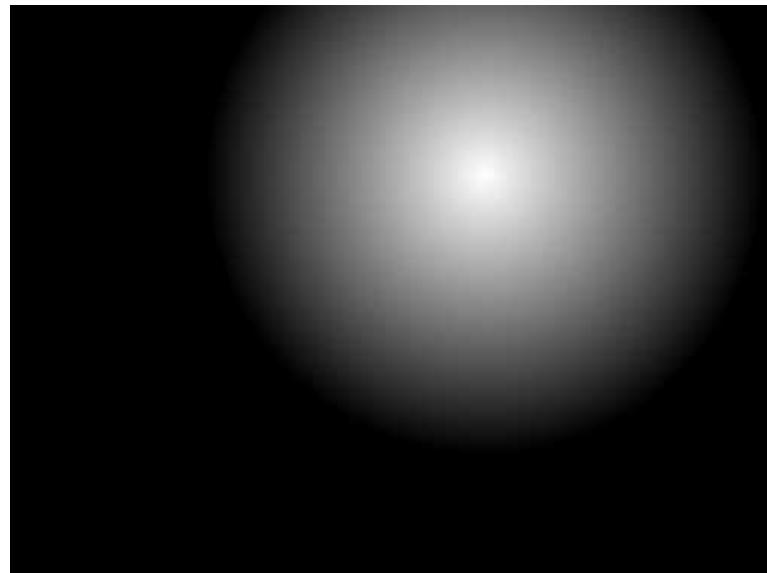
0 - czarny  
1 - biały



## Maskowanie ciągłe obrazów



\*



||



0 - czarny  
1 - biały



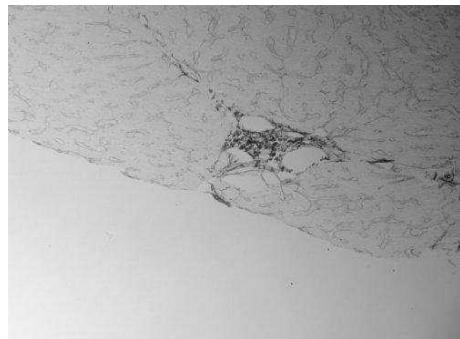
## Dzielenie obrazów

$$Image(x,y) = Image1(x,y) / Image2(x,y)$$

lub

$$Image(x,y,kanał) = Image1(x,y,kanał) / Image2(x,y,kanał)$$

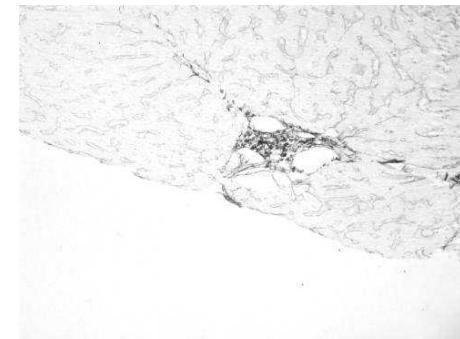
- uwaga na dzielenie przez zero !!!
- podobne zastosowanie do odejmowania obrazów (lepsze efekty)
  - dzielenie przez tło



/



=



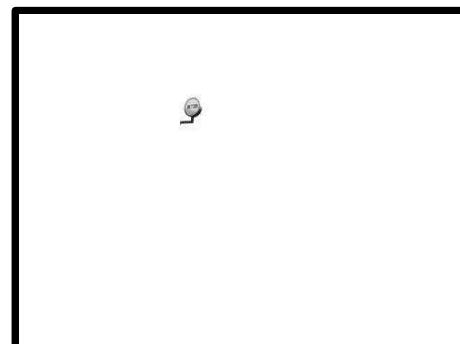
- wykrywanie w ruchu obrazu



/



=

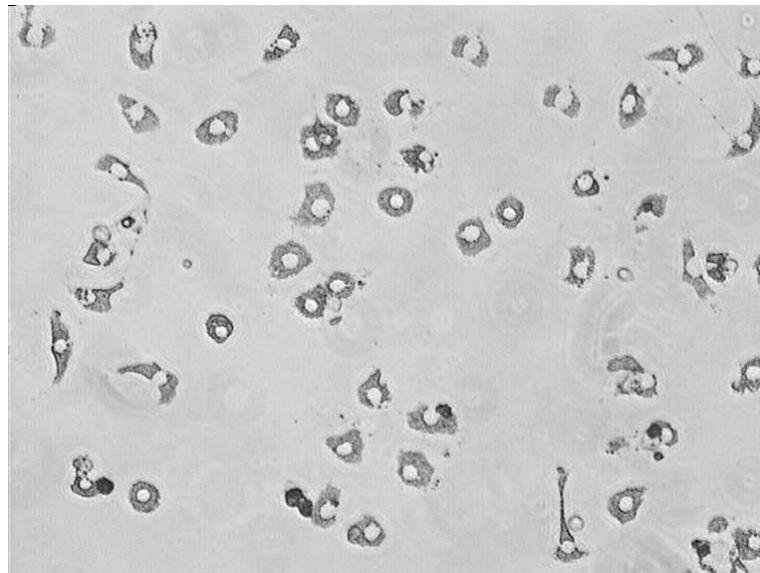




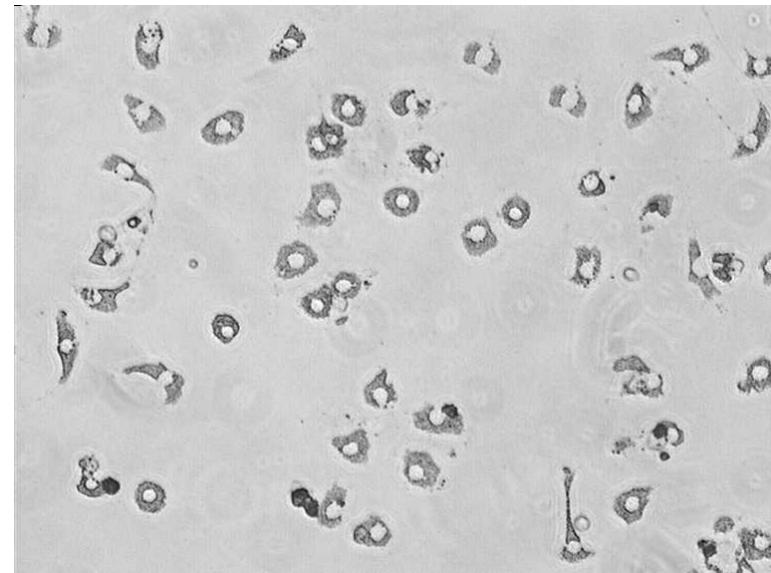
## Dzielenie obrazów w wykrywaniu ruchu

jedna z komórek poruszyła się, która ?

Obraz I1



Obraz I2





## Dzielenie obrazów w wykrywaniu ruchu

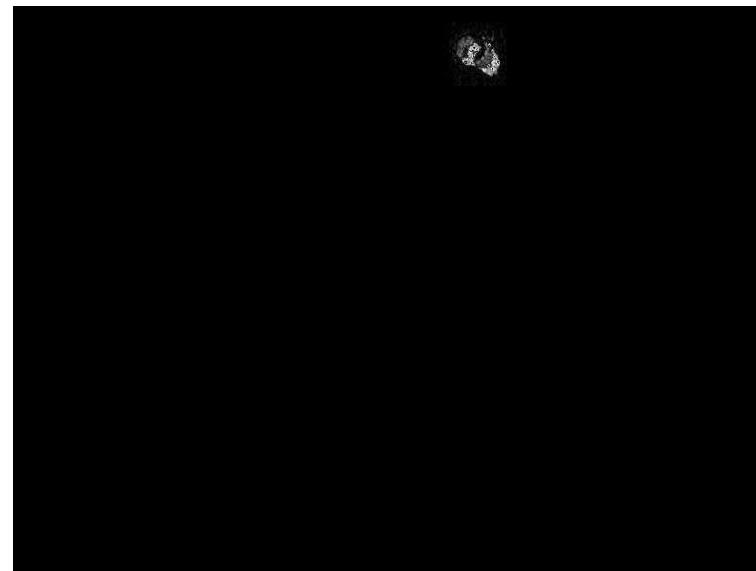
Obraz  $I_1/I_2$  - pozycja początkowa



Obraz  $I_2/I_1$  - pozycja końcowa



Obraz  $(I_1/I_2 - I_2/I_1)$   
- dynamika ruchu





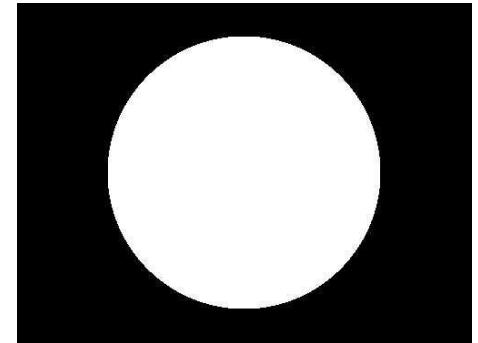
# Operacje na bitach obrazów

- najczęściej składanie z obrazami – maskami

00000000 - czarny  
11111111 - biały (255)



p	q	$p \wedge q$	p	q	$p \vee q$
0	0	0	0	0	0
0	1	0	0	1	1
1	0	0	1	0	1
1	1	1	1	1	1



operacja AND



operacja OR





## Operacja XOR

p	q	$p \oplus q$
0	0	0
0	1	1
1	0	1
1	1	0

- ze względu na fakt bycia własną operacją odwrotną ( $\oplus = \oplus^{-1}$ ,  $(X \oplus Y) \oplus Y = X$  ) była często stosowana do tymczasowego zaznaczania kształtów na obrazach
- usunięcie kształtu odbywa się poprzez jego ponowne narysowanie
- jest to szybka operacja
- algorytm opatentowany !!!: U.S. Patent 4,197,590 firma Cadtrack
- od jakiegoś czasu zniknął z rozwiązań bibliotek GUI :-(

# GNU Free Documentation License



Version 1.2, November 2002

Copyright (C) 2000,2001,2002 Free Software Foundation, Inc.  
51 Franklin St, Fifth Floor, Boston, MA 02110-1301 USA  
Everyone is permitted to copy and distribute verbatim copies  
of this license document, but changing it is not allowed.

## 0. PREAMBLE

The purpose of this License is to make a manual, textbook, or other functional and useful document "free" in the sense of freedom: to assure everyone the effective freedom to copy and redistribute it, with or without modifying it, either commercially or noncommercially. Secondly, this License preserves for the author and publisher a way to get credit for their work, while not being considered responsible for modifications made by others.

This License is a kind of "copyleft", which means that derivative works of the document must themselves be free in the same sense. It complements the GNU General Public License, which is a copyleft license designed for free software.

We have designed this License in order to use it for manuals for free software, because free software needs free documentation: a free program should come with manuals providing the same freedoms that the software does. But this License is not limited to software manuals; it can be used for any textual work, regardless of subject matter or whether it is published as a printed book. We recommend this License principally for works whose purpose is instruction or reference.

## 1. APPLICABILITY AND DEFINITIONS

This License applies to any manual or other work, in any medium, that contains a notice placed by the copyright holder saying it can be distributed under the terms of this License. Such a notice grants a world-wide, royalty-free license, unlimited in duration, to use that work under the conditions stated herein. The "Document", below, refers to any such manual or work. Any member of the public is a licensee, and is addressed as "you". You accept the license if you copy, modify or distribute the work in a way requiring permission under copyright law.

A "Modified Version" of the Document means any work containing the Document or a portion of it, either copied verbatim, or with modifications and/or translated into another language.

A "Secondary Section" is a named appendix or a front-matter section of the Document that deals exclusively with the relationship of the publishers or authors of the Document to the Document's overall subject (or to related matters) and contains nothing that could fall directly within that overall subject. (Thus, if the Document is in part a textbook of mathematics, a Secondary Section may not explain any mathematics.) The relationship could be a matter of historical connection with the subject or with related matters, or of legal, commercial, philosophical, ethical or political position regarding them.

The "Invariant Sections" are certain Secondary Sections whose titles are designated, as being those of Invariant Sections, in the notice that says that the Document is released under this License. If a section does not fit the above definition of Secondary then it is not allowed to be designated as Invariant. The Document may contain zero Invariant Sections. If the Document does not identify any Invariant Sections then there are none.

The "Cover Texts" are certain short passages of text that are listed, as Front-Cover Texts or Back-Cover Texts, in the notice that says that the Document is released under this License. A Front-Cover Text may be at most 5 words, and a Back-Cover Text may be at most 25 words.

A "Transparent" copy of the Document means a machine-readable copy, represented in a format whose specification is available to the general public, that is suitable for revising the document straightforwardly with generic text editors or (for images composed of pixels) generic paint programs or (for drawings) some widely available drawing editor, and that is suitable for input to text formatters or for automatic translation to a variety of formats suitable for input to text formatters. A copy made in an otherwise Transparent file format whose markup, or absence of markup, has been arranged to thwart or discourage subsequent modification by readers is not Transparent. An image format is not Transparent if used for any substantial amount of text. A copy that is not "Transparent" is called "Opaque".

Examples of suitable formats for Transparent copies include plain ASCII without markup, Texinfo input format, LaTeX input format, SGML or XML using a publicly available DTD, and standard-conforming simple HTML, PostScript or PDF designed for human modification. Examples of transparent image formats include PNG, XCF and JPG. Opaque formats include proprietary formats that can be read and edited only by proprietary word processors, SGML or XML for which the DTD and/or processing tools are not generally available, and the machine-generated HTML, PostScript or PDF produced by some word processors for output purposes only.

The "Title Page" means, for a printed book, the title page itself, plus such following pages as are needed to hold, legibly, the material this License requires to appear in the title page. For works in formats which do not have any title page as such, "Title Page" means the text near the most prominent appearance of the work's title, preceding the beginning of the body of the text.

A section "Entitled XYZ" means a named subunit of the Document whose title either is precisely XYZ or contains XYZ in parentheses following text that translates XYZ in another language. (Here XYZ stands for a specific section name mentioned below, such as "Acknowledgements", "Dedications", "Endorsements", or "History".) To "Preserve the Title" of such a section when you modify the Document means that it remains a section "Entitled XYZ" according to this definition.

The Document may include Warranty Disclaimers next to the notice which states that this License applies to the Document. These Warranty Disclaimers are considered to be included by reference in this License, but only as regards disclaiming warranties: any other implication that these Warranty Disclaimers may have is void and has no effect on the meaning of this License.

## 2. VERBATIM COPYING

You may copy and distribute the Document in any medium, either commercially or noncommercially, provided that this License, the copyright notices, and the license notice saying this License applies to the Document are reproduced in all copies, and that you add no other conditions whatsoever to those of this License. You may not use technical measures to obstruct or control the reading or further copying of the copies you make or distribute. However, you may accept compensation in exchange for copies. If you distribute a large enough number of copies you must also follow the conditions in section 3.

You may also lend copies, under the same conditions stated above, and you may publicly display copies.



### 3. COPYING IN QUANTITY

If you publish printed copies (or copies in media that commonly have printed covers) of the Document, numbering more than 100, and the Document's license notice requires Cover Texts, you must enclose the copies in covers that carry, clearly and legibly, all these Cover Texts: Front-Cover Texts on the front cover, and Back-Cover Texts on the back cover. Both covers must also clearly and legibly identify you as the publisher of these copies. The front cover must present the full title with all words of the title equally prominent and visible. You may add other material on the covers in addition. Copying with changes limited to the covers, as long as they preserve the title of the Document and satisfy these conditions, can be treated as verbatim copying in other respects.

If the required texts for either cover are too voluminous to fit legibly, you should put the first ones listed (as many as fit reasonably) on the actual cover, and continue the rest onto adjacent pages.

If you publish or distribute Opaque copies of the Document numbering more than 100, you must either include a machine-readable Transparent copy along with each Opaque copy, or state in or with each Opaque copy a computer-network location from which the general network-using public has access to download using public-standard network protocols a complete Transparent copy of the Document, free of added material. If you use the latter option, you must take reasonably prudent steps, when you begin distribution of Opaque copies in quantity, to ensure that this Transparent copy will remain thus accessible at the stated location until at least one year after the last time you distribute an Opaque copy (directly or through your agents or retailers) of that edition to the public.

It is requested, but not required, that you contact the authors of the Document well before redistributing any large number of copies, to give them a chance to provide you with an updated version of the Document.

### 4. MODIFICATIONS

You may copy and distribute a Modified Version of the Document under the conditions of sections 2 and 3 above, provided that you release the Modified Version under precisely this License, with the Modified Version filling the role of the Document, thus licensing distribution and modification of the Modified Version to whoever possesses a copy of it. In addition, you must do these things in the Modified Version:

- \* A. Use in the Title Page (and on the covers, if any) a title distinct from that of the Document, and from those of previous versions (which should, if there were any, be listed in the History section of the Document). You may use the same title as a previous version if the original publisher of that version gives permission.
- \* B. List on the Title Page, as authors, one or more persons or entities responsible for authorship of the modifications in the Modified Version, together with at least five of the principal authors of the Document (all of its principal authors, if it has fewer than five), unless they release you from this requirement.
- \* C. State on the Title page the name of the publisher of the Modified Version, as the publisher.
- \* D. Preserve all the copyright notices of the Document.
- \* E. Add an appropriate copyright notice for your modifications adjacent to the other copyright notices.
- \* F. Include, immediately after the copyright notices, a license notice giving the public permission to use the Modified Version under the terms of this License, in the form shown in the Addendum below.
- \* G. Preserve in that license notice the full lists of Invariant Sections and required Cover Texts given in the Document's license notice.
- \* H. Include an unaltered copy of this License.
- \* I. Preserve the section Entitled "History", Preserve its Title, and add to it an item stating at least the title, year, new authors, and publisher of the Modified Version as given on the Title Page. If there is no section Entitled "History" in the Document, create one stating the title, year, authors, and publisher of the Document as given on its Title Page, then add an item describing the Modified Version as stated in the previous sentence.
- \* J. Preserve the network location, if any, given in the Document for public access to a Transparent copy of the Document, and likewise the network locations given in the Document for previous versions it was based on. These may be placed in the "History" section. You may omit a network location for a work that was published at least four years before the Document itself, or if the original publisher of the version it refers to gives permission.
- \* K. For any section Entitled "Acknowledgements" or "Dedications", Preserve the Title of the section, and preserve in the section all the substance and tone of each of the contributor acknowledgements and/or dedications given therein.
- \* L. Preserve all the Invariant Sections of the Document, unaltered in their text and in their titles. Section numbers or the equivalent are not considered part of the section titles.
- \* M. Delete any section Entitled "Endorsements". Such a section may not be included in the Modified Version.
- \* N. Do not retitle any existing section to be Entitled "Endorsements" or to conflict in title with any Invariant Section.
- \* O. Preserve any Warranty Disclaimers.

If the Modified Version includes new front-matter sections or appendices that qualify as Secondary Sections and contain no material copied from the Document, you may at your option designate some or all of these sections as invariant. To do this, add their titles to the list of Invariant Sections in the Modified Version's license notice. These titles must be distinct from any other section titles.

You may add a section Entitled "Endorsements", provided it contains nothing but endorsements of your Modified Version by various parties--for example, statements of peer review or that the text has been approved by an organization as the authoritative definition of a standard.

You may add a passage of up to five words as a Front-Cover Text, and a passage of up to 25 words as a Back-Cover Text, to the end of the list of Cover Texts in the Modified Version. Only one passage of Front-Cover Text and one of Back-Cover Text may be added by (or through arrangements made by) any one entity. If the Document already includes a cover text for the same cover, previously added by you or by arrangement made by the same entity you are acting on behalf of, you may not add another; but you may replace the old one, on explicit permission from the previous publisher that added the old one.

The author(s) and publisher(s) of the Document do not by this License give permission to use their names for publicity for or to assert or imply endorsement of any Modified Version.



## 6. COLLECTIONS OF DOCUMENTS

You may make a collection consisting of the Document and other documents released under this License, and replace the individual copies of this License in the various documents with a single copy that is included in the collection, provided that you follow the rules of this License for verbatim copying of each of the documents in all other respects.

You may extract a single document from such a collection, and distribute it individually under this License, provided you insert a copy of this License into the extracted document, and follow this License in all other respects regarding verbatim copying of that document.

## 7. AGGREGATION WITH INDEPENDENT WORKS

A compilation of the Document or its derivatives with other separate and independent documents or works, in or on a volume of a storage or distribution medium, is called an "aggregate" if the copyright resulting from the compilation is not used to limit the legal rights of the compilation's users beyond what the individual works permit. When the Document is included in an aggregate, this License does not apply to the other works in the aggregate which are not themselves derivative works of the Document.

If the Cover Text requirement of section 3 is applicable to these copies of the Document, then if the Document is less than one half of the entire aggregate, the Document's Cover Texts may be placed on covers that bracket the Document within the aggregate, or the electronic equivalent of covers if the Document is in electronic form. Otherwise they must appear on printed covers that bracket the whole aggregate.

## 8. TRANSLATION

Translation is considered a kind of modification, so you may distribute translations of the Document under the terms of section 4. Replacing Invariant Sections with translations requires special permission from their copyright holders, but you may include translations of some or all Invariant Sections in addition to the original versions of these Invariant Sections. You may include a translation of this License, and all the license notices in the Document, and any Warranty Disclaimers, provided that you also include the original English version of this License and the original versions of those notices and disclaimers. In case of a disagreement between the translation and the original version of this License or a notice or disclaimer, the original version will prevail.

If a section in the Document is Entitled "Acknowledgements", "Dedications", or "History", the requirement (section 4) to Preserve its Title (section 1) will typically require changing the actual title.

## 9. TERMINATION

You may not copy, modify, sublicense, or distribute the Document except as expressly provided for under this License. Any other attempt to copy, modify, sublicense or distribute the Document is void, and will automatically terminate your rights under this License. However, parties who have received copies, or rights, from you under this License will not have their licenses terminated so long as such parties remain in full compliance.

## 10. FUTURE REVISIONS OF THIS LICENSE

The Free Software Foundation may publish new, revised versions of the GNU Free Documentation License from time to time. Such new versions will be similar in spirit to the present version, but may differ in detail to address new problems or concerns. See <http://www.gnu.org/copyleft/>.

Each version of the License is given a distinguishing version number. If the Document specifies that a particular numbered version of this License "or any later version" applies to it, you have the option of following the terms and conditions either of that specified version or of any later version that has been published (not as a draft) by the Free Software Foundation. If the Document does not specify a version number of this License, you may choose any version ever published (not as a draft) by the Free Software Foundation.

## 5. COMBINING DOCUMENTS

You may combine the Document with other documents released under this License, under the terms defined in section 4 above for modified versions, provided that you include in the combination all of the Invariant Sections of all of the original documents, unmodified, and list them all as Invariant Sections of your combined work in its license notice, and that you preserve all their Warranty Disclaimers.

The combined work need only contain one copy of this License, and multiple identical Invariant Sections may be replaced with a single copy. If there are multiple Invariant Sections with the same name but different contents, make the title of each such section unique by adding at the end of it, in parentheses, the name of the original author or publisher of that section if known, or else a unique number. Make the same adjustment to the section titles in the list of Invariant Sections in the license notice of the combined work.

In the combination, you must combine any sections Entitled "History" in the various original documents, forming one section Entitled "History"; likewise combine any sections Entitled "Acknowledgements", and any sections Entitled "Dedications". You must delete all sections Entitled "Endorsements."

(not as a draft) by the Free Software Foundation.

# Cyfrowe Przetwarzanie Obrazów

## Digital Image Processing

### Wykład 3



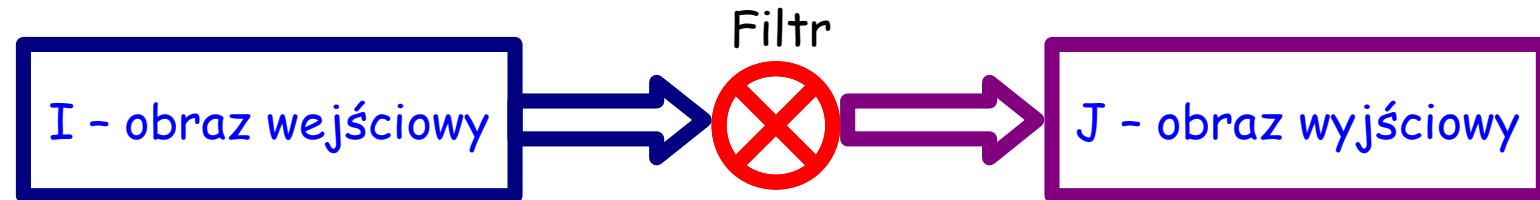
Aurelio A. Heckert

Copyright (c) 2008 Cezary Bołdak

Permission is granted to copy, distribute and/or modify this document under the terms of the GNU Free Documentation License, Version 1.2 or any later version published by the Free Software Foundation; with no Invariant Sections, no Front-Cover Texts, and no Back-Cover Texts. A copy of the license is included in the section entitled "GNU Free Documentation License"



## Cyfrowa filtracja obrazu - operacje kontekstowe



- piksel w obrazie wynikowym zależy od tego piksela w obrazie wejściowym  
i jego sąsiedztwa - kontekstu:

$$J(x, y) = f(I(x+i, y+j))$$

gdzie:  $i = -m, \dots, m$  oraz  $j = -n, \dots, n$

$(2*m+1, 2*n+1)$  - rozpatrywane otoczenie - kontekst

- różne cele, ale między innymi:
  - poprawa jakości obrazu (wystrzeganie, usuwanie szumu, ...)
  - wyodrębnianie cech z obrazu (krawędzie, struktury liniowe)
  - liczenie pewnych statystyk
- funkcja  $f$  może być dowolna, rozpatrujemy tu podział:
  - liniowa
  - nielinowa



## Filtracja liniowa obrazów

- piksel obrazu wyjściowego jest liniową kombinacją pikseli jego otoczenia w obrazie wejściowym:

$$\begin{aligned} J(x, y) = & a_{0,0} \cdot I(x-m, y-n) + a_{0,1} \cdot I(x-m+1, y-n) + \dots + a_{0,2m+1} \cdot I(x+m, y-n) + \\ & a_{1,0} \cdot I(x-m, y-n+1) + a_{1,1} \cdot I(x-m+1, y-n+1) + \dots + a_{1,2m+1} \cdot I(x+m, y-n+1) + \\ & \dots \\ & a_{2n+1,0} \cdot I(x-m, y+n) + a_{2n+1,1} \cdot I(x-m+1, y+n) + \dots + a_{2n+1,2m+1} \cdot I(x+m, y+n) \end{aligned}$$

$$J(x, y) = [a_{0,0} \ a_{0,1} \ \dots \ a_{0,2m+1}] * [I(x-m, y-n) \ I(x-m+1, y-n) \ \dots \ I(x+m, y-n)] \\ [a_{1,0} \ a_{1,1} \ \dots \ a_{1,2m+1}] * [I(x-m, y-n+1) \ I(x-m+1, y-n+1) \ \dots \ I(x+m, y-n+1)] \\ [\dots] * [\dots] \\ [a_{2n+1,0} \ a_{2n+1,1} \ \dots \ a_{2n+1,2m+1}] * [I(x-m, y+n) \ I(x-m+1, y+n) \ \dots \ I(x+m, y+n)]$$

$$J(x, y) = A * I(x-m : x+m, y-n : y+n)$$

$$J = A * I$$

- operacja **konwolucji** \*
- $A$  nazywana jest **maską** konwolucji



## Konwolucja (splot)

- w ujęciu matematycznym konwolucja (splot) 2 funkcji  $f$  i  $g$  to:

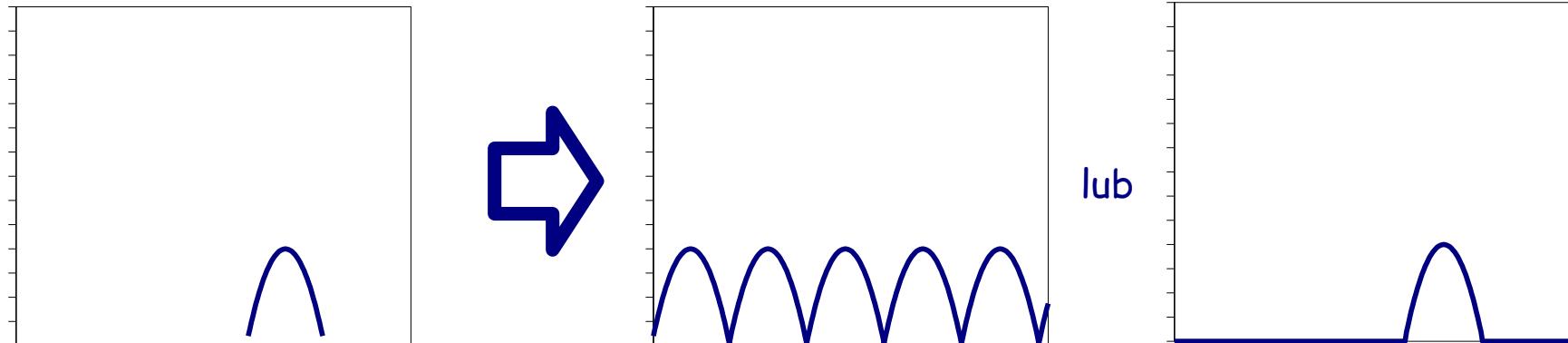
$$(f * g)(t) = \int_a^b f(\tau) g(t - \tau) d\tau$$

gdzie granice całkowania często  $a = -\infty$   $b = +\infty$

- w przypadku dyskretnym:

$$(f * g)(t) = \sum_i f(i) g(t - i)$$

- jeśli któraś z funkcji ma mniejszą dziedzinę lub chcemy policzyć splot w większym zakresie, można postąpić na 2 sposob:
  - uczynienie funkcji cykliczną – splot kołowy
  - rozszerzenie funkcji wartościami 0 – splot liniowy



- przypadek dwuwymiarowy

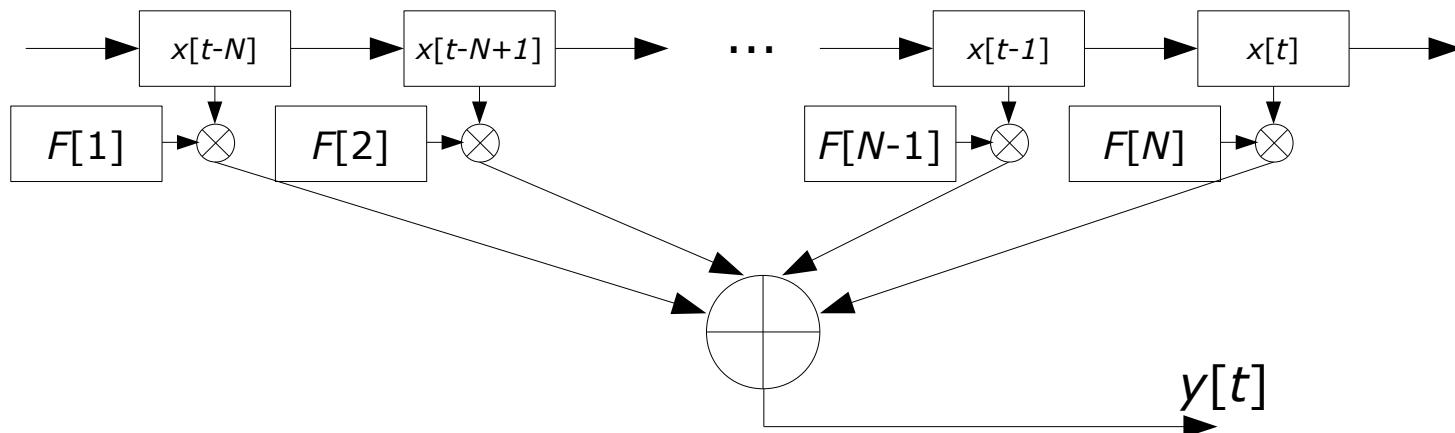
$$(f * g)(m, n) = \sum_{i,j} f(i, j) g(m - i, n - j)$$



## Konwolucja w przetwarzaniu sygnałów

$$(F * x)(t) = \sum_i F(i) x(t-i)$$

- zmienna  $t$  jest utożsamiana z czasem
- $x(t)$  lub  $x[t]$  reprezentuje sygnał cyfrowy w dyskretnym momencie czasowym  $t$
- $F$  reprezentuje filtr o  $N$  współczynnikach  $F[1], F[2], \dots, F[N]$
- filtr  $F$  sięga „wstecz”
- filtra FIR (o skończonej odpowiedzi impulsowej)



- $y[t]$  – sygnał wyjściowy w chwili  $t$



## Konwolucja w przetwarzaniu obrazów

$$(F * I)(x, y) = \sum_{i=-m, \dots, m} \sum_{j=-n, \dots, n} F(i, j) I(x+i, y+j)$$

- zmienne  $x$  i  $y$  odpowiadają współrzędnym obrazu  $I$
- filtr  $F$  o współczynnikach  $F(i,j)$  przykładana jest centralnie do każdego piksela
- macierz współczynników filtru  $F$  nazywana jest **maską konwolucji** o rozmiarach  $m \times n$
- w masce najczęściej współczynniki mają indeksy nieujemne

$$(F * I)(x, y) = \sum_{i=-m, \dots, m} \sum_{j=-n, \dots, n} F(i+m, j+n) I(x+i, y+j)$$

lub numerowane kolejno:

$$F = \begin{bmatrix} f_{0,0} & f_{1,0} & f_{2,0} \\ f_{0,1} & f_{1,1} & f_{2,1} \\ f_{0,2} & f_{1,2} & f_{2,2} \end{bmatrix} \Leftrightarrow \begin{bmatrix} f_1 & f_2 & f_3 \\ f_4 & f_5 & f_6 \\ f_7 & f_8 & f_9 \end{bmatrix}$$

- maskę można rozszerzać w dowolnym kierunku wypełniając ją zerami

$$I * \begin{bmatrix} f_1 & f_2 & f_3 \\ f_4 & f_5 & f_6 \\ f_7 & f_8 & f_9 \end{bmatrix} \Leftrightarrow I * \begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & f_1 & f_2 & f_3 & 0 \\ 0 & f_4 & f_5 & f_6 & 0 \\ 0 & f_7 & f_8 & f_9 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$



## 2 konwencje indeksowania obrazów i masek

- $I(współrzędna\_x, współrzędna\_y)$ ,  $I(x,y)$ ,  $F(x,y)$

- wzory matematyczne

$$F = \begin{bmatrix} f_{0,0} & f_{1,0} & f_{2,0} \\ f_{0,1} & f_{1,1} & f_{2,1} \\ f_{0,2} & f_{1,2} & f_{2,2} \end{bmatrix}$$

- $I(wiersz, kolumna)$ ,  $I(w,k)$ ,  $I(i,j)$ ,  $I(m,n)$ ,  $F(i,j)$

- częściej w rozwiązaniach implementacyjnych
  - Scilab

$$F = \begin{bmatrix} f_{0,0} & f_{0,1} & f_{0,2} \\ f_{1,0} & f_{1,1} & f_{1,2} \\ f_{2,0} & f_{2,1} & f_{2,2} \end{bmatrix}$$

- bardzo częsta przyczyna błędów: pomieszanie obu konwencji !!!

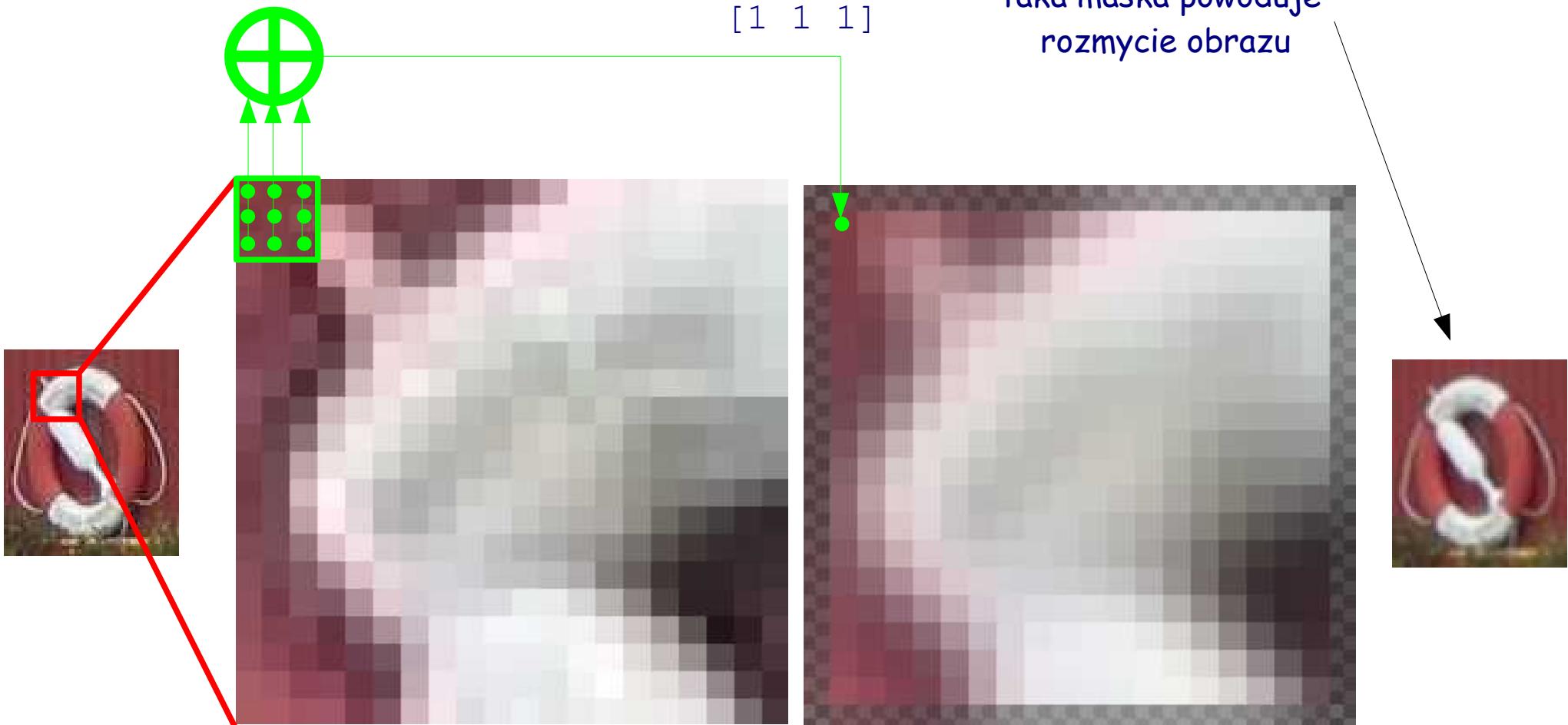


## Konwolucja w filtracji liniowej obrazów

- założymy jednolitą maskę 3x3:

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

taka maska powoduje  
rozmycie obrazu



- proces konwolucji polega na przejściu okna (maski) przez wszystkie piksele obrazu wynikowego



## Normalizacja wyniku konwolucji

- konwolucja może spowodować wyjście z zakresu intensywności obrazu:



$$* \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix} =$$



- aby tego uniknąć stosuje się dwa sposoby:

- jeżeli  $\sum M_{i,j} > 1$  dzieli się wynik przez sumę współczynników maski:

$$(F * I)(x, y) = \frac{\sum_{i=-m,..,m} \sum_{j=-n,..n} F(i+m, j+n) I(x+i, y+j)}{\sum_{i=-m,..,m} \sum_{j=-n,..n} F(i+m, j+n)}$$

- narzuca się warunek:  $\sum F_{i,j} = 1$  oraz bardzo często dodatkowo:  $\forall_{i,j} F_{i,j} \leq 1$

$$\begin{array}{ccc} [1 \ 1 \ 1] & [1/9 \ 1/9 \ 1/9] & [0 \ -1 \ 0] \\ [1 \ 1 \ 1] \Rightarrow [1/9 \ 1/9 \ 1/9] & \text{lub} & [-1 \ 5 \ -1] \\ [1 \ 1 \ 1] & [1/9 \ 1/9 \ 1/9] & [0 \ -1 \ 0] \end{array}$$

\*)taka maska nie daje 100% pewności, że wynik nie wyjdzie z zakresu



## Własności konwolucji

- konwolucja jest symetryczna:

$$f * g = g * f$$

zatem z obliczeniowego punktu widzenia nieważne, co jest maską a co obrazem, należy tylko dopasować wielkość obrazu i maski

- konwolucja jest łączna:

$$f * (g * h) = (f * g) * h$$

zatem kolejność filtrowania nie ma znaczenia  
a przyłożenie 2 mniejszych masek można zastąpić przyłożeniem 1 większej

$$I * \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix} * \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix} \Leftrightarrow I * \begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 1 & 0 \\ 0 & 1 & 1 & 1 & 0 \\ 0 & 1 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix} * \begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 1 & 0 \\ 0 & 1 & 1 & 1 & 0 \\ 0 & 1 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix} \Leftrightarrow I * \begin{bmatrix} 1 & 2 & 3 & 2 & 1 \\ 2 & 4 & 6 & 4 & 1 \\ 3 & 6 & 9 & 6 & 3 \\ 2 & 4 & 6 & 4 & 1 \\ 1 & 2 & 3 & 2 & 1 \end{bmatrix}$$

- konwolucja jest rozdzielna względem dodawania (i innych operacji skalarnych) :

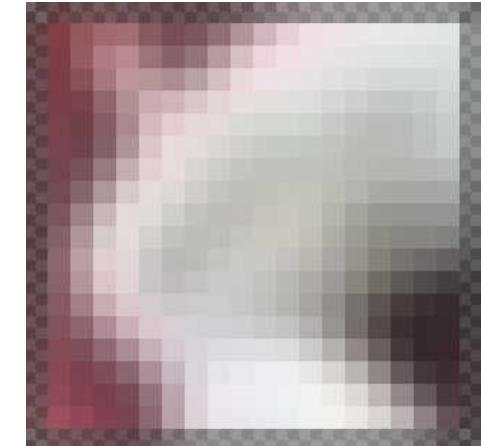
$$f * (I + J) = (f * I) + (f * J)$$

- „jedynka” konwolucji (element neutralny) to  $[1]$ :  $f * [1] = [1] * f = f$



## Elementy brzegowe

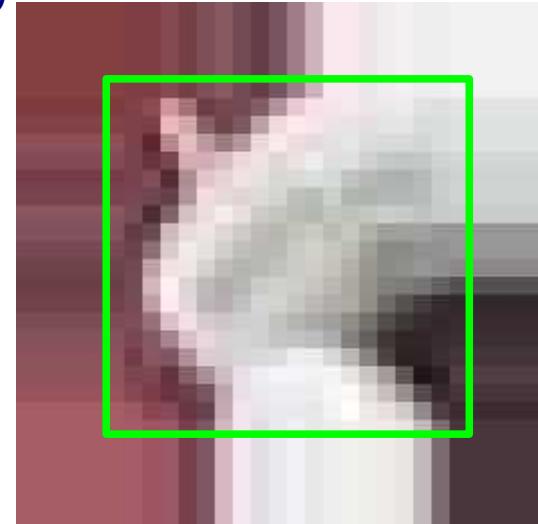
- dla elementów skrajnych nie można policzyć wartości konwolucji – nie mają one wystarczającej liczby sąsiadów
- szerokość pasa niewyliczonych wartości równy jest promieniowi maski
- czasami można sobie pozwolić na nieliczenie ich
- w przeciwnym przypadku można powiększyć obraz wejściowy uzupełniając go poza brzegi:
  - wartościami 0
  - powtarzając wartości z brzegów
  - „zawijając” obraz cyklicznie (współrzędne modulo rozmiar obrazu, kafelki)



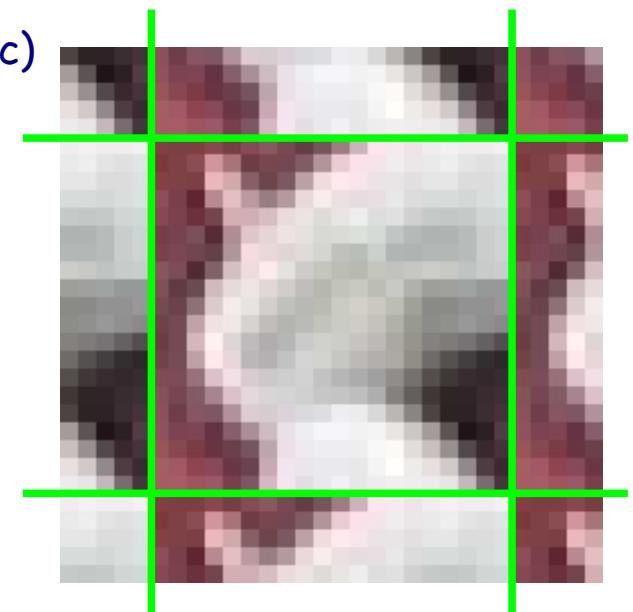
a)



b)



c)



# Obliczanie konwolucji



- przykładowy algorytm w programie Scilab:

```
[M,N] = size(Obraz); //obraz M wierszy i N kolumn  
[m,n] = size(maska); //maska mxn, m i n nieparzyste
```

- pierwszy sposób - wg definicji, wolniejszy:

```
rm = int(m,2); rn=int(n,2); //promienie maski  
Wynik = zeros(M,N);  
for y=1+rm:m-rm  
    for x=1+rn:n-rn
```

lub krócej

```
for i=-rm:rm  
    for j=-rn:rnrn  
        Wynik(y,x)=Wynik(y,x)+Obraz(y+i,x+j)  
            *maska(i+rm+1,j+rn+1);  
    end  
end
```

```
Wynik (y,x)=  
    Obraz (y-rm:y+rm,x-rn:x+rn) .*maska;
```

**end**

**end**

- drugi sposób - szybszy, zalecany, z pakietu 'SIP – Scilab Image Processing':

```
Wynik=imconv(Obraz,maska,'same');
```

trzeci parametr:

'same' - wynik ma rozmiar MxN

'full' - wynik ma rozmiar (M+rm)x(N+rn)

'valid' - wynik ma rozmiar (M-rm)x(N-rn)

# Złożoność obliczeniowa konwolucji



- zakładając  $M=N$  i  $n=n$ :
  - $N^2n^2$  mnożeń
  - $N^2n^2$  dodawań
- niewielka optymalizacja – alternatywny sposób z przesuwaniem obrazu:
  - $N^2n^2$  mnożeń
  - $N^2$  dodawań

$$I * \begin{bmatrix} f_{-1,-1} & f_{0,-1} & f_{1,-1} \\ f_{-1,0} & f_{0,0} & f_{1,0} \\ f_{-1,1} & f_{0,1} & f_{1,1} \end{bmatrix} = f_{-1,-1} \cdot T_{-1,-1}(I) + f_{0,-1} \cdot T_{0,-1}(I) + f_{1,-1} \cdot T_{1,-1}(I) + \\ f_{-1,0} \cdot T_{-1,0}(I) + f_{0,0} \cdot I + f_{1,0} \cdot T_{1,0}(I) + \\ f_{-1,1} \cdot T_{-1,1}(I) + f_{0,1} \cdot T_{0,1}(I) + f_{1,1} \cdot T_{1,1}(I)$$

gdzie  $T_{x,y}(I)$  – translacja obrazu  $I$  o wektor  $[x,y]$

$$f_{-1,-1}^* \begin{bmatrix} \square \end{bmatrix} + f_{0,-1}^* \begin{bmatrix} \square \end{bmatrix} + f_{1,-1}^* \begin{bmatrix} \square \end{bmatrix} + f_{-1,0}^* \begin{bmatrix} \square \end{bmatrix} + f_{0,0}^* \begin{bmatrix} \square \end{bmatrix} + f_{1,0}^* \begin{bmatrix} \square \end{bmatrix} + \\ + f_{-1,1}^* \begin{bmatrix} \square \end{bmatrix} + f_{0,1}^* \begin{bmatrix} \square \end{bmatrix} + f_{1,1}^* \begin{bmatrix} \square \end{bmatrix}$$

- koszt  $O(N^2 \log_2(N))$  – przy użyciu FFT (w dalszej części wykładu)



## Zastosowania: wygładzanie obrazu

- uśrednianie elementów obrazu
- różne wagi przykładane do elementu bieżącego
- różne zastosowania:
  - usuwanie szumu i zakłóceń
  - efekt filtru dolnoprzepustowego

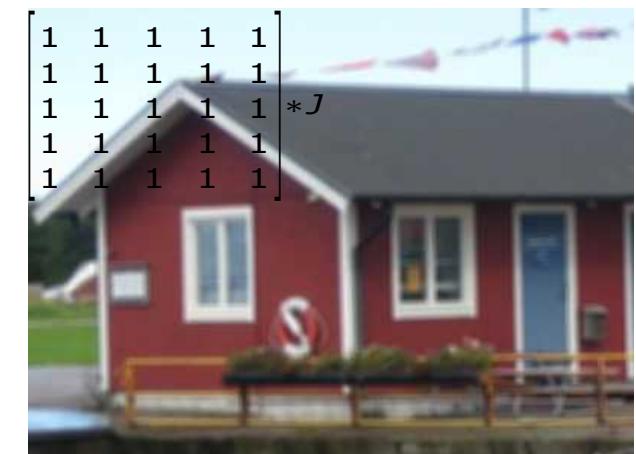
$$\begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix} \text{ lub } \begin{bmatrix} 1/9 & 1/9 & 1/9 \\ 1/9 & 1/9 & 1/9 \\ 1/9 & 1/9 & 1/9 \end{bmatrix} \quad \begin{bmatrix} 1 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 1 \end{bmatrix} \text{ lub } \begin{bmatrix} 0.125 & 0.125 & 0.125 \\ 0.125 & 0.125 & 0.125 \\ 0.125 & 0.125 & 0.125 \end{bmatrix} \quad \begin{bmatrix} 1 & 1 & 1 \\ 1 & 2 & 1 \\ 1 & 1 & 1 \end{bmatrix} \quad \begin{bmatrix} 1 & 1 & 1 \\ 1 & 4 & 1 \\ 1 & 1 & 1 \end{bmatrix}$$

- efekty uboczny:
  - rozmycie krawędzi
  - utrata ostrości
  - zamazanie detali



## Wygładzanie obrazu

obraz oryginalny  $I(x,y)$



- czym większa maska (mniejszy wpływ piksela centralnego) tym rozmycie większe

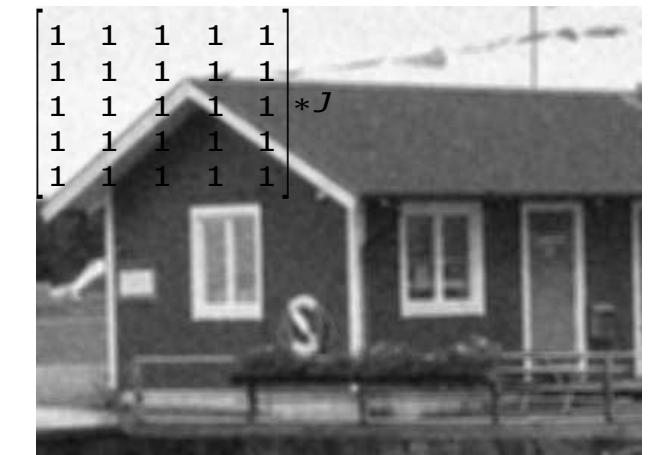
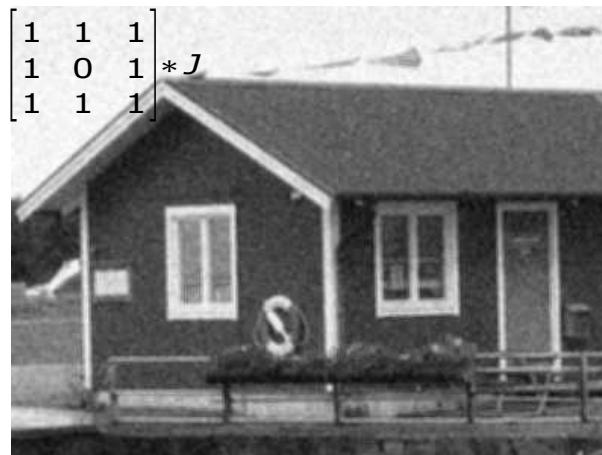


## Usuwanie szumu - szum biały

obraz oryginalny  $I(x,y)$



obraz z szumem  $J(x,y)=I(x,y)+\mathcal{N}(0,\sigma)$



- szum ulega zmniejszeniu kosztem rozmycia krawędzi i utraty szczegółów

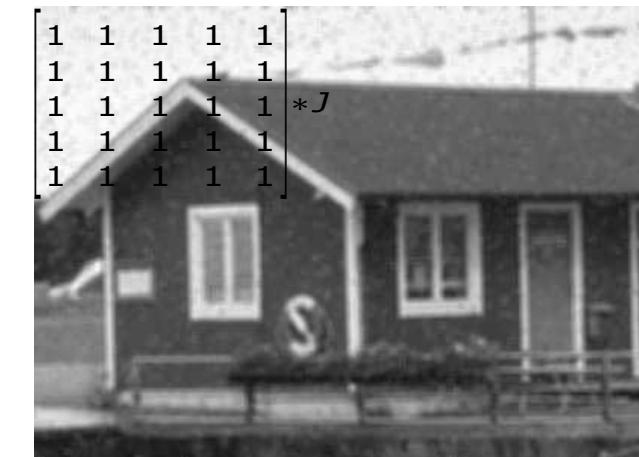
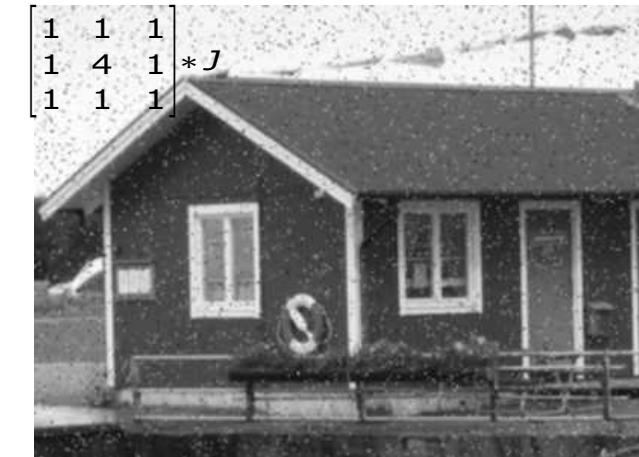
## Usuwanie szumu impulsowego (typu „salt and pepper”)



obraz oryginalny  $I(x,y)$



obraz z szumem  $J(x,y)$ : losowe piksele z  $I$  ustawione na losową wartość



- szum trudniejszy do usunięcia przez ten typ filtra



## Filtr Gaussa

- filtr Gaussa – współczynniki maski wyliczane są z równania rozkładu normalnego zakładając że centralny element ma współrzędne  $x=0$  i  $y=0$  oraz jednakową wariancję  $\sigma_x = \sigma_y = \sigma$

$$\frac{1}{2\pi\sigma^2} e^{-\frac{x^2+y^2}{2\sigma^2}}$$

$\sigma=1$

$$\begin{bmatrix} 0.06 & 0.1 & 0.06 \\ 0.1 & 0.16 & 0.1 \\ 0.06 & 0.1 & 0.06 \end{bmatrix} \sum F_{ij} = 0.78$$

$\sigma=1,5$

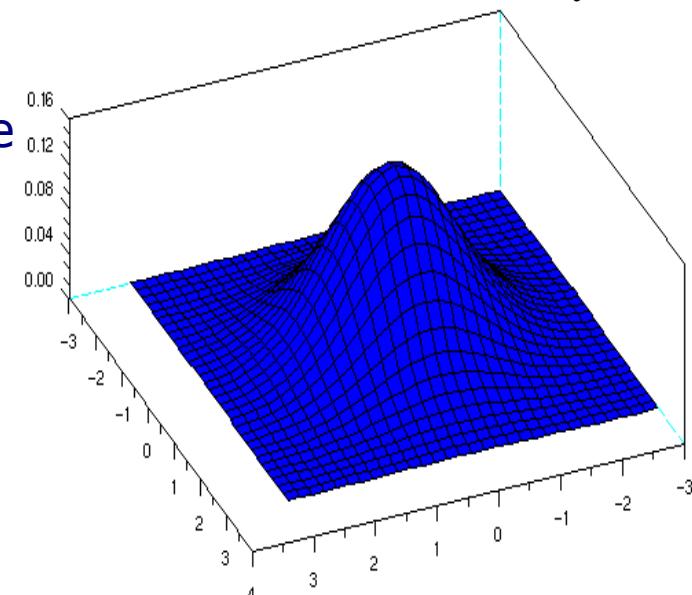
$$\begin{bmatrix} 0 & 0 & 0.01 & 0.01 & 0.01 & 0 & 0 \\ 0 & 0.01 & 0.02 & 0.03 & 0.02 & 0.01 & 0 \\ 0.01 & 0.02 & 0.05 & 0.06 & 0.05 & 0.02 & 0.01 \\ 0.01 & 0.03 & 0.06 & 0.07 & 0.06 & 0.03 & 0.01 \\ 0.01 & 0.02 & 0.05 & 0.06 & 0.05 & 0.02 & 0.01 \\ 0 & 0.01 & 0.02 & 0.03 & 0.02 & 0.01 & 0 \\ 0 & 0 & 0.01 & 0.01 & 0.01 & 0 & 0 \end{bmatrix} \sum F_{ij} = 0.97$$

$$\begin{bmatrix} 0 & 0.01 & 0.02 & 0.01 & 0 \\ 0.01 & 0.06 & 0.1 & 0.06 & 0.01 \\ 0.02 & 0.1 & 0.16 & 0.1 & 0.02 \\ 0.01 & 0.06 & 0.1 & 0.06 & 0.01 \\ 0 & 0.01 & 0.02 & 0.01 & 0 \end{bmatrix} \sum F_{ij} = 0.98$$

- im dalszy punkt tym mniejszy wpływ na wynik
- im większa wariancja rozkładu tym większe rozmycie
- często stosowana wersja ze współczynnikami -potęgami 2:

$$\begin{bmatrix} 1 & 2 & 1 \\ 2 & 4 & 2 \\ 1 & 2 & 1 \end{bmatrix}$$

$$\begin{bmatrix} 1 & 2 & 4 & 2 & 1 \\ 2 & 4 & 8 & 4 & 2 \\ 4 & 8 & 16 & 8 & 4 \\ 2 & 4 & 8 & 4 & 2 \\ 1 & 2 & 4 & 2 & 1 \end{bmatrix}$$





## Filtr Gaussa

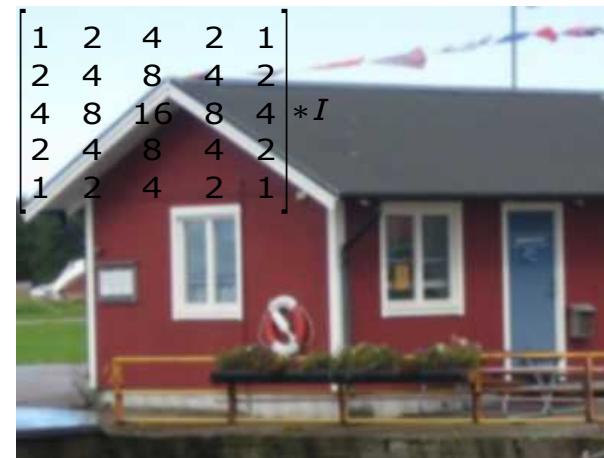
obraz oryginalny  $I(x,y)$



$$\begin{bmatrix} 1 & 2 & 1 \\ 2 & 4 & 2 \\ 1 & 2 & 1 \end{bmatrix} * I$$



$$\begin{bmatrix} 1 & 2 & 4 & 2 & 1 \\ 2 & 4 & 8 & 4 & 2 \\ 4 & 8 & 16 & 8 & 4 \\ 2 & 4 & 8 & 4 & 2 \\ 1 & 2 & 4 & 2 & 1 \end{bmatrix} * I$$



$$\begin{bmatrix} 0 & 0.01 & 0.02 & 0.01 & 0 \\ 0.01 & 0.06 & 0.1 & 0.06 & 0.01 \\ 0.02 & 0.1 & 0.16 & 0.1 & 0.02 \\ 0.01 & 0.06 & 0.1 & 0.06 & 0.01 \\ 0 & 0.01 & 0.02 & 0.01 & 0 \end{bmatrix} * I$$





## Filtr Gaussa w usuwaniu szumu

obraz oryginalny  $I(x,y)$



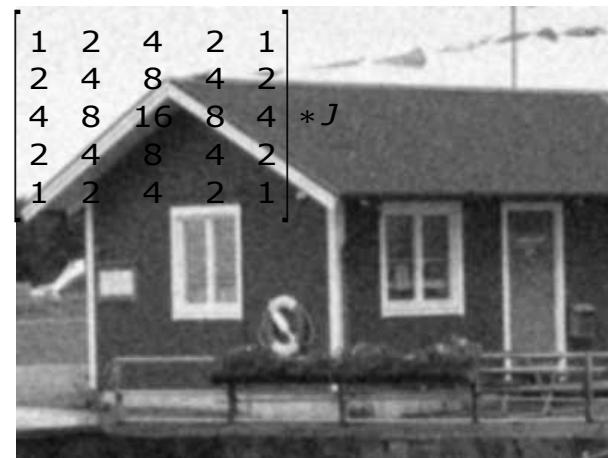
obraz z szumem  $J(x,y)$



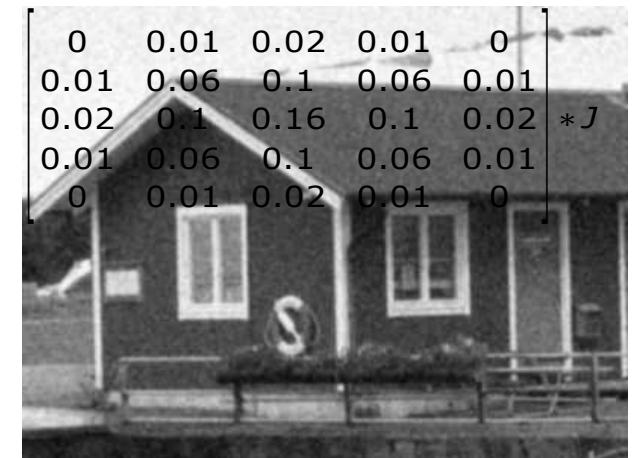
$$\begin{bmatrix} 1 & 2 & 1 \\ 2 & 4 & 2 \\ 1 & 2 & 1 \end{bmatrix} * J$$



$$\begin{bmatrix} 1 & 2 & 4 & 2 & 1 \\ 2 & 4 & 8 & 4 & 2 \\ 4 & 8 & 16 & 8 & 4 \\ 2 & 4 & 8 & 4 & 2 \\ 1 & 2 & 4 & 2 & 1 \end{bmatrix} * J$$



$$\begin{bmatrix} 0 & 0.01 & 0.02 & 0.01 & 0 \\ 0.01 & 0.06 & 0.1 & 0.06 & 0.01 \\ 0.02 & 0.1 & 0.16 & 0.1 & 0.02 \\ 0.01 & 0.06 & 0.1 & 0.06 & 0.01 \\ 0 & 0.01 & 0.02 & 0.01 & 0 \end{bmatrix} * J$$





## Maska jednolita kontra maska Gaussa

obraz oryginalny  $I(x,y)$



obraz z szumem  $J(x,y)$



Filtr z maską jednolitą



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

Filtr z maską Gaussa



$$\begin{bmatrix} 0 & 0.01 & 0.02 & 0.01 & 0 \\ 0.01 & 0.06 & 0.1 & 0.06 & 0.01 \\ 0.02 & 0.1 & 0.16 & 0.1 & 0.02 \\ 0.01 & 0.06 & 0.1 & 0.06 & 0.01 \\ 0 & 0.01 & 0.02 & 0.01 & 0 \end{bmatrix}$$

- filtr Gaussa lepiej zachowuje krawędzie i detale ale mniej skutecznie usuwa szum



## Rozmycie kierunkowe



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



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



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



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



## Wyostrzanie obrazów

obraz oryginalny  $I(x,y)$



$I * F$



$I * F * F$



$I * F * F * F$



- filtr wydobywa różnice w obrazie i dodaje je do obrazu oryginalnego (dodanie Laplasjanu do obrazu)

$$F = \begin{bmatrix} 0 & -1 & 0 \\ -1 & 5 & -1 \\ 0 & -1 & 0 \end{bmatrix}$$

$$F' = \begin{bmatrix} -1 & -1 & -1 \\ -1 & 9 & -1 \\ -1 & -1 & -1 \end{bmatrix}$$

- wielokrotne stosowanie filtra wzmacnia szumy obrazu



## Nakładanie przesuniętych obrazów



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

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



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

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





obraz oryginalny  $I(x,y)$

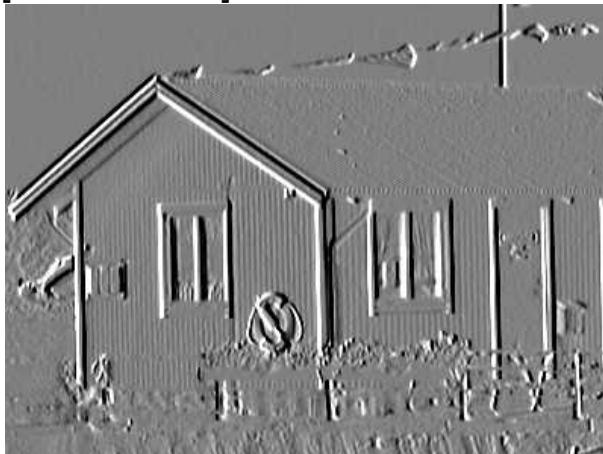


## Obliczanie gradientu

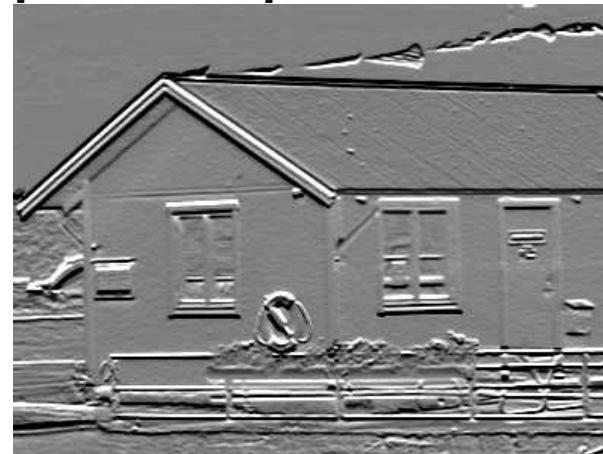
- operacja matematyczna gradientu
- gradient to pole wektorowe
- każda maska oblicza jeden gradient kierunkowy
- wartości wynikowe przesunięte o 128:
  - kolor czarny (0) – wartości ujemne
  - kolor szary (128) – zero
  - kolor biały (255) – wartości dodatnie

$$\nabla I = \left[ \frac{\partial I}{\partial x}, \frac{\partial I}{\partial y} \right] \approx \left[ \frac{I(x+h, y) - I(x, y)}{h}, \frac{I(x, y+h) - I(x, y)}{h} \right]$$

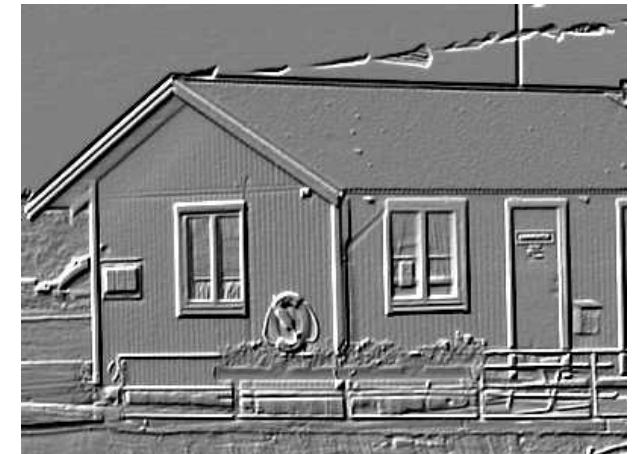
$$\begin{bmatrix} -1 & 0 & +1 \\ -1 & 0 & +1 \\ -1 & 0 & +1 \end{bmatrix}, \quad \nabla_{\bar{x}} I = \left[ \frac{\partial I}{\partial x}, 0 \right]$$



$$\begin{bmatrix} +1 & +1 & +1 \\ 0 & 0 & 0 \\ -1 & -1 & -1 \end{bmatrix}, \quad \nabla_{\bar{y}}(I) = [0, \frac{\partial I}{\partial y}]$$



$$\begin{bmatrix} -1 & -1 & 0 \\ -1 & 0 & +1 \\ 0 & +1 & +1 \end{bmatrix}, \quad \nabla_{[1, -1]} I$$

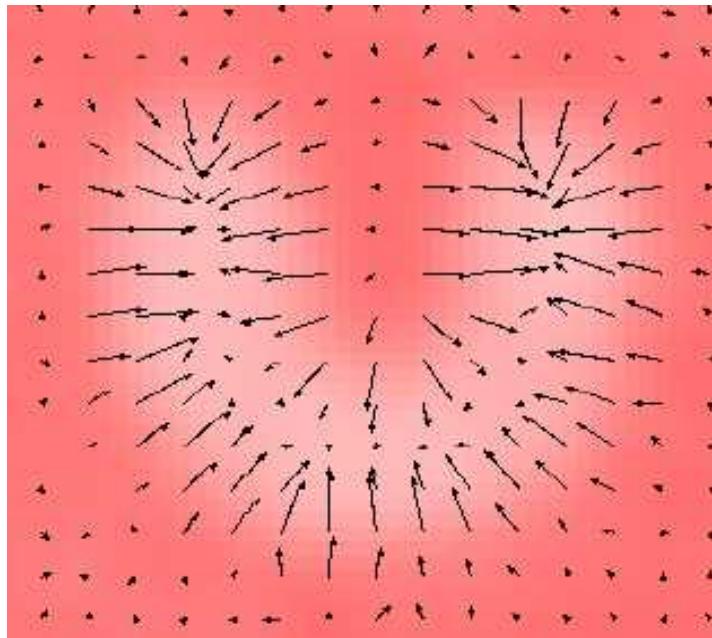




## Gradient jako pole wektorowe

$$\nabla I = \left[ \frac{\partial I}{\partial x}, \frac{\partial I}{\partial y} \right]$$

$$|\nabla I| = \sqrt{\left( \frac{\partial I}{\partial x} \right)^2 + \left( \frac{\partial I}{\partial y} \right)^2}$$



- pole wektorowe gradientu
- nałożone na obraz oryginalny
- wyliczone przez 2 konwolucje
- pokazuje kierunek wzrostu jasności



- moduł gradientu
- pole skalarne
- operacja nieliniowa !!!
- wyliczona przy pomocy 2 konwolucji



## Obliczanie laplasjanu

$$\Delta I = \frac{\partial^2 I}{\partial x^2} + \frac{\partial^2 I}{\partial y^2}$$

$$\Delta I \approx \frac{I(x+h, y) - 2 \cdot I(x, y) + I(x-h, y)}{h^2} + \\ \frac{I(x, y+h) - 2 \cdot I(x, y) + I(x, y-h)}{h^2}$$

- druga pochodna w 2D
- pole skalarne
- przybliżane przez ilorazy różnicowe
- wartości wynikowe przesunięte o 128:
  - kolor czarny (0) – wartości ujemne
  - kolor szary (128) – zero
  - kolor biały (255) – wartości dodatnie
- wykorzystanie:
  - wykrywania linii
  - wykrywanie „plam” w różnej skali  
(ang. *blob detection*)

$$\begin{bmatrix} 0 & -1 & 0 \\ -1 & 4 & -1 \\ 0 & -1 & 0 \end{bmatrix} \begin{bmatrix} 0 & 1 & 0 \\ 1 & -4 & 1 \\ 0 & 1 & 0 \end{bmatrix}$$

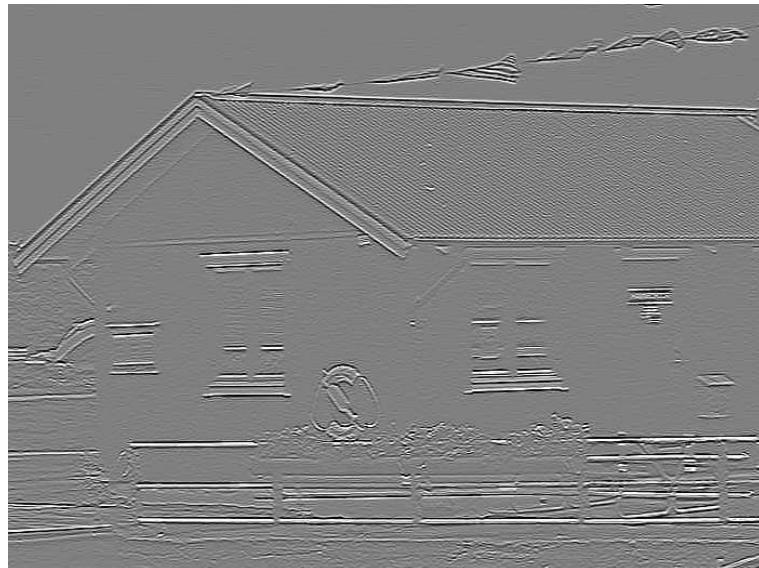
obraz oryginalny  $I(x,y)$



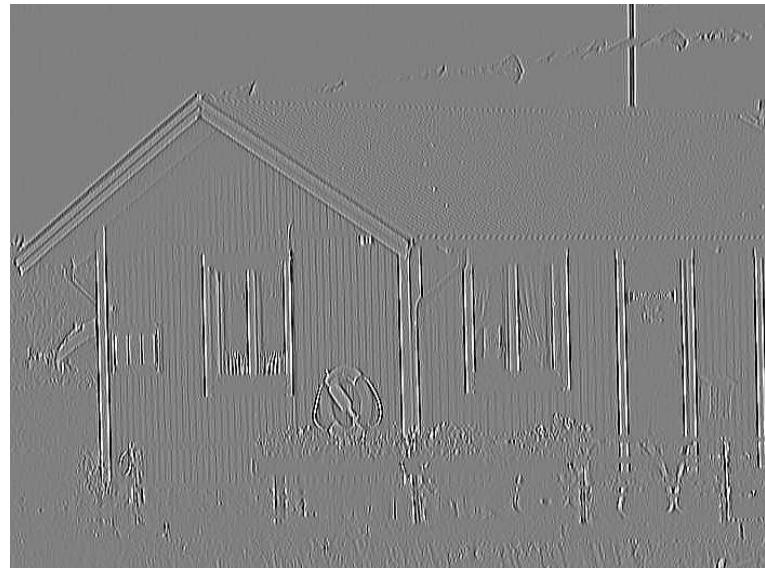


## Wykrywanie linii

$$\begin{bmatrix} -1 & -1 & -1 \\ 2 & 2 & 2 \\ -1 & -1 & -1 \end{bmatrix}$$



$$\begin{bmatrix} -1 & 2 & -1 \\ -1 & 2 & -1 \\ -1 & 2 & -1 \end{bmatrix}$$



- każda maska wrażliwa na jeden kierunek linii – możliwe inne maski w różnych kierunkach
- wrażliwość na grubość linii – potrzeba różnych rozmiarów masek
- wrażliwość na kolor linii – można wziąć wartość bezwzględną, ale to już nie będzie filtr liniowy

$$\begin{bmatrix} 2 & -1 & -1 \\ -1 & 2 & -1 \\ -1 & -1 & 2 \end{bmatrix} \quad \begin{bmatrix} 1 & 1 & -2 \\ 1 & -2 & 1 \\ -2 & 1 & 1 \end{bmatrix}$$





## Filtracja nieliniowa obrazów

$$J(x, y) = f(I(x+i, y+j))$$

gdzie:  $i = -m, \dots, m$  oraz  $j = -n, \dots, n$

$(2*m+1, 2*n+1)$  - rozpatrywane otoczenie - kontekst

$f$  - nieliniowa funkcja

- przykład

- moduł gradientu  $|\nabla I|$



- moduł laplasjanu  $|\Delta I|$





## Filtr medianowy

- funkcja nieliniowa  $f$  jest medianą z argumentów: funkcją sortującą swoje argumenty i wybierającą środkowy spośród nich
- ważna zaleta – wynik jest zawsze jednym z pikseli występujących w obrazie, co może być ważne w przypadku obrazów indeksowanych
- filtr medianowy jest często używany do usuwania drobnych zakłóceń, np szumu typu „salt and pepper”
- filtr ten nie powoduje rozmycia krawędzi
- jeśli sąsiedztwo brane pod uwagę wzrasta, filtrowanie może być wolne

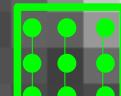
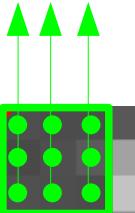
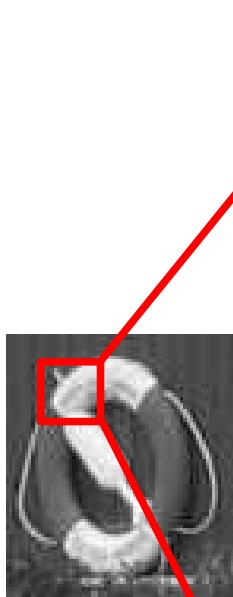


## Obliczanie mediany

$$\begin{bmatrix} 79 & 77 & 76 \\ 73 & \underline{\mathbf{79}} & 90 \\ 80 & 79 & 90 \end{bmatrix}$$



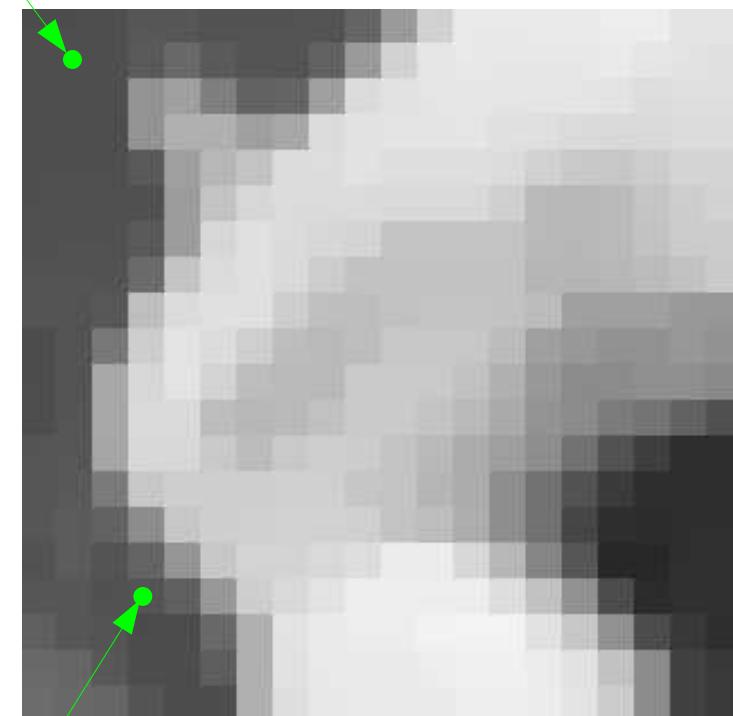
$$[73 \ 76 \ 77 \ 79 \ \underline{\mathbf{79}} \ 79 \ 80 \ 90 \ 90]$$



$$\begin{bmatrix} 98 & 96 & 152 \\ 84 & \underline{\mathbf{63}} & 105 \\ 75 & 65 & 75 \end{bmatrix}$$



$$[63 \ 65 \ 75 \ 75 \ \underline{\mathbf{84}} \ 96 \ 98 \ 105 \ 152]$$





## Efekty filtru medianowego

obraz oryginalny



filtr medianowy 3x3



filtr medianowy 5x5



filtr medianowy 11x11



filtr medianowy 21x21



- filtr usuwa detale o rozmiarze proporcjonalnym do promienia jego sąsiedztwa, zatem jest filtrem dolnoprzepustowym
- filtr nie rozmywa krawędzi
- filtr tworzy obszary o stałej jasności, bez faktury



## Filtr medianowy w usuwaniu szumu

obraz oryginalny



obraz z szumem „salt&pepper”



filtr medianowy 3x3



obraz z szumem białym



filtr medianowy 3x3



- filtr medianowy dobrze usuwa szum „salt&pepper”, gdzie tylko nieliczne piksele są zmienione
- gorzej radzi sobie z szumem białym gdzie wszystkie piksele są zmienione

## Konserwatywne wygładzanie *(conservative smoothing)*



- filtr nieliniowy przeznaczony głównie do usuwania szumu impulsowego
- w minimalnym stopniu zniekształca obraz
- wyszukuje lokalne, odizolowane piksele z wartością mocno odbiegającą od otoczenia i tylko je usuwa:
  - dla każdego sąsiedztwa piksela  $I(x,y)$  liczone są wartości *max* i *min* (bez brania pod uwagę tego piksela)
  - jeśli  $I(x,y)$  należy do przedziału  $\langle \min, \max \rangle$ , jego wartość nie jest zmieniana

$$\begin{bmatrix} 79 & 77 & 76 \\ 73 & \underline{\textbf{81}} & 90 \\ 80 & 79 & 90 \end{bmatrix} \rightarrow \begin{bmatrix} 79 & 77 & 76 \\ 73 & \underline{\textbf{81}} & 90 \\ 80 & 79 & 90 \end{bmatrix}$$

- jeśli  $I(x,y) > \max$  wtedy  $I(x,y) := \max$

$$\begin{bmatrix} 79 & 77 & 76 \\ 73 & \underline{\textbf{204}} & 90 \\ 80 & 79 & 90 \end{bmatrix} \rightarrow \begin{bmatrix} 79 & 77 & 76 \\ 73 & \underline{\textbf{90}} & 90 \\ 80 & 79 & 90 \end{bmatrix}$$

- jeśli  $I(x,y) < \min$  wtedy  $I(x,y) := \min$

$$\begin{bmatrix} 79 & 77 & 76 \\ 73 & \underline{\textbf{15}} & 90 \\ 80 & 79 & 90 \end{bmatrix} \rightarrow \begin{bmatrix} 79 & 77 & 76 \\ 73 & \underline{\textbf{73}} & 90 \\ 80 & 79 & 90 \end{bmatrix}$$

- filtr szybszy od medianowego – wyliczenie *min* i *max* jest szybsze niż posortowanie zakresu



## Konserwatywne wygładzanie *(conservative smoothing)*



- jeśli impulsy szumu są zbyt bliskie siebie, nie zostaną usunięte
- w zamian za to detale nie są wygładzane (jak to się dzieje w filtrze medianowym) →





## Filtr wartości środkowej (*midpoint*)

- wartość piksela  $I(x,y)$  jest zastępowana przez średnią z wartości maksymalnej i minimalnej z jego otoczenia:

$$J(x, y) = [\max(I(x-m : x+m, y-n : y+n)) + \min(I(x-m : x+m, y-n : y+n))] / 2$$

- stosowana głównie do białego szumu o rozkładzie normalnym lub jednorodnym





## Filtr uśredniający z pomijaniem $\alpha$ wartości skrajnych

- przyjmując kwadratowe sąsiedztwo  $n \times n$  ( $n$  nieparzyste) sortuje się wartości pikseli z tego sąsiedztwa:

$$[I_1, I_2, \dots, I_{n^2-1}, I_{n^2}]$$

- z tego zakresu usuwa się  $\alpha$  skrajnych (minimalnych i maksymalnych) wartości
- pozostałe wartości uśrednia się:

$$J(x, y) = \frac{1}{n^2 - 2\cdot\alpha} \cdot \sum_{i=\alpha+1}^{n^2-\alpha} I_i$$

- dla  $\alpha=0$  filtr jest filtrem uśredniającym o jednolitej masce
- dla  $\alpha = (n^2-1)/2$  filtr jest filtrem medianowym

obraz z szumem białym po filtrowaniu,  $\alpha=2$       obraz z szumem impulsowym po filtrowaniu,  $\alpha=2$





## Filtr uśredniający adaptacyjny

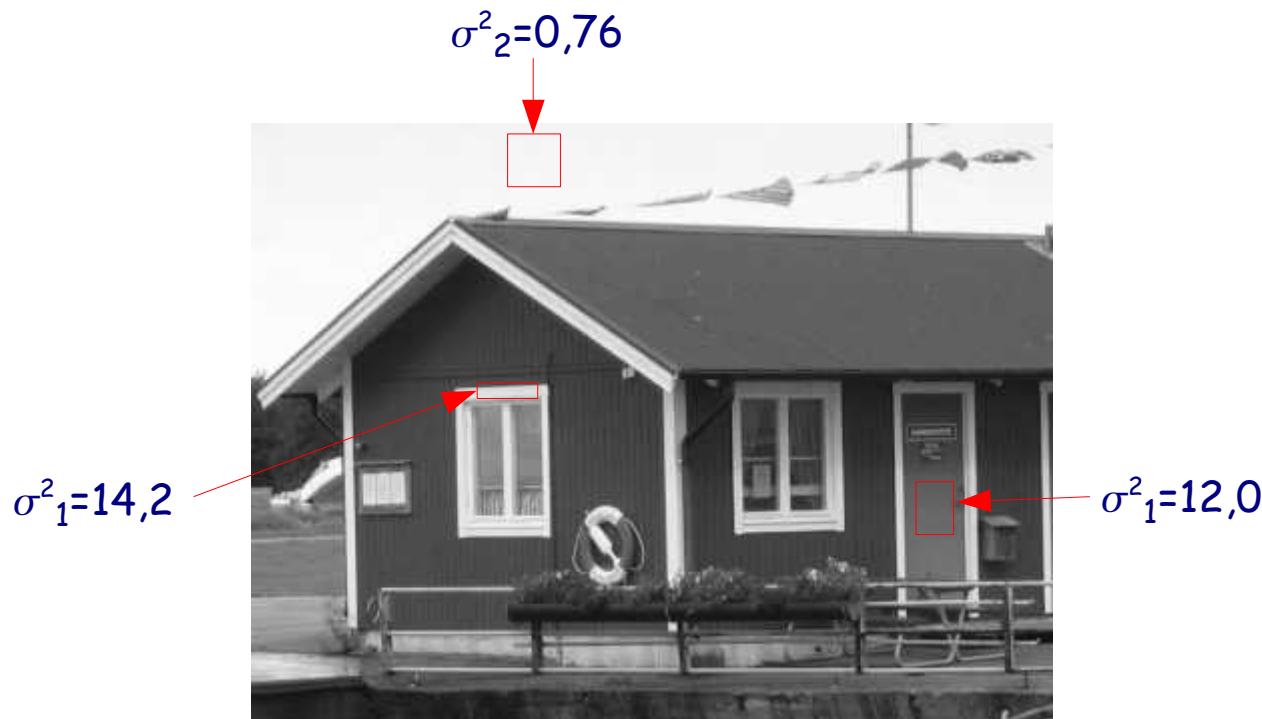
$$J(x, y) = I(x, y) - \frac{\sigma^2}{\sigma^2(x, y)} \cdot (I(x, y) - \text{mean}(x, y))$$

- $\sigma^2$  – wariancja szumu obrazu
- $\sigma^2(x, y)$  – wariancja w otoczeniu piksela  $I(x, y)$
- $\text{mean}(x, y)$  – średnia intensywność w otoczeniu piksela  $I(x, y)$
- dla obszarów jednolitych, bez detali,  $\sigma^2 = \sigma^2(x, y)$  i  $J(x, y) = \text{mean}(x, y)$  – obraz jest uśredniany
- dla obszarów ze szczegółami  $\sigma^2(x, y) \gg \sigma^2$  i  $J(x, y) = I(x, y)$  – obraz oryginalny się nie zmienia



## Eksperymentalne wyliczanie wariancji szumu

- jako wariancję szumu przyjmuje się wariancję w obszarach uznanych za jednolite





## Filtr uśredniający adaptacyjny

obraz przefiltrowany z wyliczoną wariancją i sąsiedztwem  $7 \times 7$



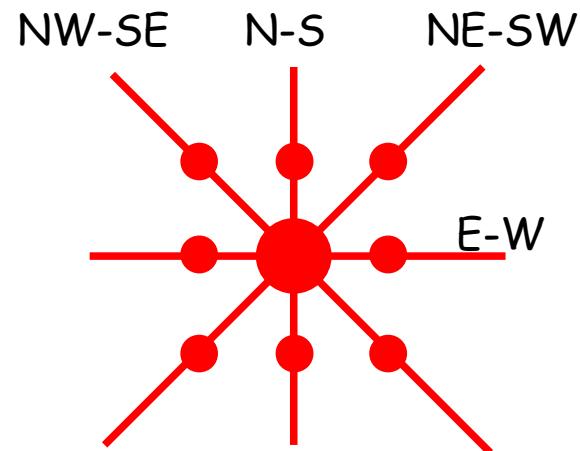
- uśrednianie odbywa się w obszarach „jednolitych”
- okolice krawędzi nie są filtrowane – nie ulegają one rozmyciu jak w uśrednianiu
- filtr kosztowny obliczeniowo





## Filtr odplamiający Crimmins'a (Crimmins speckle removal )

- nieliniowy filtr iteracyjny **stopniowo** niwelujący lokalne odstępstwa intensywności pikseli od intensywności otoczenia
- przeznaczony do usuwania szumu impulsowego
- rodzaj filtra dolnoprzepustowego – stopniowo usuwa drobne detale
  
- dla każdego piksela jego intensywność jest porównywana z 2 sąsiadami na jednym z 4 kierunków
- jeżeli od nich odbiega, jest modyfikowana:
  - jest zmniejszana jeśli jest większa
  - jest zwiększana jeśli jest mniejsza
- analizowane są po kolej po wszystkie kierunki
- algorytm jest złożony obliczeniowo !!!





## Algorytm filtru Crimmins'a

pojedyncza iteracja filtrowania:

### 1. Rozjaśnianie pikseli (usuwanie szumu „pepper”)

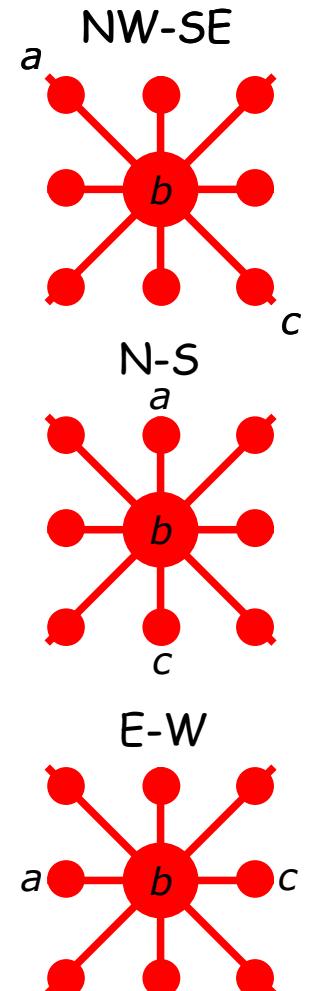
dla każdego z kierunków NW-SE, N-S, NE-SW, E-W:

- oznacz sąsiednie piksele na kierunku jako  $a$ ,  $b$ ,  $c$
- dla wszystkich trypletów  $(a, b, c)$  w obrazie:
  - jeśli  $(a \geq b+2) \Rightarrow b+=1;$
- dla wszystkich trypletów  $(a, b, c)$  w obrazie:
  - jeśli  $((a > b) \text{ i } (b \leq c)) \Rightarrow b+=1;$
- dla wszystkich trypletów  $(a, b, c)$  w obrazie:
  - jeśli  $((c > b) \text{ i } (b \leq a)) \Rightarrow b+=1;$
- dla wszystkich trypletów  $(a, b, c)$  w obrazie:
  - jeśli  $(c \geq b+2) \Rightarrow b+=1;$

### 2. Wygaszanie pikseli (usuwanie szumu „salt”)

dla każdego z kierunków NW-SE, N-S, NE-SW, E-W:

- oznacz sąsiednie piksele na kierunku jako  $a$ ,  $b$ ,  $c$
- dla wszystkich trypletów  $(a, b, c)$  w obrazie:
  - jeśli  $(a \leq b-2) \Rightarrow b-=1;$
- dla wszystkich trypletów  $(a, b, c)$  w obrazie:
  - jeśli  $((a < b) \text{ i } (b \geq c)) \Rightarrow b-=1;$
- dla wszystkich trypletów  $(a, b, c)$  w obrazie:
  - jeśli  $((c < b) \text{ i } (b \geq a)) \Rightarrow b-=1;$
- dla wszystkich trypletów  $(a, b, c)$  w obrazie:
  - jeśli  $(c \leq b-2) \Rightarrow b-=1;$





## Wyniki filtru Crimmins'a

obraz oryginalny z szumem impulsowym obraz po 1 iteracji



obraz po 3 iteracjach



obraz po 5 iteracjach



obraz po 10 iteracjach



- zmiany zachodzą stopniowo
- detale są usuwane
- krawędzie pozostają ostre
- szum zanika
- filtr bardzo wolny  
(1 iteracja - kilka minut)



## Inne filtry nieliniowe

- zwykłe filtry specjalizowane wykrywające pewne rodzaje struktur:
  - krawędzie
  - zakończenia krawędzi
  - struktury kuliste
  - rogi
  - struktury liniowe
- przykład – wykrywanie naczyń krwionośnych 3D (struktur liniowych) w oparciu o wartości momentów geometrycznych wyliczonych w otoczeniu każdego punktu

obraz oryginalny, naczynia  
zasłonięte przez kości



różne wersje filtrowania z użyciem momentów geometrycznych



K. Wiszowaty, C. Bołdak, C. Toumoulin: Detekcja form liniowych w trójwymiarowych obrazach medycznych CTA przy użyciu momentów geometrycznych. XV KBIB, 2007.

# GNU Free Documentation License



Version 1.2, November 2002

Copyright (C) 2000,2001,2002 Free Software Foundation, Inc.  
51 Franklin St, Fifth Floor, Boston, MA 02110-1301 USA  
Everyone is permitted to copy and distribute verbatim copies  
of this license document, but changing it is not allowed.

## 0. PREAMBLE

The purpose of this License is to make a manual, textbook, or other functional and useful document "free" in the sense of freedom: to assure everyone the effective freedom to copy and redistribute it, with or without modifying it, either commercially or noncommercially. Secondly, this License preserves for the author and publisher a way to get credit for their work, while not being considered responsible for modifications made by others.

This License is a kind of "copyleft", which means that derivative works of the document must themselves be free in the same sense. It complements the GNU General Public License, which is a copyleft license designed for free software.

We have designed this License in order to use it for manuals for free software, because free software needs free documentation: a free program should come with manuals providing the same freedoms that the software does. But this License is not limited to software manuals; it can be used for any textual work, regardless of subject matter or whether it is published as a printed book. We recommend this License principally for works whose purpose is instruction or reference.

## 1. APPLICABILITY AND DEFINITIONS

This License applies to any manual or other work, in any medium, that contains a notice placed by the copyright holder saying it can be distributed under the terms of this License. Such a notice grants a world-wide, royalty-free license, unlimited in duration, to use that work under the conditions stated herein. The "Document", below, refers to any such manual or work. Any member of the public is a licensee, and is addressed as "you". You accept the license if you copy, modify or distribute the work in a way requiring permission under copyright law.

A "Modified Version" of the Document means any work containing the Document or a portion of it, either copied verbatim, or with modifications and/or translated into another language.

A "Secondary Section" is a named appendix or a front-matter section of the Document that deals exclusively with the relationship of the publishers or authors of the Document to the Document's overall subject (or to related matters) and contains nothing that could fall directly within that overall subject. (Thus, if the Document is in part a textbook of mathematics, a Secondary Section may not explain any mathematics.) The relationship could be a matter of historical connection with the subject or with related matters, or of legal, commercial, philosophical, ethical or political position regarding them.

The "Invariant Sections" are certain Secondary Sections whose titles are designated, as being those of Invariant Sections, in the notice that says that the Document is released under this License. If a section does not fit the above definition of Secondary then it is not allowed to be designated as Invariant. The Document may contain zero Invariant Sections. If the Document does not identify any Invariant Sections then there are none.

The "Cover Texts" are certain short passages of text that are listed, as Front-Cover Texts or Back-Cover Texts, in the notice that says that the Document is released under this License. A Front-Cover Text may be at most 5 words, and a Back-Cover Text may be at most 25 words.

A "Transparent" copy of the Document means a machine-readable copy, represented in a format whose specification is available to the general public, that is suitable for revising the document straightforwardly with generic text editors or (for images composed of pixels) generic paint programs or (for drawings) some widely available drawing editor, and that is suitable for input to text formatters or for automatic translation to a variety of formats suitable for input to text formatters. A copy made in an otherwise Transparent file format whose markup, or absence of markup, has been arranged to thwart or discourage subsequent modification by readers is not Transparent. An image format is not Transparent if used for any substantial amount of text. A copy that is not "Transparent" is called "Opaque".

Examples of suitable formats for Transparent copies include plain ASCII without markup, Texinfo input format, LaTeX input format, SGML or XML using a publicly available DTD, and standard-conforming simple HTML, PostScript or PDF designed for human modification. Examples of transparent image formats include PNG, XCF and JPG. Opaque formats include proprietary formats that can be read and edited only by proprietary word processors, SGML or XML for which the DTD and/or processing tools are not generally available, and the machine-generated HTML, PostScript or PDF produced by some word processors for output purposes only.

The "Title Page" means, for a printed book, the title page itself, plus such following pages as are needed to hold, legibly, the material this License requires to appear in the title page. For works in formats which do not have any title page as such, "Title Page" means the text near the most prominent appearance of the work's title, preceding the beginning of the body of the text.

A section "Entitled XYZ" means a named subunit of the Document whose title either is precisely XYZ or contains XYZ in parentheses following text that translates XYZ in another language. (Here XYZ stands for a specific section name mentioned below, such as "Acknowledgements", "Dedications", "Endorsements", or "History".) To "Preserve the Title" of such a section when you modify the Document means that it remains a section "Entitled XYZ" according to this definition.

The Document may include Warranty Disclaimers next to the notice which states that this License applies to the Document. These Warranty Disclaimers are considered to be included by reference in this License, but only as regards disclaiming warranties: any other implication that these Warranty Disclaimers may have is void and has no effect on the meaning of this License.

## 2. VERBATIM COPYING

You may copy and distribute the Document in any medium, either commercially or noncommercially, provided that this License, the copyright notices, and the license notice saying this License applies to the Document are reproduced in all copies, and that you add no other conditions whatsoever to those of this License. You may not use technical measures to obstruct or control the reading or further copying of the copies you make or distribute. However, you may accept compensation in exchange for copies. If you distribute a large enough number of copies you must also follow the conditions in section 3.

You may also lend copies, under the same conditions stated above, and you may publicly display copies.



### 3. COPYING IN QUANTITY

If you publish printed copies (or copies in media that commonly have printed covers) of the Document, numbering more than 100, and the Document's license notice requires Cover Texts, you must enclose the copies in covers that carry, clearly and legibly, all these Cover Texts: Front-Cover Texts on the front cover, and Back-Cover Texts on the back cover. Both covers must also clearly and legibly identify you as the publisher of these copies. The front cover must present the full title with all words of the title equally prominent and visible. You may add other material on the covers in addition. Copying with changes limited to the covers, as long as they preserve the title of the Document and satisfy these conditions, can be treated as verbatim copying in other respects.

If the required texts for either cover are too voluminous to fit legibly, you should put the first ones listed (as many as fit reasonably) on the actual cover, and continue the rest onto adjacent pages.

If you publish or distribute Opaque copies of the Document numbering more than 100, you must either include a machine-readable Transparent copy along with each Opaque copy, or state in or with each Opaque copy a computer-network location from which the general network-using public has access to download using public-standard network protocols a complete Transparent copy of the Document, free of added material. If you use the latter option, you must take reasonably prudent steps, when you begin distribution of Opaque copies in quantity, to ensure that this Transparent copy will remain thus accessible at the stated location until at least one year after the last time you distribute an Opaque copy (directly or through your agents or retailers) of that edition to the public.

It is requested, but not required, that you contact the authors of the Document well before redistributing any large number of copies, to give them a chance to provide you with an updated version of the Document.

### 4. MODIFICATIONS

You may copy and distribute a Modified Version of the Document under the conditions of sections 2 and 3 above, provided that you release the Modified Version under precisely this License, with the Modified Version filling the role of the Document, thus licensing distribution and modification of the Modified Version to whoever possesses a copy of it. In addition, you must do these things in the Modified Version:

- \* A. Use in the Title Page (and on the covers, if any) a title distinct from that of the Document, and from those of previous versions (which should, if there were any, be listed in the History section of the Document). You may use the same title as a previous version if the original publisher of that version gives permission.
- \* B. List on the Title Page, as authors, one or more persons or entities responsible for authorship of the modifications in the Modified Version, together with at least five of the principal authors of the Document (all of its principal authors, if it has fewer than five), unless they release you from this requirement.
- \* C. State on the Title page the name of the publisher of the Modified Version, as the publisher.
- \* D. Preserve all the copyright notices of the Document.
- \* E. Add an appropriate copyright notice for your modifications adjacent to the other copyright notices.
- \* F. Include, immediately after the copyright notices, a license notice giving the public permission to use the Modified Version under the terms of this License, in the form shown in the Addendum below.
- \* G. Preserve in that license notice the full lists of Invariant Sections and required Cover Texts given in the Document's license notice.
- \* H. Include an unaltered copy of this License.
- \* I. Preserve the section Entitled "History", Preserve its Title, and add to it an item stating at least the title, year, new authors, and publisher of the Modified Version as given on the Title Page. If there is no section Entitled "History" in the Document, create one stating the title, year, authors, and publisher of the Document as given on its Title Page, then add an item describing the Modified Version as stated in the previous sentence.
- \* J. Preserve the network location, if any, given in the Document for public access to a Transparent copy of the Document, and likewise the network locations given in the Document for previous versions it was based on. These may be placed in the "History" section. You may omit a network location for a work that was published at least four years before the Document itself, or if the original publisher of the version it refers to gives permission.
- \* K. For any section Entitled "Acknowledgements" or "Dedications", Preserve the Title of the section, and preserve in the section all the substance and tone of each of the contributor acknowledgements and/or dedications given therein.
- \* L. Preserve all the Invariant Sections of the Document, unaltered in their text and in their titles. Section numbers or the equivalent are not considered part of the section titles.
- \* M. Delete any section Entitled "Endorsements". Such a section may not be included in the Modified Version.
- \* N. Do not retitle any existing section to be Entitled "Endorsements" or to conflict in title with any Invariant Section.
- \* O. Preserve any Warranty Disclaimers.

If the Modified Version includes new front-matter sections or appendices that qualify as Secondary Sections and contain no material copied from the Document, you may at your option designate some or all of these sections as invariant. To do this, add their titles to the list of Invariant Sections in the Modified Version's license notice. These titles must be distinct from any other section titles.

You may add a section Entitled "Endorsements", provided it contains nothing but endorsements of your Modified Version by various parties--for example, statements of peer review or that the text has been approved by an organization as the authoritative definition of a standard.

You may add a passage of up to five words as a Front-Cover Text, and a passage of up to 25 words as a Back-Cover Text, to the end of the list of Cover Texts in the Modified Version. Only one passage of Front-Cover Text and one of Back-Cover Text may be added by (or through arrangements made by) any one entity. If the Document already includes a cover text for the same cover, previously added by you or by arrangement made by the same entity you are acting on behalf of, you may not add another; but you may replace the old one, on explicit permission from the previous publisher that added the old one.

The author(s) and publisher(s) of the Document do not by this License give permission to use their names for publicity for or to assert or imply endorsement of any Modified Version.



## 6. COLLECTIONS OF DOCUMENTS

You may make a collection consisting of the Document and other documents released under this License, and replace the individual copies of this License in the various documents with a single copy that is included in the collection, provided that you follow the rules of this License for verbatim copying of each of the documents in all other respects.

You may extract a single document from such a collection, and distribute it individually under this License, provided you insert a copy of this License into the extracted document, and follow this License in all other respects regarding verbatim copying of that document.

## 7. AGGREGATION WITH INDEPENDENT WORKS

A compilation of the Document or its derivatives with other separate and independent documents or works, in or on a volume of a storage or distribution medium, is called an "aggregate" if the copyright resulting from the compilation is not used to limit the legal rights of the compilation's users beyond what the individual works permit. When the Document is included in an aggregate, this License does not apply to the other works in the aggregate which are not themselves derivative works of the Document.

If the Cover Text requirement of section 3 is applicable to these copies of the Document, then if the Document is less than one half of the entire aggregate, the Document's Cover Texts may be placed on covers that bracket the Document within the aggregate, or the electronic equivalent of covers if the Document is in electronic form. Otherwise they must appear on printed covers that bracket the whole aggregate.

## 8. TRANSLATION

Translation is considered a kind of modification, so you may distribute translations of the Document under the terms of section 4. Replacing Invariant Sections with translations requires special permission from their copyright holders, but you may include translations of some or all Invariant Sections in addition to the original versions of these Invariant Sections. You may include a translation of this License, and all the license notices in the Document, and any Warranty Disclaimers, provided that you also include the original English version of this License and the original versions of those notices and disclaimers. In case of a disagreement between the translation and the original version of this License or a notice or disclaimer, the original version will prevail.

If a section in the Document is Entitled "Acknowledgements", "Dedications", or "History", the requirement (section 4) to Preserve its Title (section 1) will typically require changing the actual title.

## 9. TERMINATION

You may not copy, modify, sublicense, or distribute the Document except as expressly provided for under this License. Any other attempt to copy, modify, sublicense or distribute the Document is void, and will automatically terminate your rights under this License. However, parties who have received copies, or rights, from you under this License will not have their licenses terminated so long as such parties remain in full compliance.

## 10. FUTURE REVISIONS OF THIS LICENSE

The Free Software Foundation may publish new, revised versions of the GNU Free Documentation License from time to time. Such new versions will be similar in spirit to the present version, but may differ in detail to address new problems or concerns. See <http://www.gnu.org/copyleft/>.

Each version of the License is given a distinguishing version number. If the Document specifies that a particular numbered version of this License "or any later version" applies to it, you have the option of following the terms and conditions either of that specified version or of any later version that has been published (not as a draft) by the Free Software Foundation. If the Document does not specify a version number of this License, you may choose any version ever published (not as a draft) by the Free Software Foundation.

## 5. COMBINING DOCUMENTS

You may combine the Document with other documents released under this License, under the terms defined in section 4 above for modified versions, provided that you include in the combination all of the Invariant Sections of all of the original documents, unmodified, and list them all as Invariant Sections of your combined work in its license notice, and that you preserve all their Warranty Disclaimers.

The combined work need only contain one copy of this License, and multiple identical Invariant Sections may be replaced with a single copy. If there are multiple Invariant Sections with the same name but different contents, make the title of each such section unique by adding at the end of it, in parentheses, the name of the original author or publisher of that section if known, or else a unique number. Make the same adjustment to the section titles in the list of Invariant Sections in the license notice of the combined work.

In the combination, you must combine any sections Entitled "History" in the various original documents, forming one section Entitled "History"; likewise combine any sections Entitled "Acknowledgements", and any sections Entitled "Dedications". You must delete all sections Entitled "Endorsements."

(not as a draft) by the Free Software Foundation.

# Cyfrowe Przetwarzanie Obrazów

## Digital Image Processing



Aurelio A. Heckert

Copyright (c) 2008 Cezary Bołdak

Permission is granted to copy, distribute and/or modify this document under the terms of the GNU Free Documentation License, Version 1.2 or any later version published by the Free Software Foundation; with no Invariant Sections, no Front-Cover Texts, and no Back-Cover Texts. A copy of the license is included in the section entitled "GNU Free Documentation License"



## Wykład 4

# Wykrywanie cech w obrazach cyfrowych (*feature detection*)

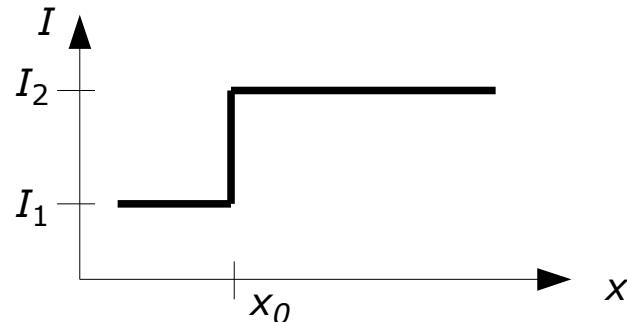


- cechy (ang. *features*) rozumiane są jako specyficzne konfiguracje pikseli układające się w pewne struktury:
  - krawędzie
  - linie
  - zakończenia linii
  - rogi
  - skupiska (blobs) - koła/kule
  - tekstury
  - grzbiety (ridges), szkielety
- zwykle wykrywa się je poprzez filtrowanie obraz
- odpowiedź filtru jest najczęściej liczba rzeczywista
- dobry filtr powinien dawać:
  - wysokie odpowiedzi dla pikseli należących do szukanych struktur
  - niskie odpowiedzi dla pikseli nie należących do szukanych struktur  
(choć mogących należeć do innych struktur)
- filtry mogą być liniowe i nieliniowe



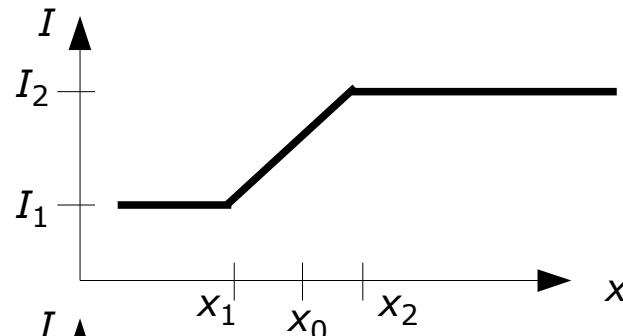
## Pojęcie krawędzi

- krawędź jest to nagła zmiana intensywności (próg)

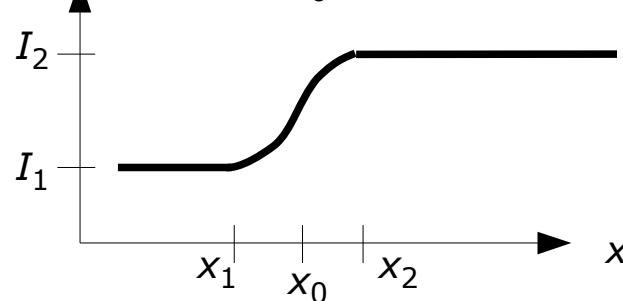


$$I(x) = \begin{cases} I_1 & \text{jeśli } x < x_0 \\ I_2 & \text{jeśli } x \geq x_0 \end{cases}$$

- w rzeczywistych przypadkach zmiana intensywności nigdy nie jest punktowa
  - zawsze istnieje strefa przejściowa, liniowa lub nielinowa



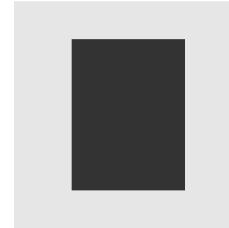
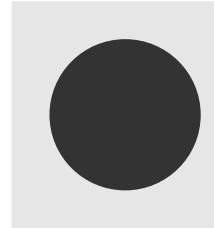
$$I(x) = \begin{cases} I_1 & \text{jeśli } x \leq x_1 \\ I_1 + (x - x_1) \cdot \frac{I_2 - I_1}{x_2 - x_1} & \text{jeśli } x_1 < x < x_2 \\ I_2 & \text{jeśli } x \geq x_2 \end{cases}$$



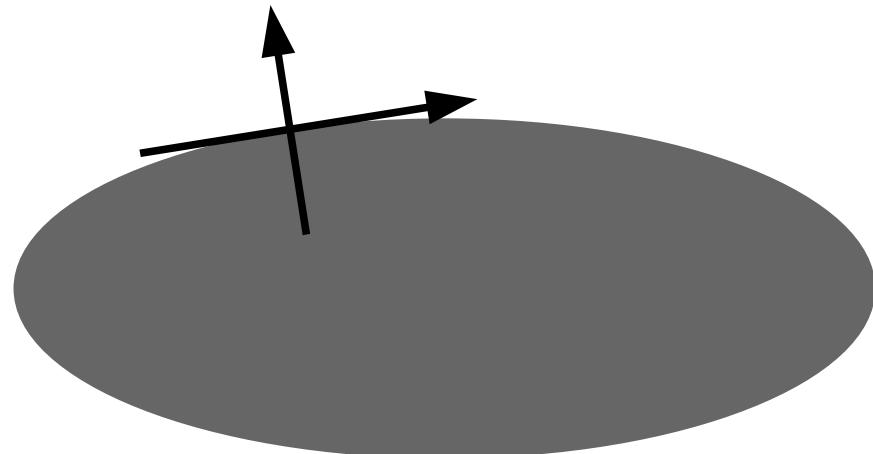


## Krawędź dwuwymiarowa

- występuje na granicy 2 obiektów (lub obiektu i tła)
- może być prostoliniowa i krzywoliniowa



- można wyróżnić 2 ortogonalne kierunki:
  - prostopadły do krawędzi – tu intensywność zmienia się gwałtownie, profil intensywności jak dla krawędzi 1D
  - styczny do krawędzi – tu intensywność lokalnie się nie zmienia, profil intensywności stały jest stały (lokalnie), jak by krawędzi nie było





## Krawędź dwuwymiarowa dyskretna

krawędź pionowa idealna

50	50	50	100	100	100	100
50	50	50	100	100	100	100
50	50	50	100	100	100	100
50	50	50	100	100	100	100

krawędź pionowa ze strefą przejściową

50	50	60	<u>75</u>	90	100	100
50	50	60	<u>75</u>	90	100	100
50	50	60	<u>75</u>	90	100	100
50	50	60	<u>75</u>	90	100	100

krawędź pozioma idealna

50	50	50	50	50	50	50
50	50	50	50	50	50	50
100	100	100	100	100	100	100
100	100	100	100	100	100	100

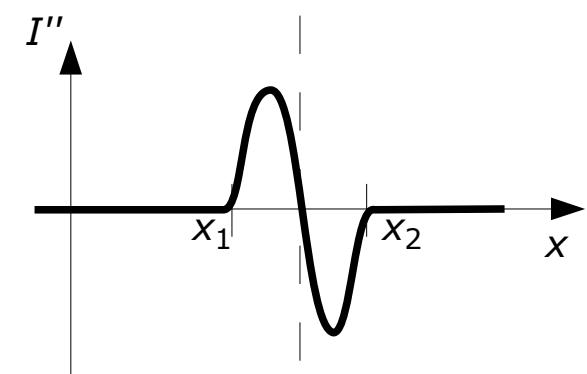
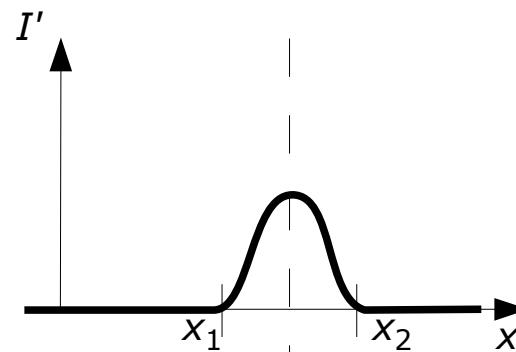
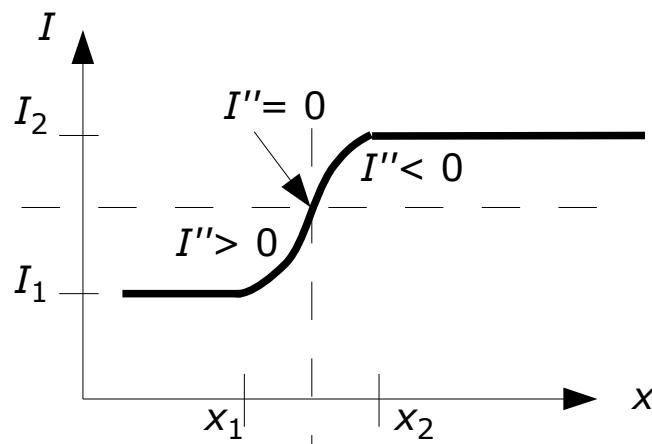
krawędź skośna ze strefą przejściową

50	50	<b>55</b>	60	<u>75</u>	90	95
<b>50</b>	55	60	<u>75</u>	90	95	<b>100</b>
55	60	<u>75</u>	90	95	<b>100</b>	100
60	<u>75</u>	90	95	<b>100</b>	100	100



## Wykrywanie krawędzi

- należy wykryć nagłą zmianę intensywności
- szum może mieć podobną właściwość, więc zmiana intensywności nie może być izolowana
- zmiana intensywności dotyczy jednego kierunku, kierunek prostopadły lokalnie nie wykazuje zmiany intensywności
- trudności z lokalizacją krawędzi – powinna być ona precyjnie umieszczona pomimo strefy przejściowej (zbocza w profilu intensywności):
  - geometryczny środek zbocza
  - przekroczenie wartości średniej lub np 30% różnicy  $I_2 - I_1$
  - punkt przegięcia (maksimum 1. pochodnej, zmiana znaku 2.pochodnej)





## Operator krzyżowy Robertsa *(Roberts Cross operator)*

- jeden z pierwszych detektorów krawędzi
- wyznacza krawędź jako maksimum 1. pochodnej 2D – gradientu
- gradient jest wektorem, który jest aproksymowany przez różnice skończone:

$$\nabla_x I(x, y) = \frac{I(x+h, y) - I(x, y)}{h}, \quad \nabla_y I(x, y) = \frac{I(x, y+h) - I(x, y)}{h}$$

- w tym operatorze uwzględniane są gradienty kierunkowe ukośne:

$$\nabla_{xy} I(x, y) = \frac{I(x+h, y+h) - I(x, y)}{\sqrt{2}h}, \quad \nabla_{\perp xy} I(x, y) = \frac{I(x, y+h) - I(x+h, y)}{\sqrt{2}h}$$

- mianowniki zwykle pomija się
- filtr ten uzyskuje się przez konwolucję z 2 maskami 2x2:

$$G_1 = \begin{bmatrix} 0 & +1 \\ -1 & 0 \end{bmatrix} * I \quad G_2 = \begin{bmatrix} +1 & 0 \\ 0 & -1 \end{bmatrix} * I$$

- następnie pole wektorowe zamienia się na skalar za pomocą modułu wektora gradientu:

$$\sqrt{G_1^2 + G_2^2}$$

lub za pomocą szybszej do policzenia sumy wartości bezwzględnej

$$|G_1| + |G_2|$$



## Operator krzyżowy Robertsa



- operator ten jest dość wrażliwy na szum
- jest to skutek jego lokalności – brane są pod uwagę jasność jedynie pojedynczych pikseli (jądro konwolucji jest małe)
- mimo tego często używany ze względu na prostotę i szybkość obliczania, szczególnie w wersji z wartościami bezwzględnymi (jądro konwolucji jest małe)
- czasami wizualizowane są oddzielnie wartości  $G_1$  i  $G_2$



## Operator Sobela

- kolejna wersja operatora który widzi krawędź jako maksimum gradientu
- gradienty poziomy i pionowy wyznaczane są przez różnice centralne

$$\nabla_x I(x, y) = \frac{I(x+h, y) - I(x-h, y)}{2h}, \quad \nabla_y I(x, y) = \frac{I(x, y+h) - I(x, y-h)}{2h}$$

- aby zmniejszyć wpływ szumu **uśrednia** się wartości w pionie (przy gradiencie poziomym) i w poziomie (przy gradiencie pionowym) z większymi wagami przyłożonymi dla kierunku przechodzącego przez analizowany punkt
- tu również można zaniedbać mianowniki i liczyć gradientu jako konwolucję

$$G_x = \begin{bmatrix} -1 & 0 & +1 \\ -2 & 0 & +2 \\ -1 & 0 & +1 \end{bmatrix} * I \quad G_y = \begin{bmatrix} +1 & +2 & +1 \\ 0 & 0 & 0 \\ -1 & -2 & -1 \end{bmatrix} * I$$

- wartość filtru można liczyć jak moduł lub sumę wartości bezwzględnych

$$\sqrt{G_1^2 + G_2^2} \quad |G_1| + |G_2|$$

- operator tej jest o wiele mniej wrażliwy na szum niż operator Roberts (wpływ różnic centralnych – większe maski konwolucji – i uśredniania)

## Operator Sobela



- mniejsza wrażliwość na szum niż operator Robertsa →
- krawędzie nadal są dość grube  
(nie są precyzyjnie zlokalizowane)  
i mają różną intensywność





## Kąt gradientu

$$\alpha = \tan^{-1} \frac{G_y}{G_x}$$





## Kierunek gradientu

- zaniedbując zwrot gradientu:





## Operator kompasowy

- wynik – **indeks** tej maski  $M_i$  z którą konwolucja da maksymalną odpowiedź
- nie ma problemów dla kąta  $0^\circ$
- możliwa wersja z 4 kierunkami

$$\text{compass} = \arg \max_i (I * M_i)$$



$$\begin{bmatrix} -1 & -1 & -1 \\ +1 & -2 & +1 \\ +1 & +1 & +1 \\ +1 & +1 & -1 \\ +1 & -2 & -1 \\ +1 & +1 & -1 \\ +1 & +1 & +1 \\ +1 & -2 & +1 \\ -1 & -1 & -1 \\ -1 & +1 & +1 \\ -1 & -2 & +1 \\ -1 & +1 & +1 \end{bmatrix} \begin{bmatrix} +1 & -1 & -1 \\ +1 & -2 & -1 \\ +1 & +1 & +1 \\ +1 & +1 & +1 \\ +1 & -2 & -1 \\ +1 & -1 & -1 \\ +1 & +1 & +1 \\ -1 & -2 & +1 \\ -1 & -1 & +1 \\ -1 & -1 & +1 \\ -1 & -2 & +1 \\ +1 & +1 & +1 \end{bmatrix}$$



## Maski Prewitta

- filtr oparty o maski Prewitta działa identycznie jak operator Sobela
- jedyną różnicę jest użycie innych masek do liczenia gradientu:

$$G_x = \begin{bmatrix} -1 & 0 & +1 \\ -1 & 0 & +1 \\ -1 & 0 & +1 \end{bmatrix} * I \quad G_y = \begin{bmatrix} +1 & +1 & +1 \\ 0 & 0 & 0 \\ -1 & -1 & -1 \end{bmatrix} * I$$

- dają one nieco gorsze rezultaty (nie są tak izotropowe)





## Operator Kirscha

- operator ten, również oparty jest o maksimum gradientu
- aby uniknąć podnoszenia do potęgi i pierwiastkowania proponuje wybór wartości maksymalnej spośród 8 wyników konwolucji z 8 maskami kierunkowymi  $M_i$ :

$$\begin{bmatrix} +5 & +5 & +5 \\ -3 & 0 & -3 \\ -3 & -3 & -3 \end{bmatrix} \quad \begin{bmatrix} -3 & +5 & +5 \\ -3 & 0 & +5 \\ -3 & -3 & -3 \end{bmatrix} \quad \begin{bmatrix} -3 & -3 & +5 \\ -3 & 0 & +5 \\ -3 & -3 & +5 \end{bmatrix} \quad \begin{bmatrix} -3 & -3 & -3 \\ -3 & 0 & +5 \\ +5 & +5 & +5 \end{bmatrix}$$
$$\begin{bmatrix} -3 & -3 & -3 \\ -3 & 0 & -3 \\ +5 & +5 & -3 \end{bmatrix} \quad \begin{bmatrix} +5 & -3 & -3 \\ +5 & 0 & -3 \\ +5 & -3 & -3 \end{bmatrix} \quad \begin{bmatrix} +5 & +5 & -3 \\ +5 & 0 & -3 \\ -3 & -3 & -3 \end{bmatrix}$$

$$J(x, y) = (\max_i(M_i * I))(x, y)$$

- dochodzi jednak koszt policzenia aż 8 operacji konwolucji



## Operator Kirscha

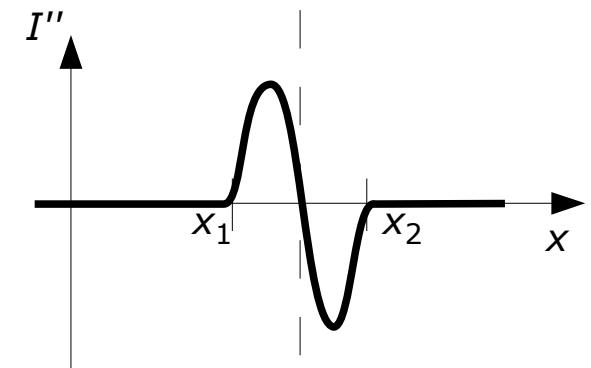
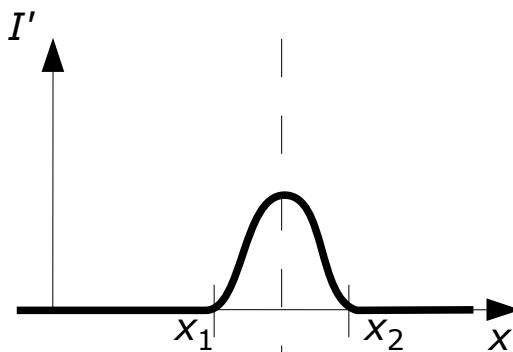
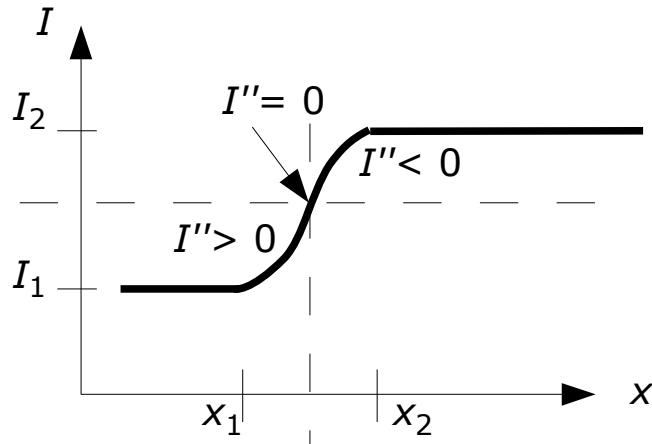


- wyniki nie odbiegają zbytnio od wcześniejszych operatorów



## Detekcja krawędzi oparta na laplasjanie

- maksimum gradientu odpowiada zwykle zerowaniu drugiej pochodnej
- w „okolicach” krawędzi laplasjan osiąga wartości ekstremalne



- laplasjan jest odpowiednikiem drugiej pochodnej 2D

$$\Delta I(x, y) = \frac{\partial^2 I(x, y)}{\partial x^2} + \frac{\partial^2 I(x, y)}{\partial y^2}$$

- laplasjan można przybliżać za pomocą różnic skończonych

$$\Delta I(x, y) = \frac{I(x+h, y) - 2 \cdot I(x, y) + I(x-h, y)}{h^2} + \frac{I(x, y+h) - 2 \cdot I(x, y) + I(x, y-h)}{h^2}$$

- zakładając  $h=1$ , sprowadza się to do konwolucji z maską:

$$\begin{bmatrix} 0 & +1 & 0 \\ +1 & -4 & +1 \\ 0 & +1 & 0 \end{bmatrix}$$



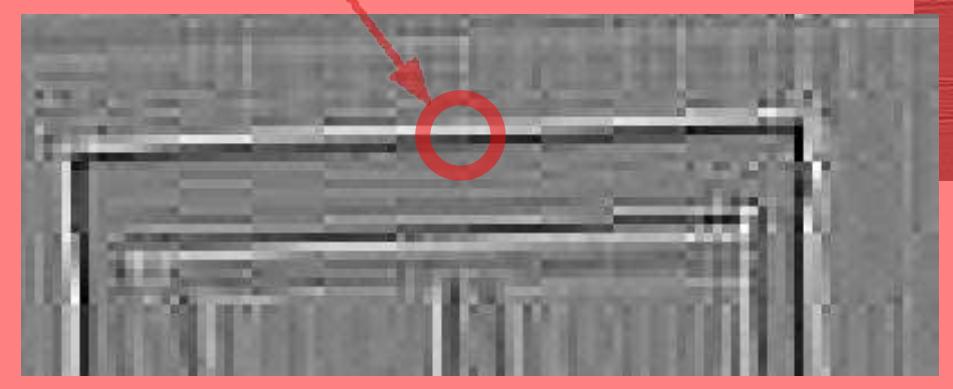
## Laplajian obrazu



$$* \begin{bmatrix} 0 & +1 & 0 \\ +1 & -4 & +1 \\ 0 & +1 & 0 \end{bmatrix} =$$



krawędź między obszarem ciemnym na górze  
i jasnym na dole



- kolor szary = 0,  $>0$  – biały,  $<0$  – czarny
- krawędź rozdziela lapl. ujemny i dodatni
- duża wrażliwość na szum



## Laplasjan obrazu



- inna maska używana do obliczania laplasjanu

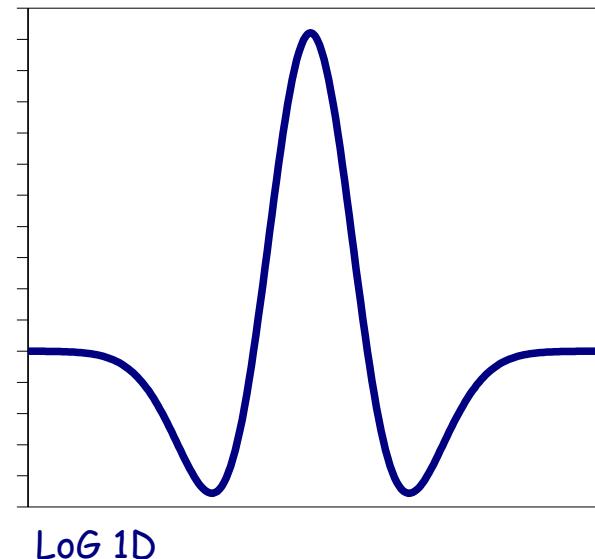
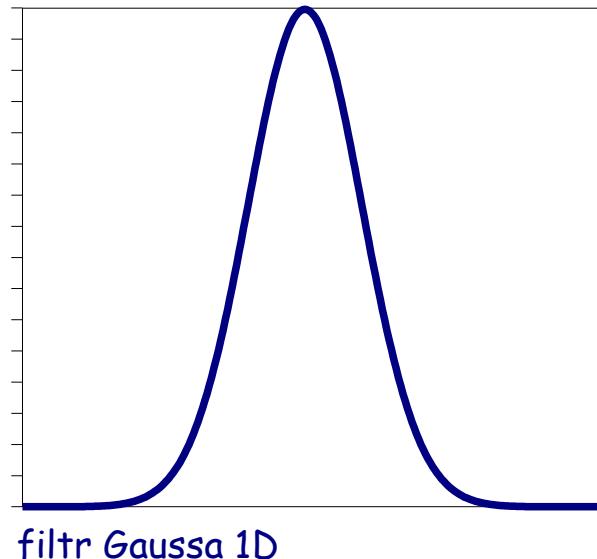
$$\begin{bmatrix} +1 & +1 & +1 \\ +1 & -8 & +1 \\ +1 & +1 & +1 \end{bmatrix}$$



## Laplasjan filtru Gaussa (LoG) (Laplacian of Gaussian)

- z powodu dużej wrażliwości laplasjanu na szum często poprzedza się go wygładzeniem filtrem Gaussa
- obie operacje realizowane są przez konwolucje, można je więc zastąpić jedną operacją konwolucji:

$$(I^* \text{Gauss})^*\text{laplasjan} = I^*(\text{Gauss}^*\text{laplasjan})$$



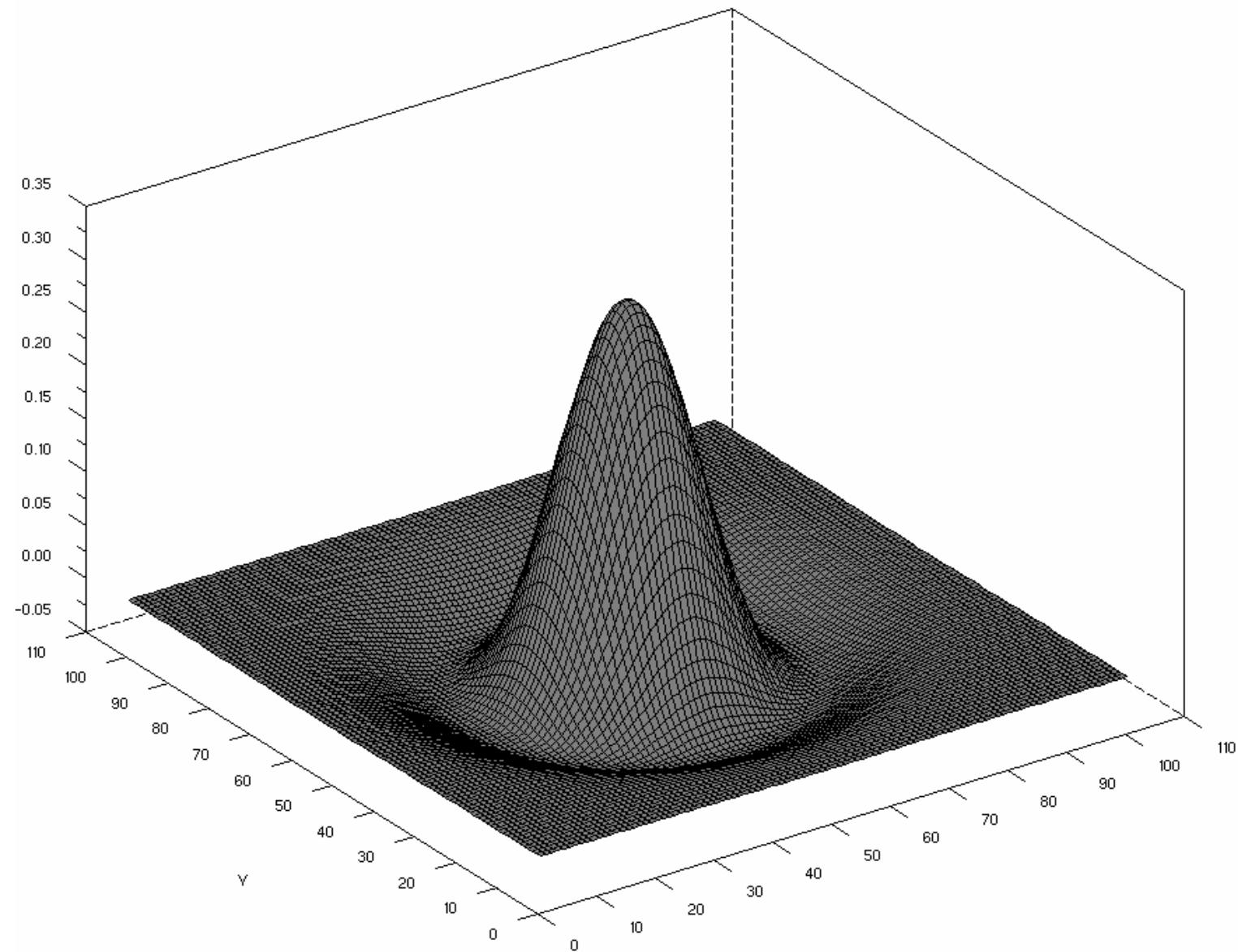
- jedna z masek używana do wyznaczenia LoG:

$$\Delta \text{Gauss}(x, y) = -\frac{1}{\pi \sigma^2} \left( 1 - \frac{x^2 + y^2}{2\sigma^2} \right) e^{-\frac{x^2 + y^2}{2\sigma^2}}$$

$$= \begin{bmatrix} 0 & 1 & 1 & 2 & 2 & 2 & 2 & 1 & 1 & 0 \\ 1 & 2 & 4 & 5 & 5 & 5 & 5 & 4 & 2 & 0 \\ 1 & 4 & 5 & 3 & 0 & 0 & 3 & 5 & 4 & 1 \\ 2 & 5 & 3 & -12 & -24 & -12 & -12 & 3 & 5 & 2 \\ 2 & 5 & 0 & -24 & -50 & -24 & -24 & 0 & 5 & 2 \\ 2 & 5 & 3 & -12 & -24 & -12 & -12 & 3 & 5 & 2 \\ 1 & 4 & 5 & 3 & 0 & 0 & 3 & 5 & 4 & 1 \\ 1 & 2 & 4 & 5 & 5 & 5 & 5 & 4 & 2 & 0 \\ 0 & 1 & 1 & 2 & 2 & 2 & 2 & 1 & 1 & 0 \end{bmatrix}$$



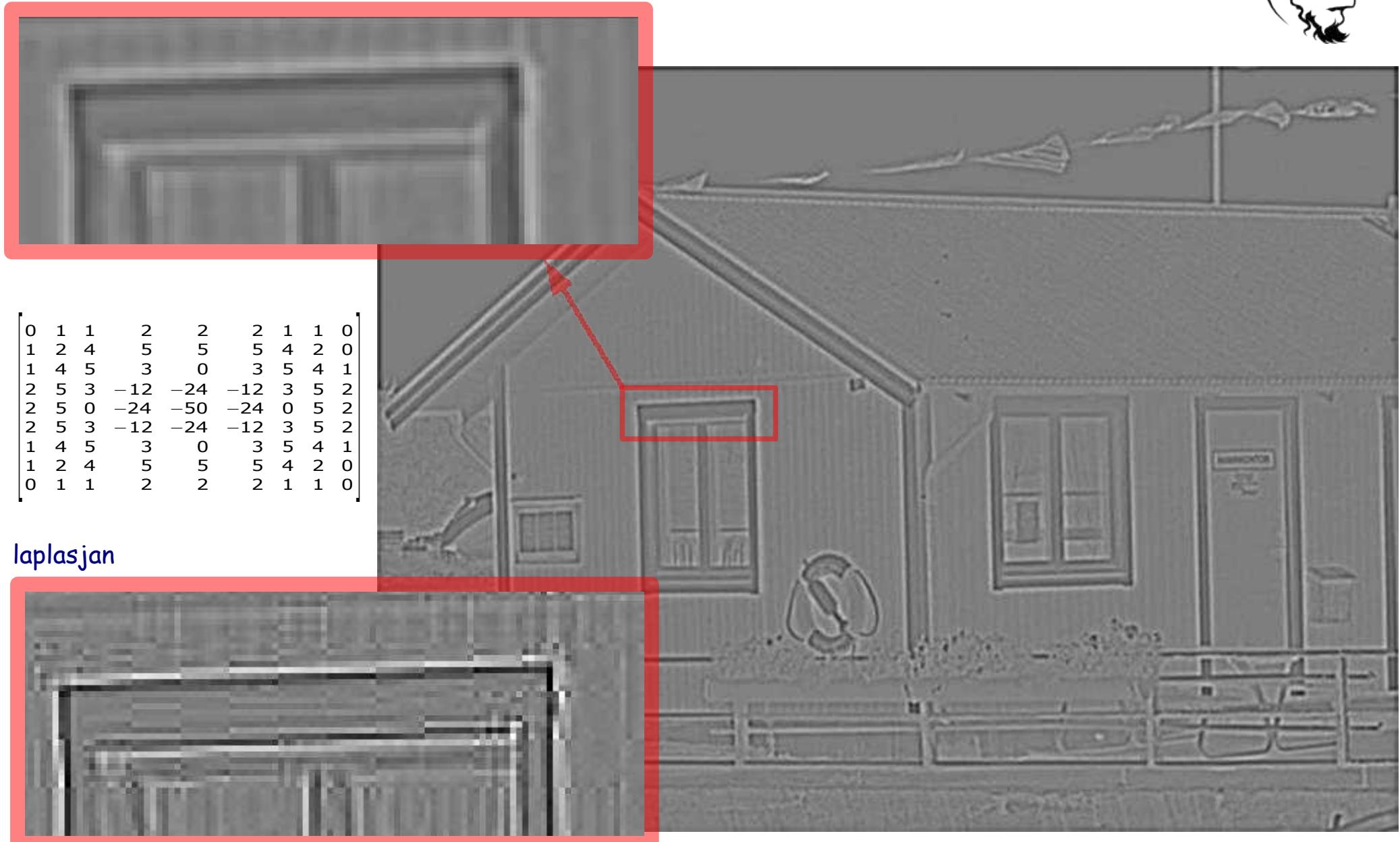
## Maska LoG - Mexican Hat





LoG

## Laplasjan filtru Gaussa (LoG)



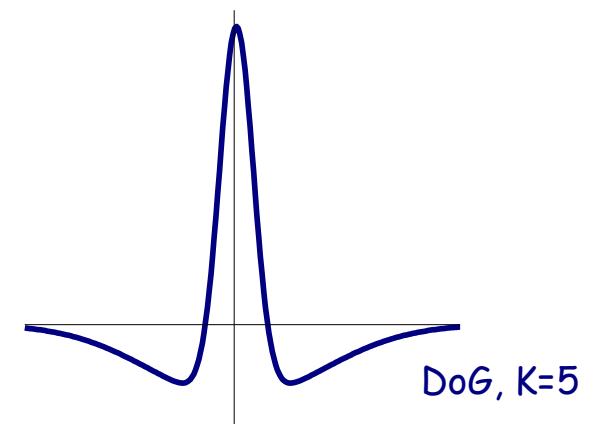
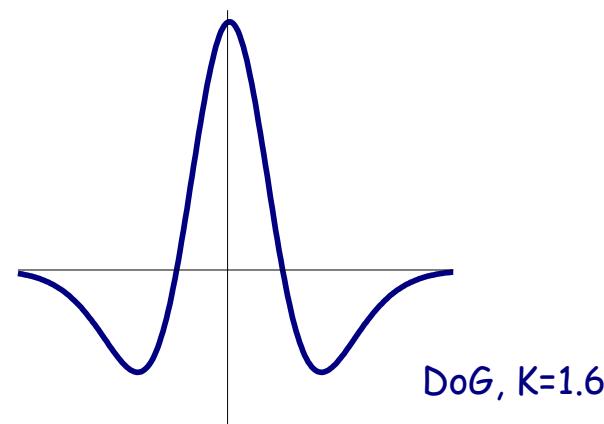
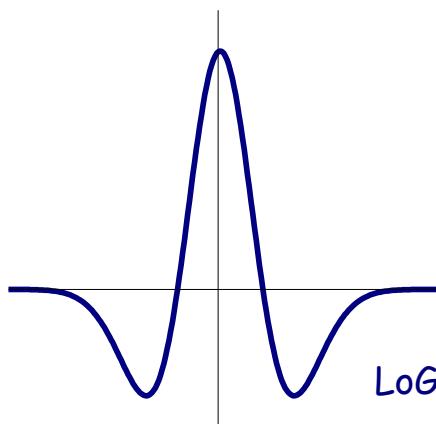


## Różnica filtrów Gaussa (DoG) (Difference of Gaussians)

- LoG można przybliżać różnicą 2 filtrów Gausa o różnym odchyleniu standardowym

$$DoG(x, y, \sigma, K) = \frac{1}{2\pi\sigma^2} e^{-\frac{x^2+y^2}{2\sigma^2}} - \frac{1}{2\pi K^2 \sigma_2^2} e^{-\frac{x^2+y^2}{2K^2\sigma_2^2}}$$

- dla  $K \approx 1.6$  DoG najlepiej przybliża LoG
- dla  $K \approx 5$  DoG, wg pewnych teorii, przybliża sposób widzenia krawędzi przez człowieka





$\text{DoG}, \sigma=1, K=5$



Dwa obrazy wynikowe:

- po lewej wartość bezwględna laplasjanu
- po prawej obraz z dodaną wartością 127
  - piksele szare – wartość zero
  - piksele czarne – wartości ujemne
  - piksele białe – wartości dodatnie

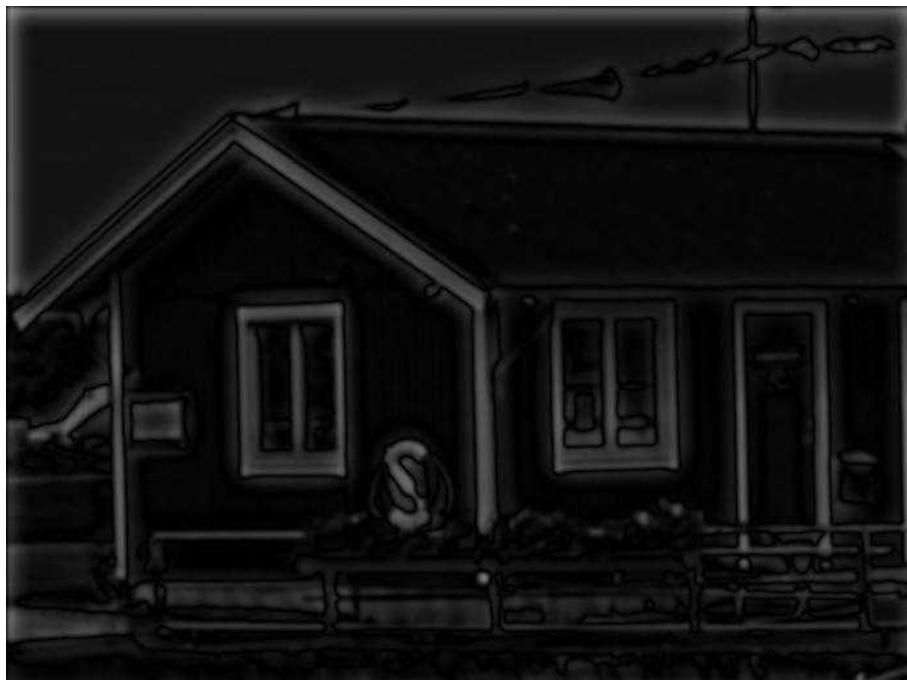




DoG,  $\sigma=2$ , K=5



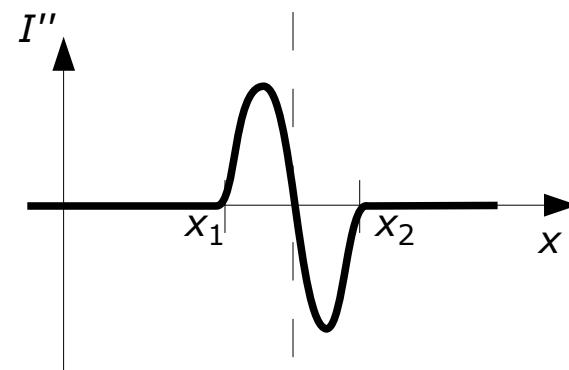
- na lewym obrazie wyraźnie widać czarne kontury krawędzi – zerową wartość laplasjanu
- można to wykorzystać do detekcji krawędzi





## Detekcja przejścia przez zero (zero crossing detector)

- na obrazach typu laplasjan, LoG, DoG krawędzie rozdzielają obszary ujemne i dodatnie (patrz poprzednie obrazy)
- ze względu na wrażliwość 2. pochodnej na szum obraz wygładzany (LoG)
- można zlokalizować piksele leżące na przejściu laplasjanu przez wartość 0
- powstanie w ten sposób obraz krawędzi obrazu
- różne metody detekcji takich pikseli:
  - wystarczy znaleźć parę pikseli o przeciwnych znakach, ten który ma mniejszą wartość bezwzględną laplasjanu (LoG) jest uznawany za krawędź
  - obraz laplasjanu (LoG) po progowaniu z wartością 0 – piksele graniczne obszarów białych/czarnych uznaje się za krawędzie
  - metody analityczne z dokładnością „podpikselową” (level 0)





## Detekcja przejścia przez zero – porównywanie sąsiednich pikseli

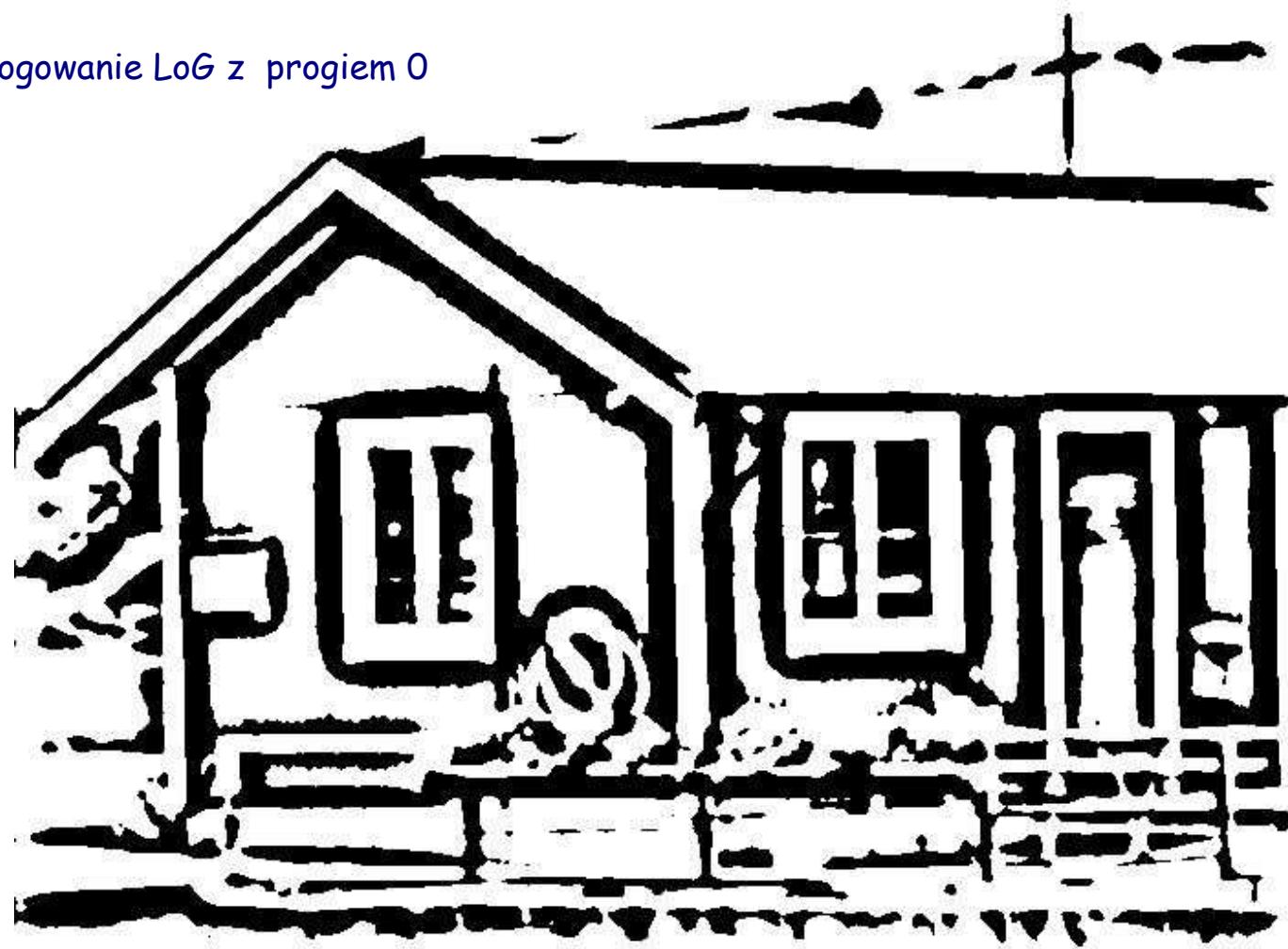
jeżeli  $\text{LoG}(x, y) \cdot \text{LoG}(x+i, y+j) < 0 \wedge |\text{LoG}(x, y)| < |\text{LoG}(x+i, y+j)| \Rightarrow \text{piksel } (x, y) \in \text{"krawędź"}$





## Detekcja przejścia przez zero – granica obszaru po progowaniu

Progowanie LoG z progiem 0





## Detekcja przejścia przez zero – granica obszaru po progowaniu

- detekcja granicy obszaru poprzez operacje morfologii matematycznej
- granica  $\in$  obszar czarny





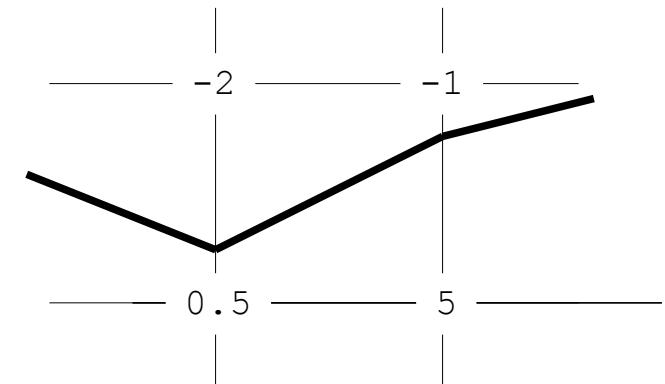
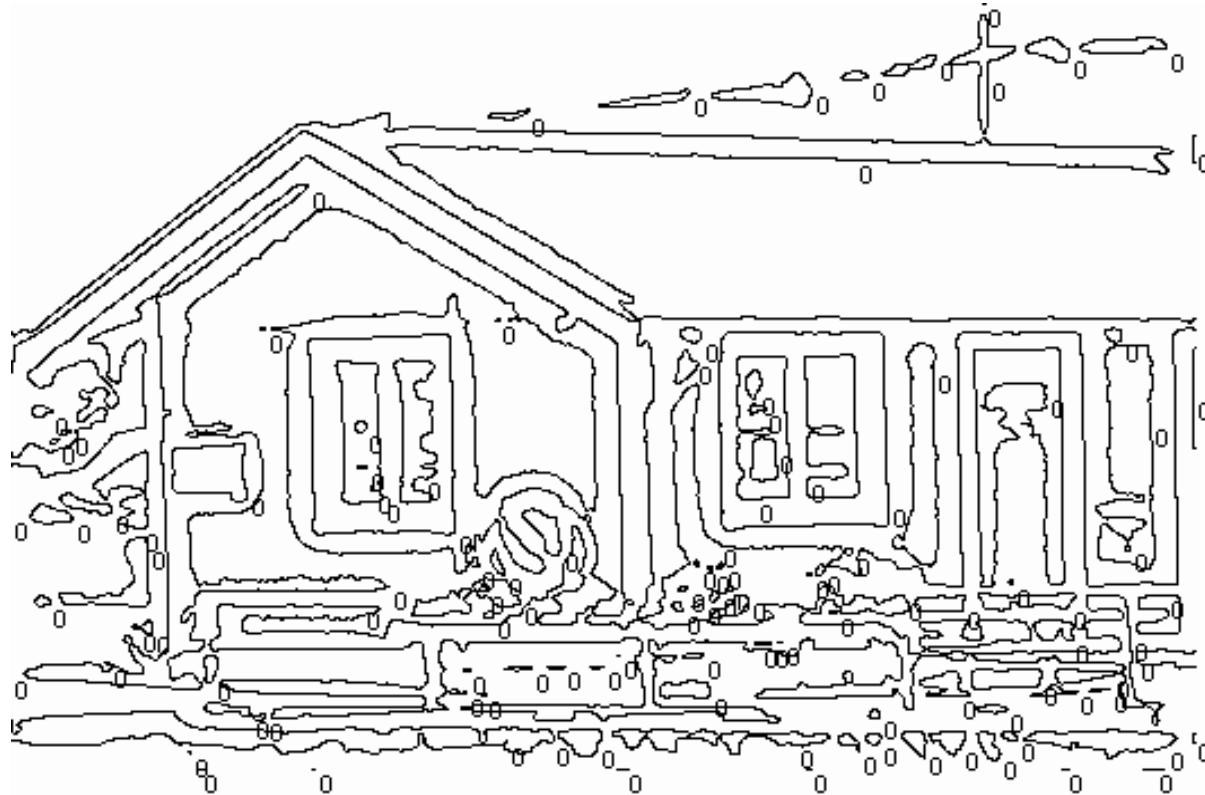
## Detekcja przejścia przez zero – granica obszaru po progowaniu

- detekcja granicy obszaru poprzez operacje morfologii matematycznej
- granica  $\in$  obszar biały





## Detekcja przejścia przez zero - analiza pod pikselowa (level zero)



- dokładność subpikselowa dzięki interpolacji liniowej
- kod w Scilabie:

```
[h w]=size(Image);  
contour2d(1:h, 1:w, ImageLoG', [0 0]);
```



## Detektor Frei'a i Chen'a

- wykonuje się 9 konwolucji (projekcji, rzutwania) z następującymi maskami:

**0.**  $\begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix}$

**1.**  $\begin{bmatrix} -1 & -\sqrt{2} & -1 \\ 0 & 0 & 0 \\ 1 & \sqrt{2} & 1 \end{bmatrix}$

**2.**  $\begin{bmatrix} -1 & 0 & -1 \\ -\sqrt{2} & 0 & \sqrt{2} \\ -1 & 0 & 1 \end{bmatrix}$

**3.**  $\begin{bmatrix} 0 & -1 & \sqrt{2} \\ 1 & 0 & -1 \\ -\sqrt{2} & 1 & 0 \end{bmatrix}$

**4.**  $\begin{bmatrix} \sqrt{2} & -1 & 0 \\ -1 & 0 & 1 \\ 0 & 1 & -\sqrt{2} \end{bmatrix}$

**5.**  $\begin{bmatrix} 0 & 1 & 0 \\ -1 & 0 & 1 \\ 0 & -1 & 0 \end{bmatrix}$

**6.**  $\begin{bmatrix} -1 & 0 & 1 \\ 0 & 0 & 0 \\ 1 & 0 & -1 \end{bmatrix}$

**7.**  $\begin{bmatrix} 1 & -2 & 1 \\ -2 & 4 & -2 \\ 1 & -2 & 1 \end{bmatrix}$

**8.**  $\begin{bmatrix} -2 & 1 & -2 \\ 1 & 4 & 1 \\ -2 & 1 & -2 \end{bmatrix}$

- maski te odgrywają rolę ortogonalnej bazy 9-wymiarowej przestrzeni
- maski 1.-4. stanowią podprzestrzeń odpowiedzialną za krawędzie
- maski 5.-8. stanowią podprzestrzeń odpowiedzialną za linie
- kąt rzutowania na podprzestrzeń krawędzi:      na podprzestrzeń linii

$$\theta_{krawędzi} = \cos^{-1} \sqrt{\frac{\sum_{k=1}^4 (I * M_k)^2}{\sum_{k=0}^8 (I * M_k)^2}}$$

$$\theta_{linii} = \cos^{-1} \sqrt{\frac{\sum_{k=5}^8 (I * M_k)^2}{\sum_{k=0}^8 (I * M_k)^2}}$$

- im mniejszy kąt, tym większa zgodność (obecność szukanej struktury)

## Detektor krawędzi Frei'a i Chen'a



## Detektor linii Frei'a i Chen'a



- jest on zależny od skali – wykrywa linie grubości 1 !!!



## Detektor krawędzi Canny'ego

- zaproponowany w 1986
- Canny, J., A Computational Approach To Edge Detection,  
IEEE Trans. Pattern Analysis and Machine Intelligence, 8:679-714, 1986
- celem było zaproponowanie optymalnego detektora krawędzi spełniającego trzy założenia:
  - dobra detekcja – algorytm powinien wykrywać jak najwięcej rzeczywistych krawędzi
  - dobra lokalizacja – wykryte krawędzie powinny być zlokalizowane maksymalnie blisko tych rzeczywistych
  - minimalna odpowiedź – krawędzie powinny dawać tylko jedną odpowiedź, zaś szum nie powinien powodować fałszywej detekcji
- algorytm złożony, 3 lub 4 etapy
- wymaga podania 3 parametrów, co czasem może utrudniać jego użycie
- kroki algorytmu mogą być wymienne (Template Method ☺ )  $\Rightarrow$  różne implementacje



## Detektor Canny'ego: 1. etap - wygładzanie

- ponieważ detektor jest wrażliwy na szum, w pierwszym kroku obraz jest filtrowany filtrem Gaussa (konwolucja)
- odchylenie standardowe filtru Gaussa jest pierwszym parametrem metody
- tym większe odchylenie standardowe, tym mniej „fałszywych” krawędzi
- ale i czas konwolucji dłuższy





## Detektor Canny'ego: 2. etap – obliczanie gradientu

- kolejna konwolucja: np. z maskami Sobela
- jako obraz wynikowy tego etapu – moduł gradientu





## Detektor Canny'ego: 2. etap - obliczanie gradientu

- ale przyda się też informacja o kierunku gradientu =  $\text{arc tg} (\nabla_x I / \nabla_y I)$
- wystarczy kierunek z dokładnością do  $45^\circ$  : N-S, E-W, SE-NW, SW-NE



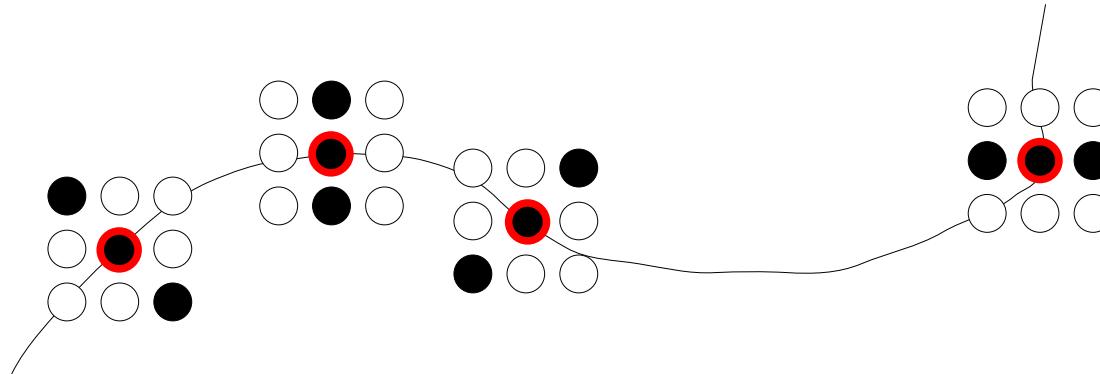
white square	gradient SW-NE
light gray square	gradient E-W
dark gray square	gradient SE-NW
black square	gradient N-S

- kierunek krawędzi jest prostopadły do krawędzi
- brak tu informacji i intensywności

## Detektor Canny'ego: 3. etap – usuwanie niemaksymalnych pikseli (non-maximum suppression)



- dla każdego piksela rozpatrywane są 2 piksele sąsiednie
- **które** ? zakładając, że dany piksel leży na krawędzi, **te**, które powinny być mniej intensywne (gradient!)  $\Rightarrow$  leżące na prostej prostopadłej do krawędzi
- ten kierunek już znamy z obliczenia gradientu !



- jeśli piksel nie jest większy od tych sąsiadów, jest zerowany
- porównanie może być:  $p \geq \text{sąsiad}_1$  i  $p > \text{sąsiad}_2$  aby wziąć pod uwagę równe wartości gradientu
- w ten sposób uzyskuje się linie złożone z pojedynczych pikseli



## Detektor Canny'ego: 3. etap

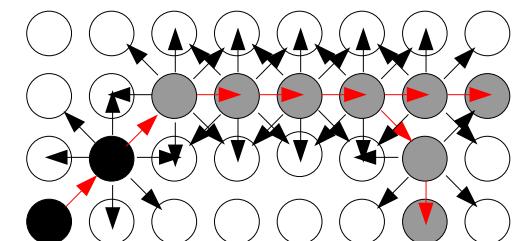
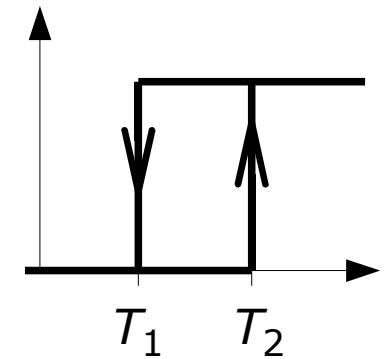


- linie krawędzi mają grubość 1
- niektóre z nich są bardzo słabe – chcemy się ich pozbyć



## Detektor Canny'ego: 4. etap – progowanie z histerezą

- chcemy zachować jasne krawędzie – gdzie gradient ma dużą wartość
- chcemy usunąć ciemne krawędzie – gdzie gradient ma małą wartość
- czasami jasne krawędzie mają ciemniejsze fragmenty (szum, zakłócenia)
- ponieważ chcemy mieć **ciągłe** krawędzie, ciemniejsze fragmenty chcemy włączyć do wyniku
- progowanie z **histerezą**: 2 progi  $T_1 < T_2$
- jeśli intensywność krawędzi (moduł gradientu) jest większa niż  $T_2$ , uznaje się ją za krawędź ostateczną
- jeśli intensywność krawędzi jest mniejsza od  $T_1$ , usuwana jest ona z obrazu wynikowego
- jeśli intensywność krawędzi jest między  $T_1$  i  $T_2$ , to krawędź jest usuwana jeśli nie przylega do już wyodrębnionej krawędzi
- odbywa się to przez rekurencyjne śledzenie krawędzi: dla bieżącego piksela krawędzi wszyscy sąsiedzi o wartości gradientu powyżej  $T_1$  stają się nowymi początkami śledzenia
- 2 kolejne parametry metody





## Detektor Canny'ego: wyniki



$$\sigma=3, T_1=0.25, T_2=125$$



## Detektor Canny'ego: wyniki



$$\sigma=3, T_1=75, T_2=125$$



## Detektor Canny'ego: wyniki



$$\sigma=3, T_1=25, T_2=75$$



## Detektor Canny'ego: wyniki



$$\sigma=1, T_1=75, T_2=125$$



## Detektor Canny'ego: wyniki



$\sigma=1, T_1=75, T_2=250$



## Detektor Canny'ego: wyniki



$$\sigma=2, T_1=50, T_2=150$$



## Detekcja linii

- linie a krawędzie
- ważna jest:
  - grubość linii
  - jej intensywność względem tła (linie jasne i ciemne)
  - czasami również orientacja
- linie 2D i 3D
- najprostszy detektor: konwolucja z maskami kierunkowymi:

jasna pozioma linia  
na ciemnym tle

$$\begin{bmatrix} -1 & -1 & -1 \\ +2 & +2 & +2 \\ -1 & -1 & -1 \end{bmatrix}$$

ciemna ukośna linia  
na jasnym tle

$$\begin{bmatrix} +1 & +1 & -2 \\ +1 & -2 & +1 \\ -2 & +1 & +1 \end{bmatrix}$$

- słabe efekty: ścisłe dopasowanie do wielkości i orientacji
- nawet małe odstępstwa w grubości znacznie obniża odpowiedź
- wrażliwy również na krawędzie
- transformata Hough'a – ale to bardziej lokalizowanie

## Detekcja linii przez konwolucję



$$\begin{bmatrix} +1 & -2 & +1 \\ +1 & -2 & +1 \\ +1 & -2 & +1 \end{bmatrix}$$



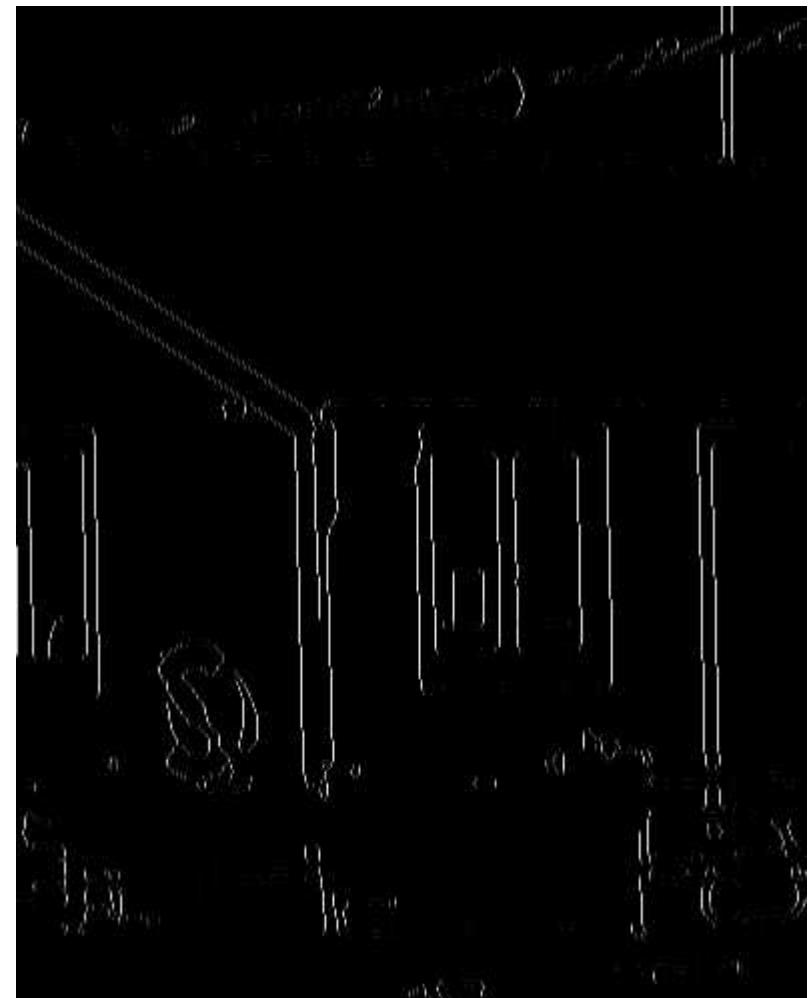
$$\begin{bmatrix} -1 & +2 & -1 \\ -1 & +2 & -1 \\ -1 & +2 & -1 \end{bmatrix}$$



# Detekcja linii przez konwolucję



- lepsze wyniki dla obrazów szkieletowych



$$\begin{bmatrix} -1 & +2 & -1 \\ -1 & +2 & -1 \\ -1 & +2 & -1 \end{bmatrix}$$



## Detekcja skupisk (blob detection)

- **interesujące obszary, obszary szczególne**
- najczęściej jasne (lub ciemne), zwarte (kołowe, kuliste) obszary na ciemnym (lub odpowiednio jasnym) tle
- jeden z tradycyjnych detektorów takich obszarów: laplasjan (DoG, LoG)
  - silnie uzależniony od skali obszaru, wykrywa też krawędzie



skala=2 (maska 5x5)

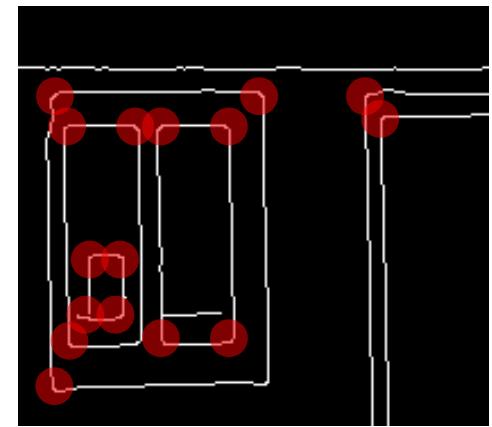
- wskazane wcześniejsze wygładzenie obrazu
- wiele innych algorytmów (patrz: [wikipedia.org](https://wikipedia.org) 'corner detection')



## Detekcja rogów

- róg w pojęciu „gwałtownie zakręcającej krawędzi”
  - przecięcie 2 krawędzi, linii
  - punkty, które w swoim otoczeniu mają więcej niż jedne dominujący kierunek
- rodzaj **interesujących punktów, punktów szczególnych** (inne: zakończenia linii, jasne lub ciemne punkty)
- pierwsze algorytmy wykrywające rogi:
  - detekcja krawędzi
  - śledzenie wyodrębnionych krawędzi i detekcja nagłej zmiany ich kierunku:
- nowsza generacja algorytmów:
  - wysoka krzywizna (*curvature*) gradientu

$$\kappa = \frac{I_{xx} \cdot I_y^2 - 2 \cdot I_{xy} \cdot I_x \cdot I_y + I_{yy} \cdot I_x^2}{|\nabla I|}$$



- wskazane wcześniejsze wygładzenie obrazu
- wiele innych algorytmów (patrz: wikipedia.org 'corner detection')

## Regiony z wysoką wartością bezwzględną krzywizny





## Detektory oparte na wartościach własnych hesjanu

- hesjan (macierz Hess'a) to macierz drugich pochodnych

$$H = \begin{bmatrix} \frac{\partial^2 I}{\partial x^2} & \frac{\partial^2 I}{\partial x \partial y} \\ \frac{\partial^2 I}{\partial x \partial y} & \frac{\partial^2 I}{\partial y^2} \end{bmatrix} = \begin{bmatrix} I_{xx} & I_{xy} \\ I_{xy} & I_{yy} \end{bmatrix}$$

- pochodne cząstkowe można wyliczyć za pomocą konwolucji

$$I_{xx} = I * \begin{bmatrix} 0 & 0 & 0 \\ 1 & -2 & 1 \\ 0 & 0 & 0 \end{bmatrix}, \quad I_{yy} = I * \begin{bmatrix} 0 & 1 & 0 \\ 0 & -2 & 0 \\ 0 & 1 & 0 \end{bmatrix}, \quad I_{xy} = I * \begin{bmatrix} -\frac{1}{4} & 0 & \frac{1}{4} \\ 0 & 0 & 0 \\ \frac{1}{4} & 0 & -\frac{1}{4} \end{bmatrix}$$

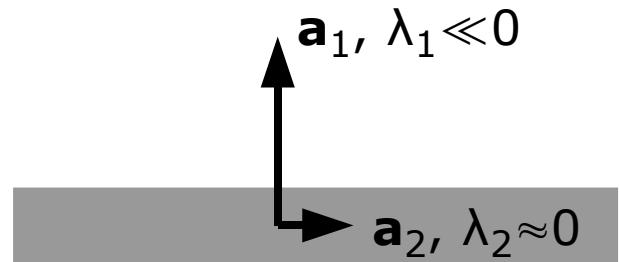
dla kroku (skali) 1

- jeśli uporządkujemy wektory hesjanu  $\mathbf{a}_1$  i  $\mathbf{a}_2$  (które są prostopadłe do siebie) i odpowiadające im wartości  $\lambda_1$  i  $\lambda_2$  tak, aby  $|\lambda_1| > |\lambda_2|$ , to  $\mathbf{a}_1$  wskazuje kierunek największej zmienności obrazu (jego gradientu) a  $\lambda_1$  jest proporcjonalna do tej zmiany:
  - $>0$  jeśli w tym kierunku obraz jaśnieje
  - $<0$  jeśli w tym kierunku obraz ciemnieje

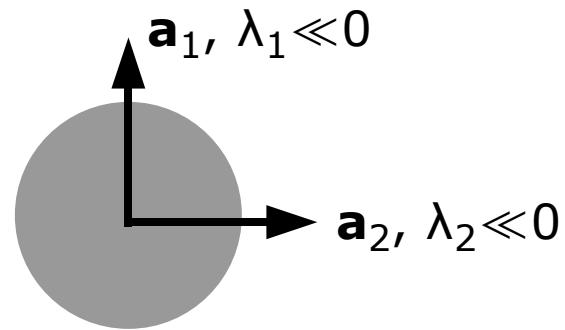


## Konfiguracje wartości własnych hesjanu

- $\lambda_1 \approx 0$  i  $\lambda_2 \approx 0 \Rightarrow$  obraz jest jednolity (nie wykazuje zmienności)
- $\lambda_1 \gg 0$  i  $\lambda_2 \approx 0 \Rightarrow$  ciemna linia



- $\lambda_1 \ll 0$  i  $\lambda_2 \approx 0 \Rightarrow$  jasna linia
- $\lambda_1 \gg 0$  i  $\lambda_2 \gg 0$  i  $\lambda_1 \approx \lambda_2 \Rightarrow$  ciemne skupisko (*bulb*)



- $\lambda_1 \ll 0$  i  $\lambda_2 \ll 0$  i  $\lambda_1 \approx \lambda_2 \Rightarrow$  jasne skupisko



## Filtr wykrywający linie

- filtr wykrywający linie jasne ( $\lambda_1 \ll 0$  i  $\lambda_2 \approx 0$ )

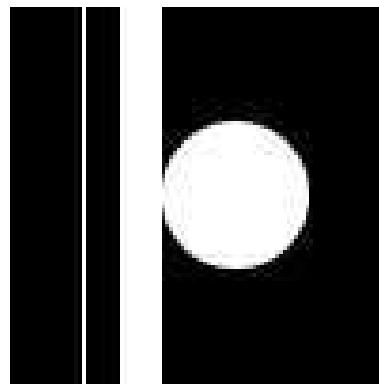
$$\lambda_{max} = \max(\lambda_1, \lambda_2); \quad \lambda_{min} = \min(\lambda_1, \lambda_2)$$

$$odp = -\frac{\lambda_{min}}{|\lambda_{max}| + 1}$$

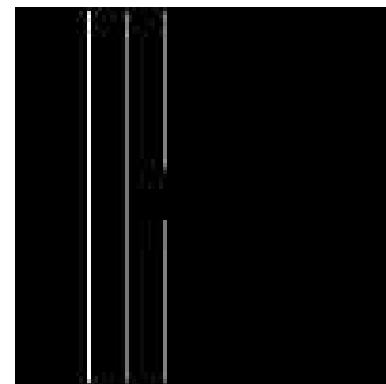
- filtr wykrywający linie ciemne ( $\lambda_1 \gg 0$  i  $\lambda_2 \approx 0$ )

$$\lambda_{max} = \max(\lambda_1, \lambda_2); \quad \lambda_{min} = \min(\lambda_1, \lambda_2)$$

$$odp = \frac{\lambda_{max}}{|\lambda_{min}| + 1}$$



skala=1 (maska 3x3)

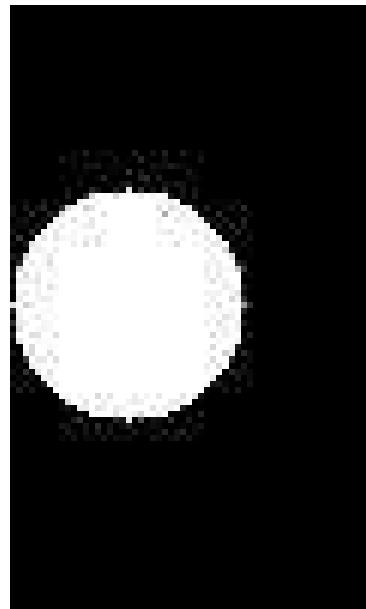
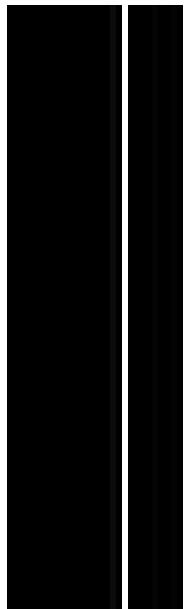


- jak widać słabsze odpowiedzi są też na granicach płaskich obszarów (powierzchniach)



## Problemy

- duże odpowiedzi na inne konfiguracje
- przykład – kiedy koło nie jest rozmyte na brzegach tylko binarne:



skala=1 (maska 3x3)



- wartości własne mają tam zbliżone do szukanej wartości

## Linie ciemne o różnej grubości



skala=1 (maska 3x3)



skala=5 (maska 11x11)



skala=9 (maska 18x18)



rzet

## Linie jasne o różnej grubości



skala=1 (maska 3x3)



skala=5 (maska 11x11)



skala=9 (maska 18x18)





## Detekcja skupień

- przykładowa odpowiedź mierzy symetryczność wartości własnych, ich jednakowy znak, premiuje większe różnice, wykrywa jasne i ciemne skupiska

$$odp = \begin{cases} \frac{\lambda_2}{\lambda_1} \cdot (|\lambda_1| + |\lambda_2|) & \text{jeżeli } (|\lambda_1| > |\lambda_2|) \wedge (\lambda_1 \neq 0) \\ \frac{\lambda_1}{\lambda_2} \cdot (|\lambda_1| + |\lambda_2|) & \text{jeżeli } (|\lambda_2| > |\lambda_1|) \wedge (\lambda_2 \neq 0) \\ 0 & \text{w przeciwnym przypadku} \end{cases}$$

skala=1 (maska 3x3)



skala=5 (maska 11x11)



# GNU Free Documentation License



Version 1.2, November 2002

Copyright (C) 2000,2001,2002 Free Software Foundation, Inc.  
51 Franklin St, Fifth Floor, Boston, MA 02110-1301 USA  
Everyone is permitted to copy and distribute verbatim copies  
of this license document, but changing it is not allowed.

## 0. PREAMBLE

The purpose of this License is to make a manual, textbook, or other functional and useful document "free" in the sense of freedom: to assure everyone the effective freedom to copy and redistribute it, with or without modifying it, either commercially or noncommercially. Secondly, this License preserves for the author and publisher a way to get credit for their work, while not being considered responsible for modifications made by others.

This License is a kind of "copyleft", which means that derivative works of the document must themselves be free in the same sense. It complements the GNU General Public License, which is a copyleft license designed for free software.

We have designed this License in order to use it for manuals for free software, because free software needs free documentation: a free program should come with manuals providing the same freedoms that the software does. But this License is not limited to software manuals; it can be used for any textual work, regardless of subject matter or whether it is published as a printed book. We recommend this License principally for works whose purpose is instruction or reference.

## 1. APPLICABILITY AND DEFINITIONS

This License applies to any manual or other work, in any medium, that contains a notice placed by the copyright holder saying it can be distributed under the terms of this License. Such a notice grants a world-wide, royalty-free license, unlimited in duration, to use that work under the conditions stated herein. The "Document", below, refers to any such manual or work. Any member of the public is a licensee, and is addressed as "you". You accept the license if you copy, modify or distribute the work in a way requiring permission under copyright law.

A "Modified Version" of the Document means any work containing the Document or a portion of it, either copied verbatim, or with modifications and/or translated into another language.

A "Secondary Section" is a named appendix or a front-matter section of the Document that deals exclusively with the relationship of the publishers or authors of the Document to the Document's overall subject (or to related matters) and contains nothing that could fall directly within that overall subject. (Thus, if the Document is in part a textbook of mathematics, a Secondary Section may not explain any mathematics.) The relationship could be a matter of historical connection with the subject or with related matters, or of legal, commercial, philosophical, ethical or political position regarding them.

The "Invariant Sections" are certain Secondary Sections whose titles are designated, as being those of Invariant Sections, in the notice that says that the Document is released under this License. If a section does not fit the above definition of Secondary then it is not allowed to be designated as Invariant. The Document may contain zero Invariant Sections. If the Document does not identify any Invariant Sections then there are none.

The "Cover Texts" are certain short passages of text that are listed, as Front-Cover Texts or Back-Cover Texts, in the notice that says that the Document is released under this License. A Front-Cover Text may be at most 5 words, and a Back-Cover Text may be at most 25 words.

A "Transparent" copy of the Document means a machine-readable copy, represented in a format whose specification is available to the general public, that is suitable for revising the document straightforwardly with generic text editors or (for images composed of pixels) generic paint programs or (for drawings) some widely available drawing editor, and that is suitable for input to text formatters or for automatic translation to a variety of formats suitable for input to text formatters. A copy made in an otherwise Transparent file format whose markup, or absence of markup, has been arranged to thwart or discourage subsequent modification by readers is not Transparent. An image format is not Transparent if used for any substantial amount of text. A copy that is not "Transparent" is called "Opaque".

Examples of suitable formats for Transparent copies include plain ASCII without markup, Texinfo input format, LaTeX input format, SGML or XML using a publicly available DTD, and standard-conforming simple HTML, PostScript or PDF designed for human modification. Examples of transparent image formats include PNG, XCF and JPG. Opaque formats include proprietary formats that can be read and edited only by proprietary word processors, SGML or XML for which the DTD and/or processing tools are not generally available, and the machine-generated HTML, PostScript or PDF produced by some word processors for output purposes only.

The "Title Page" means, for a printed book, the title page itself, plus such following pages as are needed to hold, legibly, the material this License requires to appear in the title page. For works in formats which do not have any title page as such, "Title Page" means the text near the most prominent appearance of the work's title, preceding the beginning of the body of the text.

A section "Entitled XYZ" means a named subunit of the Document whose title either is precisely XYZ or contains XYZ in parentheses following text that translates XYZ in another language. (Here XYZ stands for a specific section name mentioned below, such as "Acknowledgements", "Dedications", "Endorsements", or "History".) To "Preserve the Title" of such a section when you modify the Document means that it remains a section "Entitled XYZ" according to this definition.

The Document may include Warranty Disclaimers next to the notice which states that this License applies to the Document. These Warranty Disclaimers are considered to be included by reference in this License, but only as regards disclaiming warranties: any other implication that these Warranty Disclaimers may have is void and has no effect on the meaning of this License.

## 2. VERBATIM COPYING

You may copy and distribute the Document in any medium, either commercially or noncommercially, provided that this License, the copyright notices, and the license notice saying this License applies to the Document are reproduced in all copies, and that you add no other conditions whatsoever to those of this License. You may not use technical measures to obstruct or control the reading or further copying of the copies you make or distribute. However, you may accept compensation in exchange for copies. If you distribute a large enough number of copies you must also follow the conditions in section 3.

You may also lend copies, under the same conditions stated above, and you may publicly display copies.



### 3. COPYING IN QUANTITY

If you publish printed copies (or copies in media that commonly have printed covers) of the Document, numbering more than 100, and the Document's license notice requires Cover Texts, you must enclose the copies in covers that carry, clearly and legibly, all these Cover Texts: Front-Cover Texts on the front cover, and Back-Cover Texts on the back cover. Both covers must also clearly and legibly identify you as the publisher of these copies. The front cover must present the full title with all words of the title equally prominent and visible. You may add other material on the covers in addition. Copying with changes limited to the covers, as long as they preserve the title of the Document and satisfy these conditions, can be treated as verbatim copying in other respects.

If the required texts for either cover are too voluminous to fit legibly, you should put the first ones listed (as many as fit reasonably) on the actual cover, and continue the rest onto adjacent pages.

If you publish or distribute Opaque copies of the Document numbering more than 100, you must either include a machine-readable Transparent copy along with each Opaque copy, or state in or with each Opaque copy a computer-network location from which the general network-using public has access to download using public-standard network protocols a complete Transparent copy of the Document, free of added material. If you use the latter option, you must take reasonably prudent steps, when you begin distribution of Opaque copies in quantity, to ensure that this Transparent copy will remain thus accessible at the stated location until at least one year after the last time you distribute an Opaque copy (directly or through your agents or retailers) of that edition to the public.

It is requested, but not required, that you contact the authors of the Document well before redistributing any large number of copies, to give them a chance to provide you with an updated version of the Document.

### 4. MODIFICATIONS

You may copy and distribute a Modified Version of the Document under the conditions of sections 2 and 3 above, provided that you release the Modified Version under precisely this License, with the Modified Version filling the role of the Document, thus licensing distribution and modification of the Modified Version to whoever possesses a copy of it. In addition, you must do these things in the Modified Version:

- \* A. Use in the Title Page (and on the covers, if any) a title distinct from that of the Document, and from those of previous versions (which should, if there were any, be listed in the History section of the Document). You may use the same title as a previous version if the original publisher of that version gives permission.
- \* B. List on the Title Page, as authors, one or more persons or entities responsible for authorship of the modifications in the Modified Version, together with at least five of the principal authors of the Document (all of its principal authors, if it has fewer than five), unless they release you from this requirement.
- \* C. State on the Title page the name of the publisher of the Modified Version, as the publisher.
- \* D. Preserve all the copyright notices of the Document.
- \* E. Add an appropriate copyright notice for your modifications adjacent to the other copyright notices.
- \* F. Include, immediately after the copyright notices, a license notice giving the public permission to use the Modified Version under the terms of this License, in the form shown in the Addendum below.
- \* G. Preserve in that license notice the full lists of Invariant Sections and required Cover Texts given in the Document's license notice.
- \* H. Include an unaltered copy of this License.
- \* I. Preserve the section Entitled "History", Preserve its Title, and add to it an item stating at least the title, year, new authors, and publisher of the Modified Version as given on the Title Page. If there is no section Entitled "History" in the Document, create one stating the title, year, authors, and publisher of the Document as given on its Title Page, then add an item describing the Modified Version as stated in the previous sentence.
- \* J. Preserve the network location, if any, given in the Document for public access to a Transparent copy of the Document, and likewise the network locations given in the Document for previous versions it was based on. These may be placed in the "History" section. You may omit a network location for a work that was published at least four years before the Document itself, or if the original publisher of the version it refers to gives permission.
- \* K. For any section Entitled "Acknowledgements" or "Dedications", Preserve the Title of the section, and preserve in the section all the substance and tone of each of the contributor acknowledgements and/or dedications given therein.
- \* L. Preserve all the Invariant Sections of the Document, unaltered in their text and in their titles. Section numbers or the equivalent are not considered part of the section titles.
- \* M. Delete any section Entitled "Endorsements". Such a section may not be included in the Modified Version.
- \* N. Do not retitle any existing section to be Entitled "Endorsements" or to conflict in title with any Invariant Section.
- \* O. Preserve any Warranty Disclaimers.

If the Modified Version includes new front-matter sections or appendices that qualify as Secondary Sections and contain no material copied from the Document, you may at your option designate some or all of these sections as invariant. To do this, add their titles to the list of Invariant Sections in the Modified Version's license notice. These titles must be distinct from any other section titles.

You may add a section Entitled "Endorsements", provided it contains nothing but endorsements of your Modified Version by various parties--for example, statements of peer review or that the text has been approved by an organization as the authoritative definition of a standard.

You may add a passage of up to five words as a Front-Cover Text, and a passage of up to 25 words as a Back-Cover Text, to the end of the list of Cover Texts in the Modified Version. Only one passage of Front-Cover Text and one of Back-Cover Text may be added by (or through arrangements made by) any one entity. If the Document already includes a cover text for the same cover, previously added by you or by arrangement made by the same entity you are acting on behalf of, you may not add another; but you may replace the old one, on explicit permission from the previous publisher that added the old one.

The author(s) and publisher(s) of the Document do not by this License give permission to use their names for publicity for or to assert or imply endorsement of any Modified Version.



## 6. COLLECTIONS OF DOCUMENTS

You may make a collection consisting of the Document and other documents released under this License, and replace the individual copies of this License in the various documents with a single copy that is included in the collection, provided that you follow the rules of this License for verbatim copying of each of the documents in all other respects.

You may extract a single document from such a collection, and distribute it individually under this License, provided you insert a copy of this License into the extracted document, and follow this License in all other respects regarding verbatim copying of that document.

## 7. AGGREGATION WITH INDEPENDENT WORKS

A compilation of the Document or its derivatives with other separate and independent documents or works, in or on a volume of a storage or distribution medium, is called an "aggregate" if the copyright resulting from the compilation is not used to limit the legal rights of the compilation's users beyond what the individual works permit. When the Document is included in an aggregate, this License does not apply to the other works in the aggregate which are not themselves derivative works of the Document.

If the Cover Text requirement of section 3 is applicable to these copies of the Document, then if the Document is less than one half of the entire aggregate, the Document's Cover Texts may be placed on covers that bracket the Document within the aggregate, or the electronic equivalent of covers if the Document is in electronic form. Otherwise they must appear on printed covers that bracket the whole aggregate.

## 8. TRANSLATION

Translation is considered a kind of modification, so you may distribute translations of the Document under the terms of section 4. Replacing Invariant Sections with translations requires special permission from their copyright holders, but you may include translations of some or all Invariant Sections in addition to the original versions of these Invariant Sections. You may include a translation of this License, and all the license notices in the Document, and any Warranty Disclaimers, provided that you also include the original English version of this License and the original versions of those notices and disclaimers. In case of a disagreement between the translation and the original version of this License or a notice or disclaimer, the original version will prevail.

If a section in the Document is Entitled "Acknowledgements", "Dedications", or "History", the requirement (section 4) to Preserve its Title (section 1) will typically require changing the actual title.

## 9. TERMINATION

You may not copy, modify, sublicense, or distribute the Document except as expressly provided for under this License. Any other attempt to copy, modify, sublicense or distribute the Document is void, and will automatically terminate your rights under this License. However, parties who have received copies, or rights, from you under this License will not have their licenses terminated so long as such parties remain in full compliance.

## 10. FUTURE REVISIONS OF THIS LICENSE

The Free Software Foundation may publish new, revised versions of the GNU Free Documentation License from time to time. Such new versions will be similar in spirit to the present version, but may differ in detail to address new problems or concerns. See <http://www.gnu.org/copyleft/>.

Each version of the License is given a distinguishing version number. If the Document specifies that a particular numbered version of this License "or any later version" applies to it, you have the option of following the terms and conditions either of that specified version or of any later version that has been published (not as a draft) by the Free Software Foundation. If the Document does not specify a version number of this License, you may choose any version ever published (not as a draft) by the Free Software Foundation.

## 5. COMBINING DOCUMENTS

You may combine the Document with other documents released under this License, under the terms defined in section 4 above for modified versions, provided that you include in the combination all of the Invariant Sections of all of the original documents, unmodified, and list them all as Invariant Sections of your combined work in its license notice, and that you preserve all their Warranty Disclaimers.

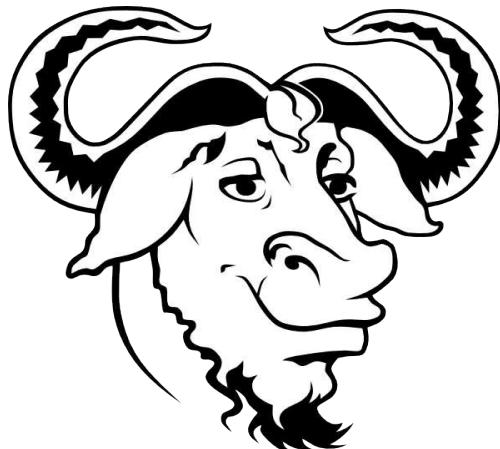
The combined work need only contain one copy of this License, and multiple identical Invariant Sections may be replaced with a single copy. If there are multiple Invariant Sections with the same name but different contents, make the title of each such section unique by adding at the end of it, in parentheses, the name of the original author or publisher of that section if known, or else a unique number. Make the same adjustment to the section titles in the list of Invariant Sections in the license notice of the combined work.

In the combination, you must combine any sections Entitled "History" in the various original documents, forming one section Entitled "History"; likewise combine any sections Entitled "Acknowledgements", and any sections Entitled "Dedications". You must delete all sections Entitled "Endorsements."

(not as a draft) by the Free Software Foundation.

# Cyfrowe Przetwarzanie Obrazów

## Digital Image Processing



Aurelio A. Heckert

Copyright (c) 2008 Cezary Bołdak

Permission is granted to copy, distribute and/or modify this document under the terms of the GNU Free Documentation License, Version 1.2 or any later version published by the Free Software Foundation; with no Invariant Sections, no Front-Cover Texts, and no Back-Cover Texts. A copy of the license is included in the section entitled "GNU Free Documentation License"



## Wykład 5

# Morfologia matematyczna (mathematical morphology)



- **morfologia matematyczna** jest teoretycznym modelem przetwarzania obrazów cyfrowych w nieliniowy sposób
- swoje źródła ma w teorii zbiorów Minkowskiego (suma Minkowskiego), topologii
- podstawowe operacje:
  - **dylacja** (*dilation*),
  - **erozja** (*erosion*),
  - **szkieletyzacja** (*skeletonization*)
- pozostałe operacje:
  - **otwarcie** (*opening*),
  - **zamknięcie** (*closing*),
  - **zmniejszanie** (*shrinking*),
  - **ścienianie** (*thinning*),
  - **pogrubianie** (*thickening*),
  - **obcinanie** (*pruning*),
  - **transformata odległościowa** (*distance transform*)
- ma zastosowanie głównie do **obrazów binarnych**
- ale istnieje rozszerzenie dla obraz w skali szarości
- różne cele: filtracja obrazu, segmentacja, detekcja cech, ...
- swoje źródła ma we Francji



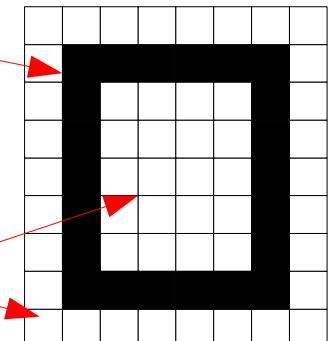
## Obraz binarny - tło i pierwszy plan (background, foreground)

- obraz cyfrowy, którego piksele mogą przyjmować tylko 2 wartości
  - (0,1), (0,255), (1,256), (-1,1), (True, False), ...
- jedna z wartości uznawana jest za tło, druga za pierwszy plan (obiekt w obrazie)
- kwestia umowna, który
  - tło: 0 (czarny), obiekt: 1(biały)
  - tło: ~~1~~<sup>A</sup>(biały), obiekt: ~~0~~<sup>A</sup>(czarny) – tak będzie na wykładzie
- konwencji raz wybranej trzeba się trzymać
- **suportem** (wsparciem, *support*) nazywamy zbiór wszystkich pikseli o wartości odpowiadającej pierwszemu planowi<sup>A</sup>

$$supp(I) = \{p = (x, y) \in I : I(p) = fg\}$$

- **dopełnieniem** (*complement*) suportu nazywamy zbiór wszystkich pozostałych pikseli (o wartościach tła)<sup>A</sup>

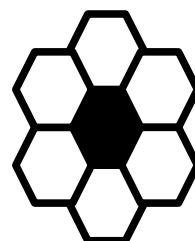
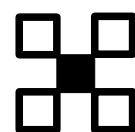
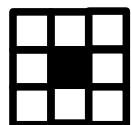
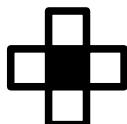
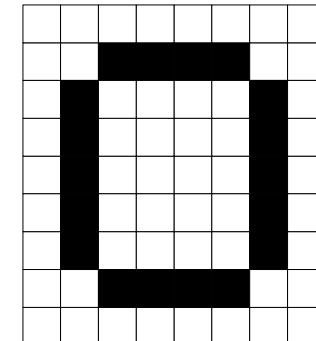
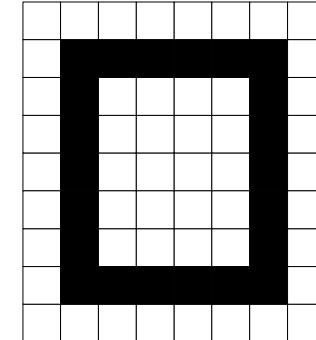
$$\{supp(I)\}^c = \{p = (x, y) \in I : I(p) = bg\}$$





## Pojecie łączności

- Co właściwie przedstawia obraz obok ?
  - wszyscy się zgodzą, że na jasnym jest czarny prostokąt
  - obraz podzielony jest na 3 części:
    - zewnętrzne
    - wnętrze
    - kontur
- A kolejny obraz ?
  - jeden obiekt – prostokąt bez rogów ?
  - 4 linie ?
- Wszystko zależy od przyjętego kryterium łączności
  - łączność poczwórnej,  
*(4-connectivity)*
  - łączność ośmiokrotna  
*(8-connectivity)*
- można wyobrazić inne kryteria łączności, w innych topologiach:





## Suma Minkowskiego

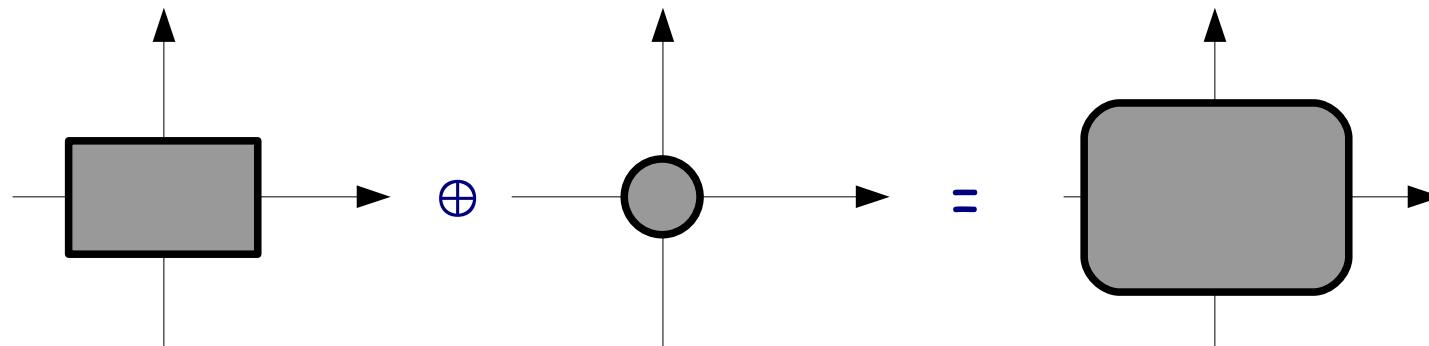
- określona na 2 zbiorach punktów (pikseli)

$$A \oplus B = \{p+q : p \in A \wedge q \in B\}$$

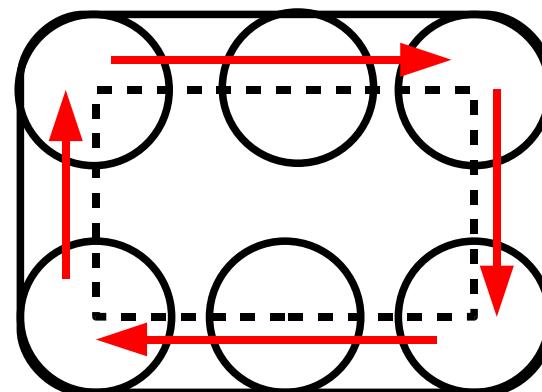
lub

$$A \oplus B = \{z \in \mathbb{R}^n : A \cap (z - B) \neq \emptyset\}$$

- przykład 2D



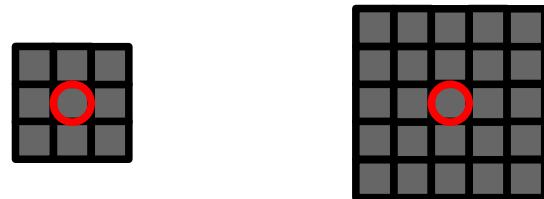
- sposób konstrukcji:





## Element strukturalny (SE, structural element)

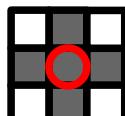
- w morfologii matematycznej operacje są dwuargumentowe:
  - pierwszym argumentem jest obraz do przetworzenia
  - drugim argumentem jest mniejszy obraz: element strukturalny
- element strukturalny jest dużej części zastosowań kwadratem o boku  $2N+1$  ( $3 \times 3$ ,  $5 \times 5$ ) i wszystkich wartościach '1' (*foreground*)



- jeden punkt/piksel jest wyróżniony i nosi nazwę punktu centralnego lub początkowego (*origin*) – nie musi on leżeć w środku (choć b. często leży)
- element strukturalny może mieć inny kształt i położenie punktu początkowego:



- w niektórych zastosowaniach pewne piksele mogą mieć wartość tła

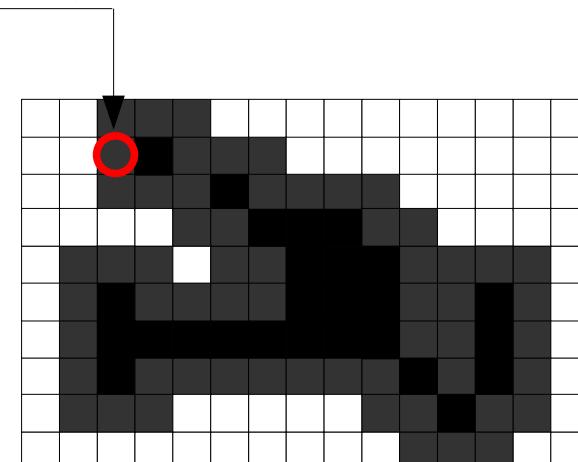
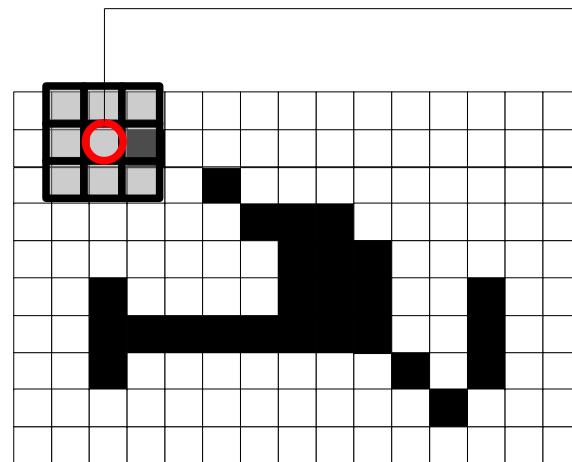




## Operator dylacji (dilation)

- operacja dylacji jest zastosowaniem sumy Minkowskiego do obrazów cyfrowych
- pierwszym argumentem jest obraz
- drugim element strukturalny
- zasada działania:
  - do każdego piksela obrazu przykłada się element strukturalny (SE) w jego punkcie centralnym
  - jeśli choć jeden piksel sąsiedztwa przykryty przez SE ma wartość '1', piksel bieżący też przyjmuje wartość '1' (obiektu pierwszoplanowego)

w otoczeniu zdefiniowanym przez SE jest co najmniej 1 piksel '1'

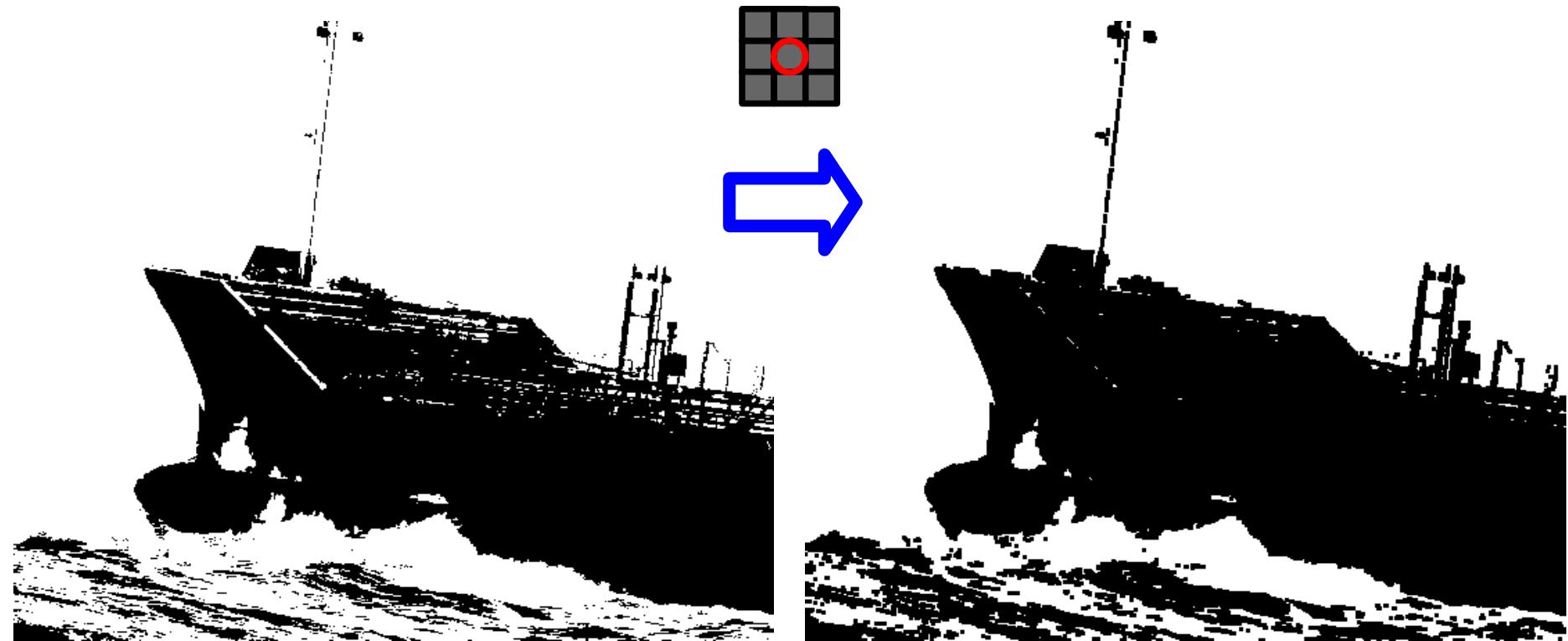




## Dylacja



- jednorodny SE  $3 \times 3$
- jednokrotna dylacja





## Dylacja wielokrotna

- obiekty zwiększają rozmiary
- zanikają detale
- „dziury” są wypełniane
- obiekt staje się lepiej widoczny - rośnie

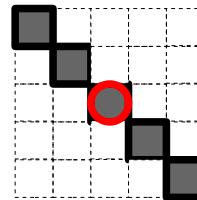
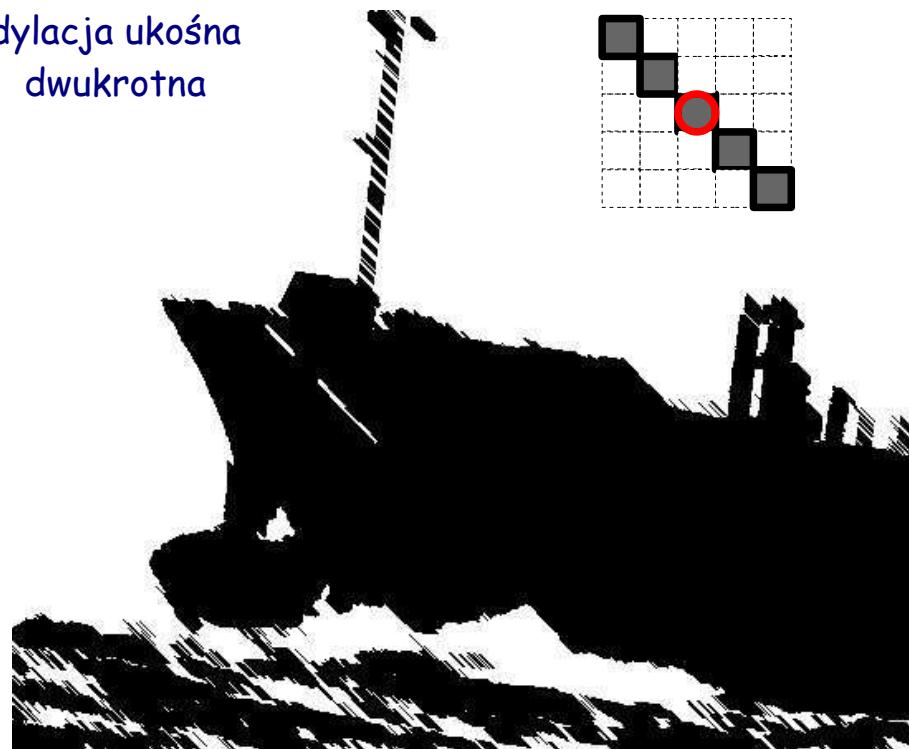
dylacja dwukrotna



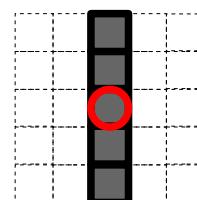
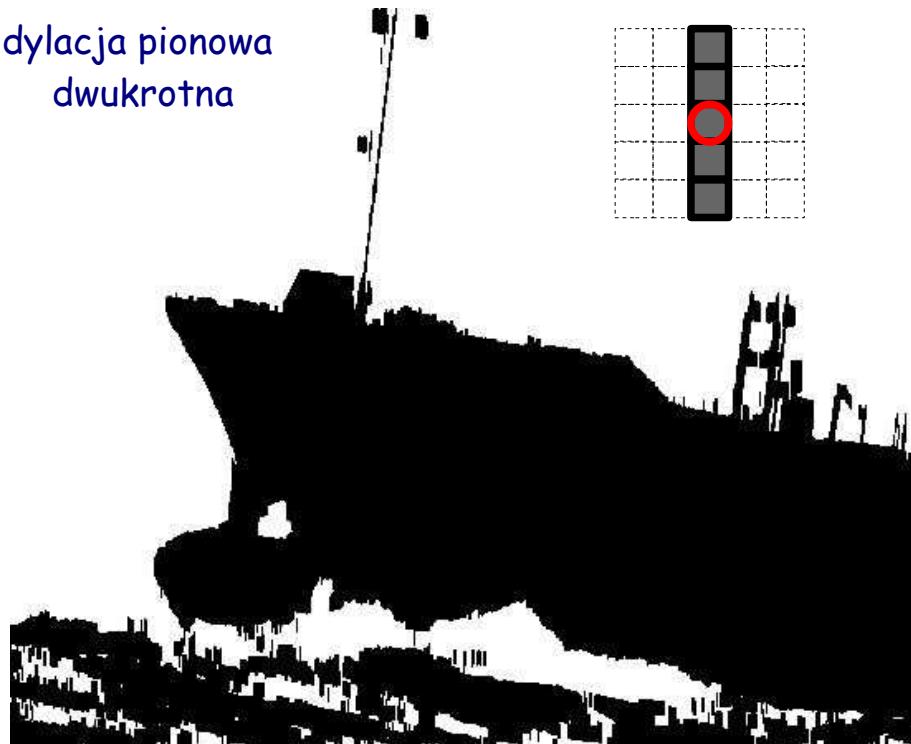
dylacja pięciokrotna



dylacja ukośna  
dwukrotna



dylacja pionowa  
dwukrotna

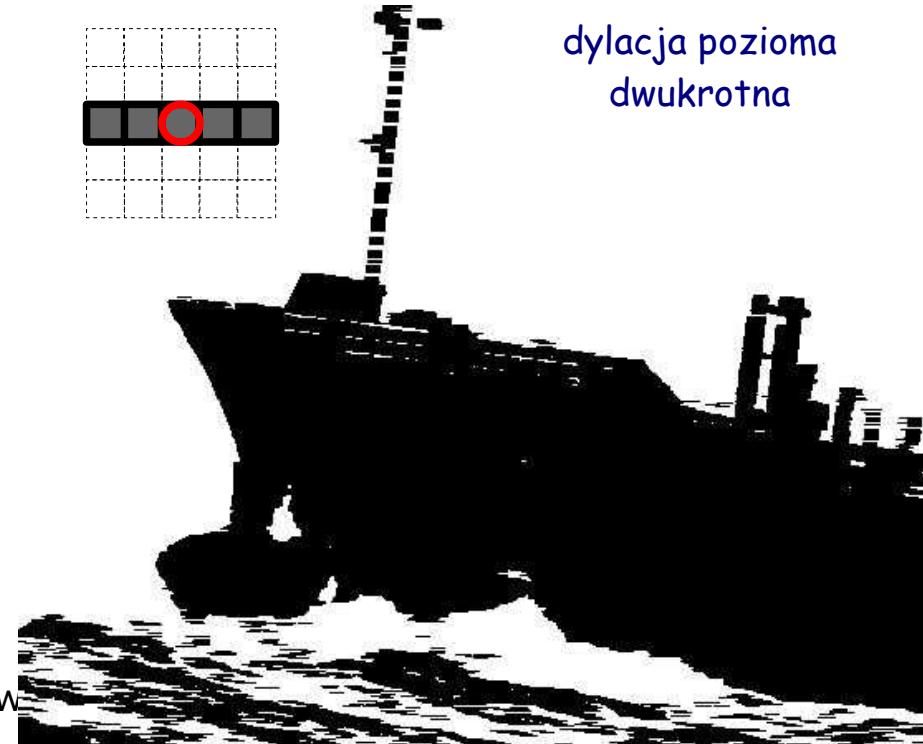


## Dylacja z niestandardowymi SE



- rozrost obiektu może być kierunkowy
- pozycja początkowa może nie leżeć w środku obszaru – rozrost niesymetryczny

dylacja pozioma  
dwukrotna

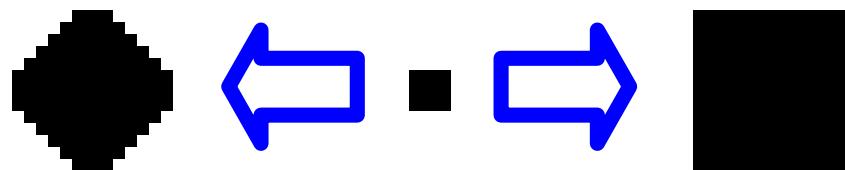


tw



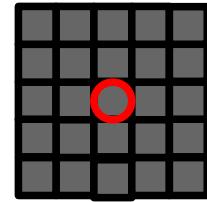
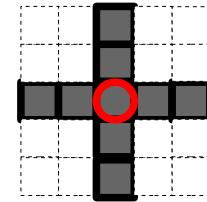
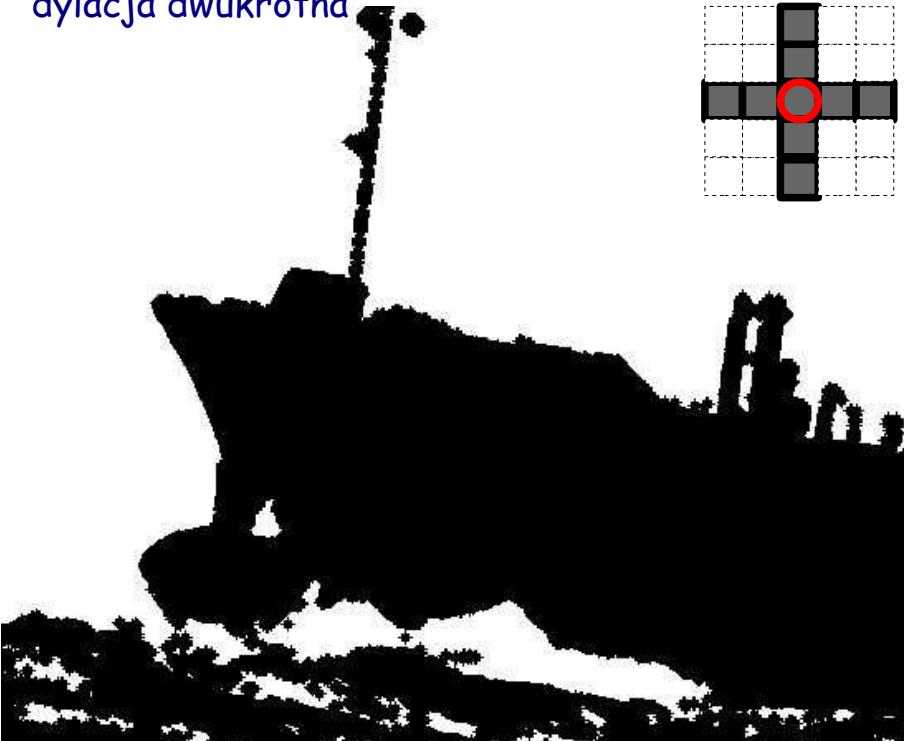
## Porównanie łączności 4-krotnej i 8-krotnej

dylacja pięciokrotna

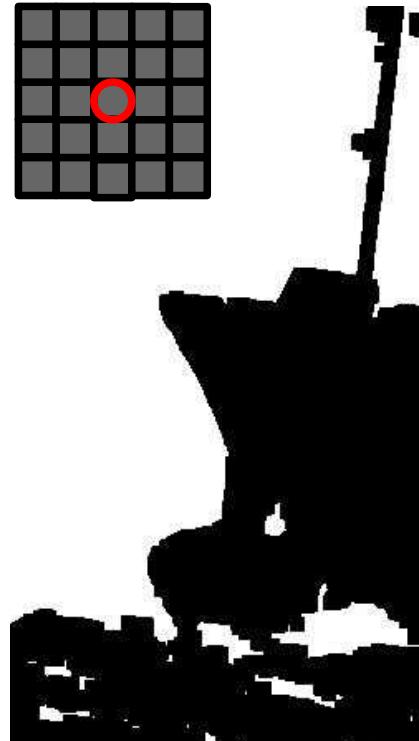


dylacja pięciokrotna

dylacja dwukrotna

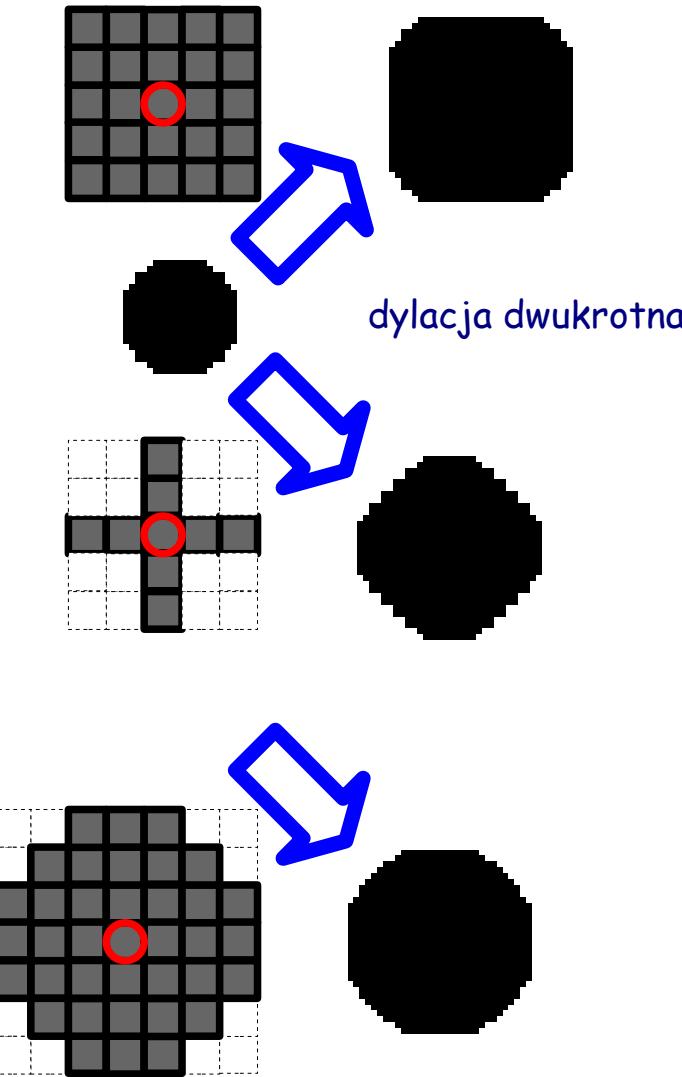


dylacja dwukrotna

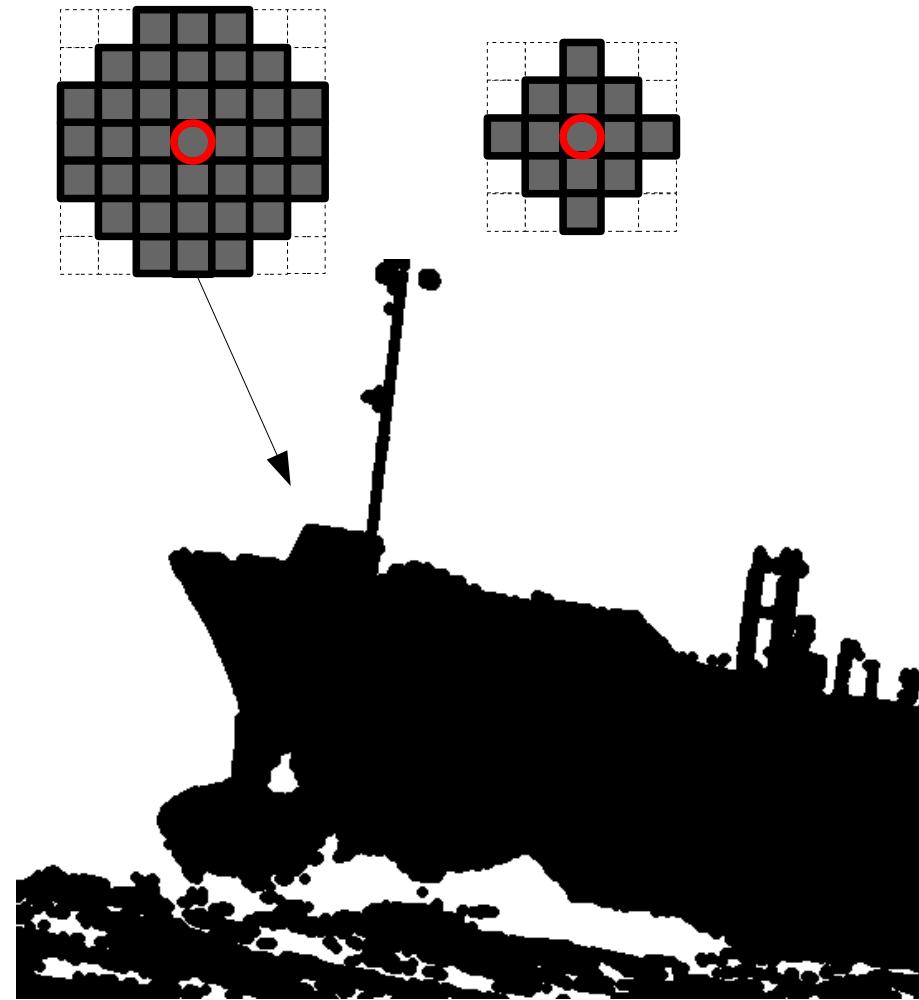




## Problemy z zachowaniem kształtów



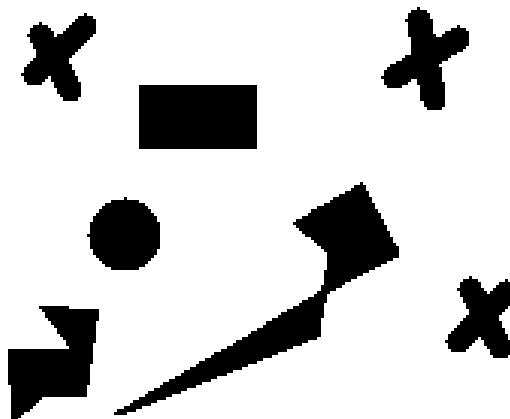
- duży wpływ kształtu SE na wynik dylacji
- neutralny SE – zbliżony do koła



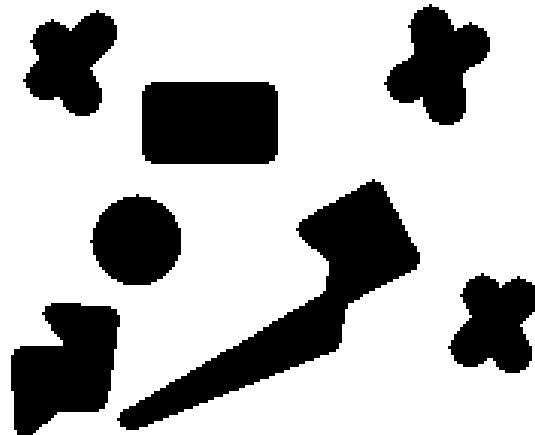


## Dylacja z „kołowym” SE

obraz oryginalny



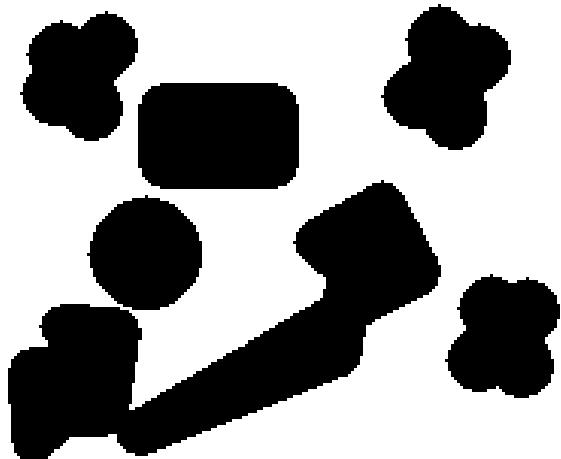
SE 7x7



SE 11x11



SE 15x15



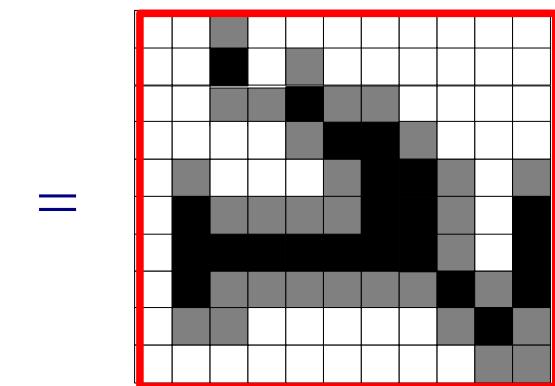
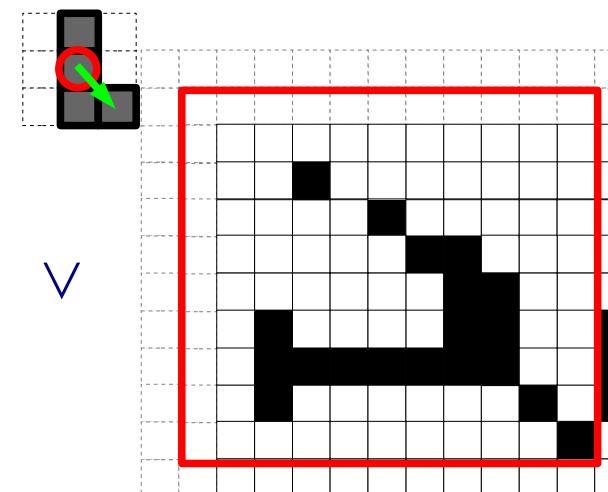
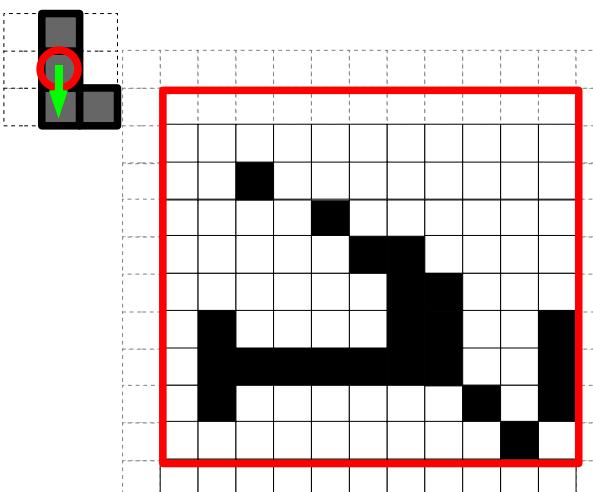
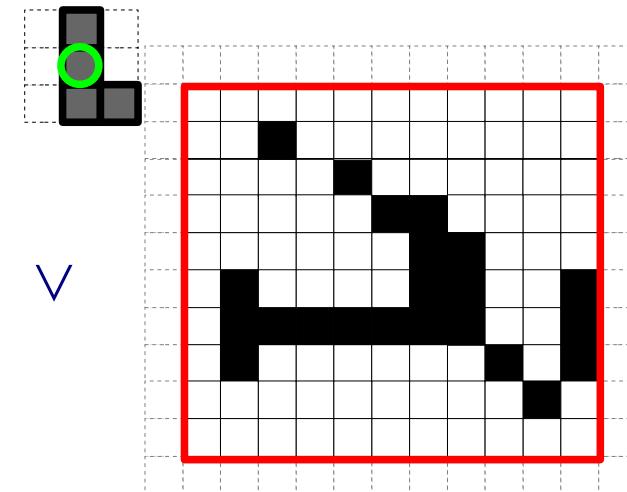
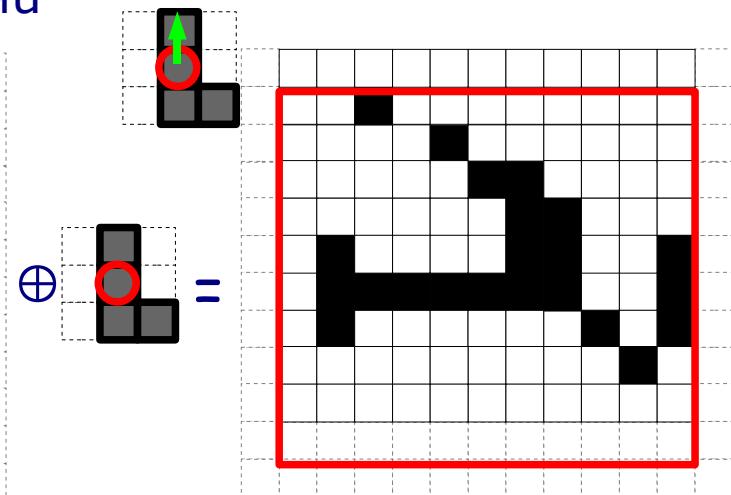
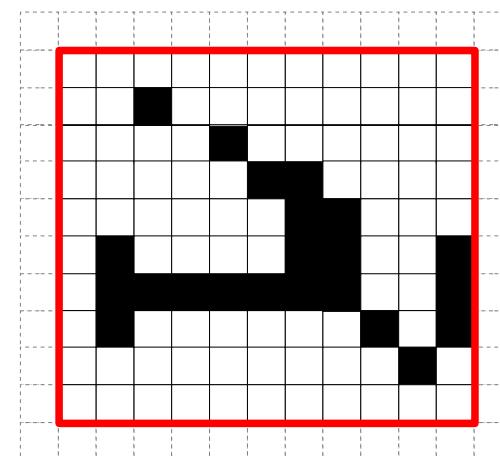
SE 21x21





## Efektywne liczenie dylacji

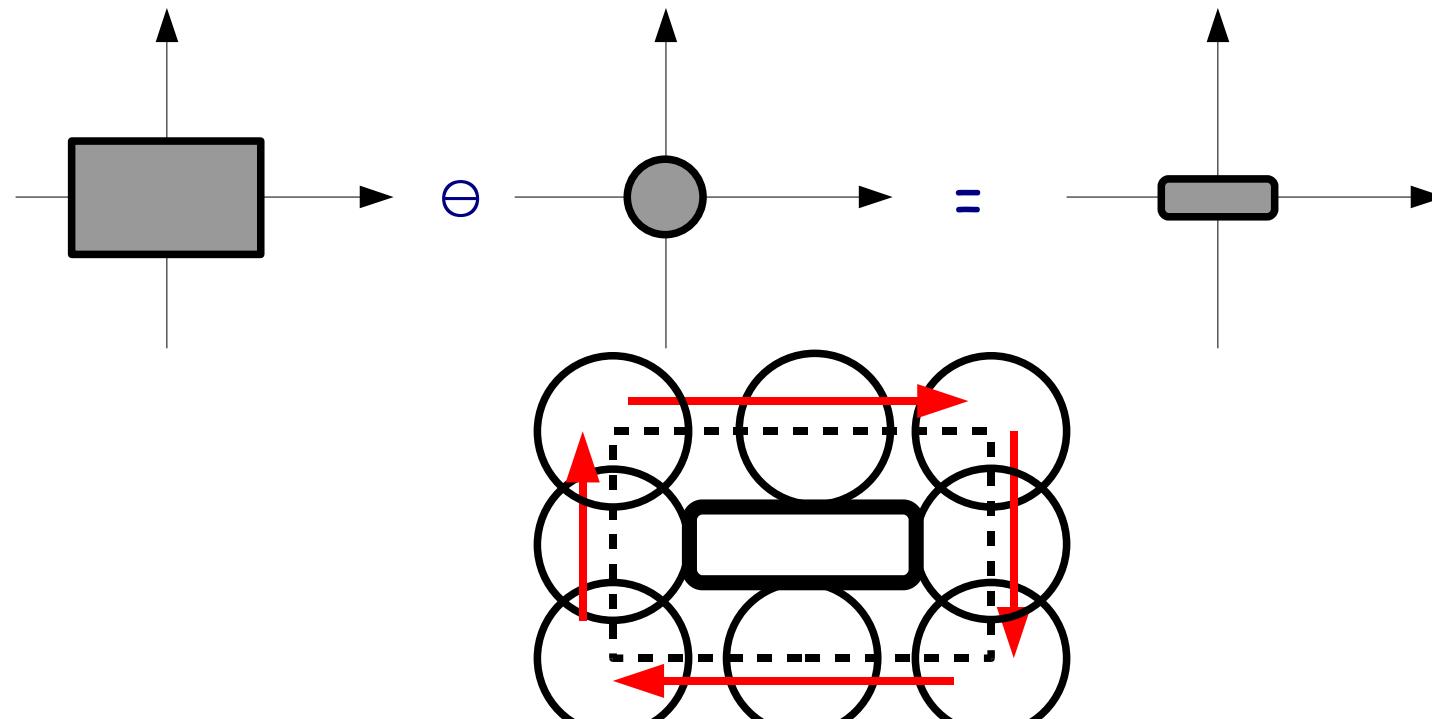
- schemat „przesuń i dodaj” (logiczne OR)
- obraz powiększa się o marginesy
- każdy piksel SE powoduje uwzględnienie jednego obrazu  $I'$  przesuniętego względem oryginału





## Operator erozji (erosion)

- analogiczny do dylacji
- pierwszym argumentem jest obraz, drugim element strukturalny
- zasada działania:
  - do każdego piksela obrazu przykłada się element strukturalny (SE) w jego punkcie centralnym
  - jeśli choć jeden piksel sąsiedztwa przykryty przez SE ma wartość '0', piksel bieżący też przyjmuje wartość '0' (tła)

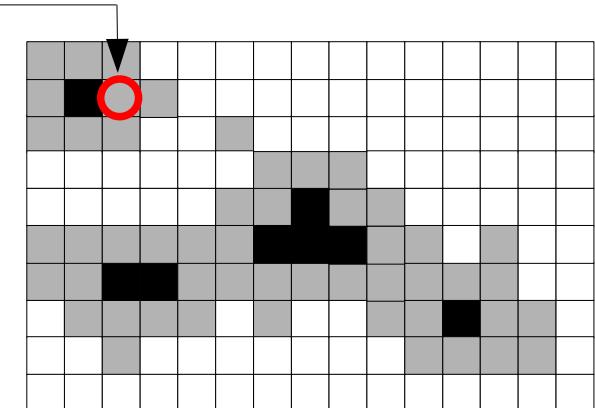
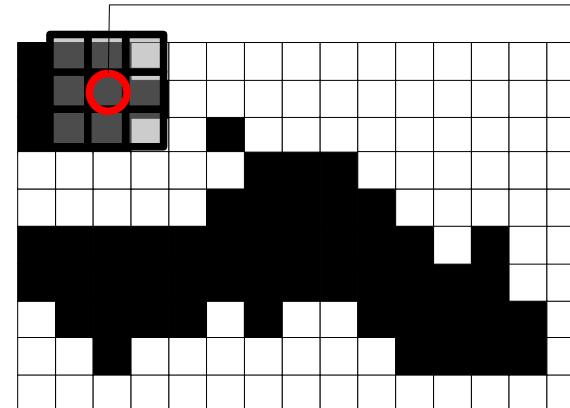




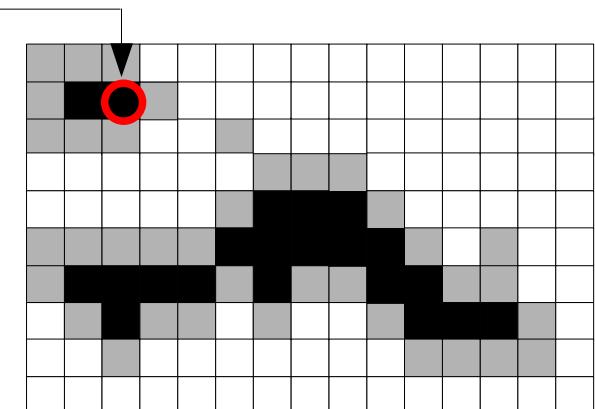
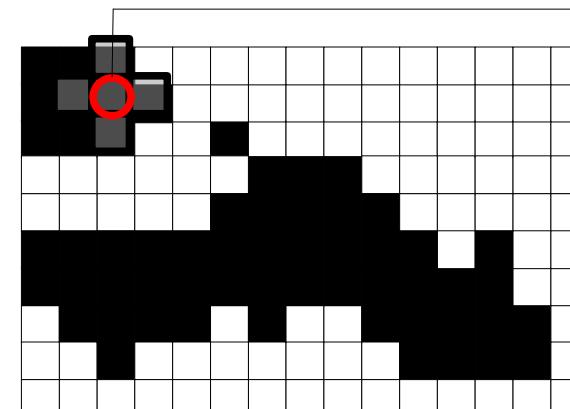
## Obliczanie erozji

- analogiczny do dylacji
- działanie zależy od kształtu elementu strukturalnego
  - łączność 4-krotna
  - łączność 8-krotna
- możliwość efektywnego obliczania erozji (jak dylacji): zamiast sumy logicznej (alternatywy) obrazów używamy iloczynu logicznego (konjunkcji)
- znika szum
- jednocześnie obiekt zmniejsza się
- niektóre gałęzie zanikają

w otoczeniu zdefiniowanym przez SE jest co najmniej 1 piksel '0'



w otoczeniu zdefiniowanym przez SE nie ma żadnego piksela '0'

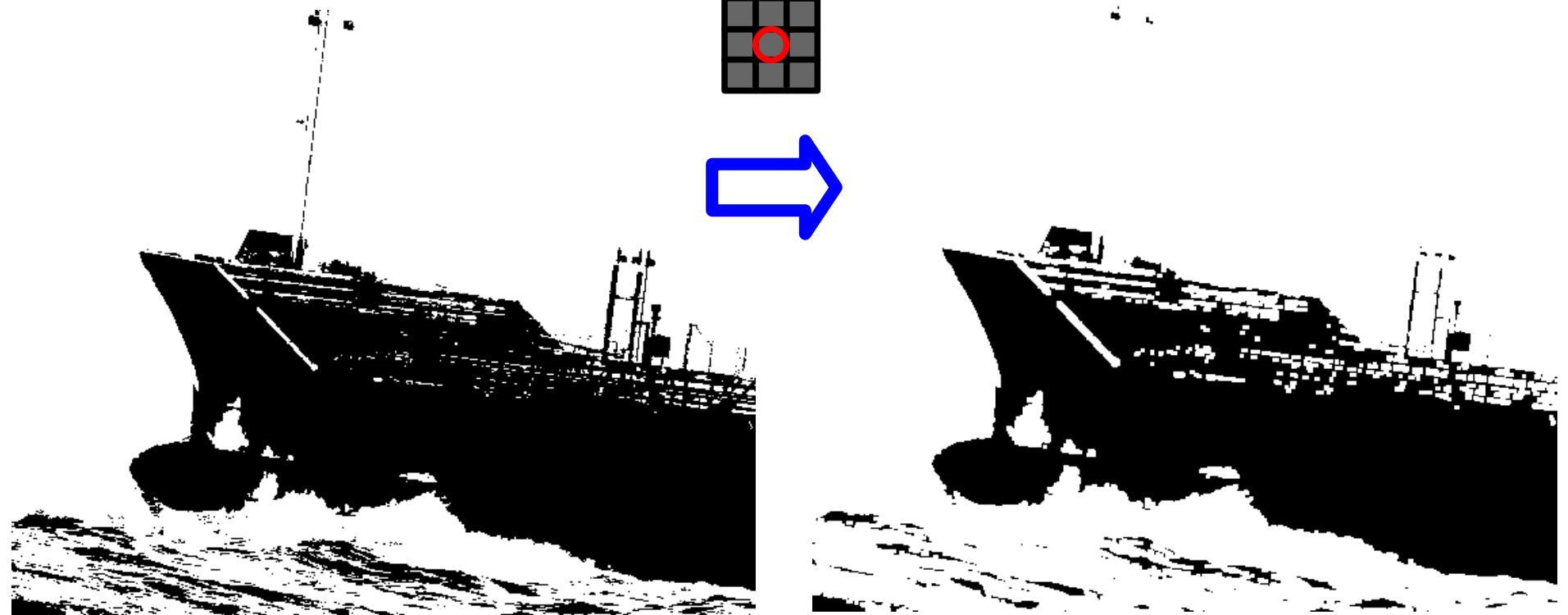
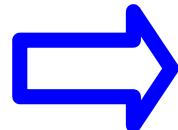
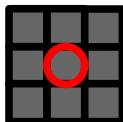




## Erozja



- jednorodny SE  $3 \times 3$
- jednokrotna erozja





## Erozja wielokrotna

- obiekty zmniejszają rozmiary
- również zanikają detale, ale inaczej niż w dylacji
- „dziury” rozszerzają się
- obiekt znika
- widać wpływ kwadratowego SE: dziury mają podobny kształt

erozja dwukrotna

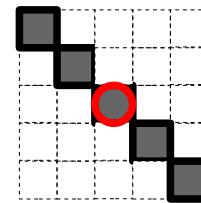


erozja pięciokrotna

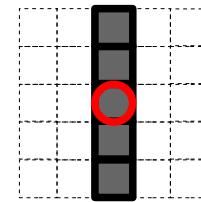




erozja ukośna  
dwukrotna



erozja pionowa  
dwukrotna

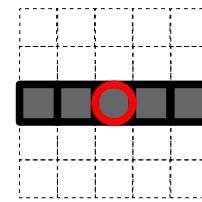


## Erozja kierunkowa

obraz oryginalny



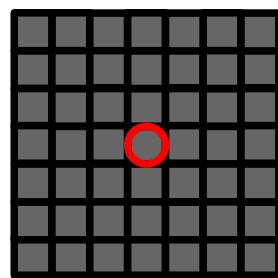
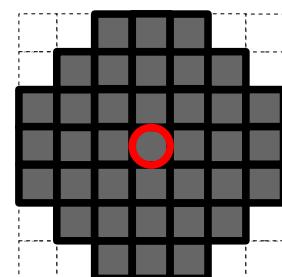
erozja pozioma  
dwukrotna





## Erozja z „kołowym” SE

- SE o kształcie zbliżonym do koła daje bardziej naturalne wyniki
- aby dobrze oddać kształt koła, SE musi mieć większy rozmiar
- większy rozmiar powoduje bardziej „drastyczną” erozję (dylację)





## Wyodrębnianie granic obiektu

obraz 1. - początkowy



obraz 2. - po dylacjji



(obraz 1. - obraz 2.) - granica zewnętrzna obiektu



# Wyodrębnianie granic obiektu



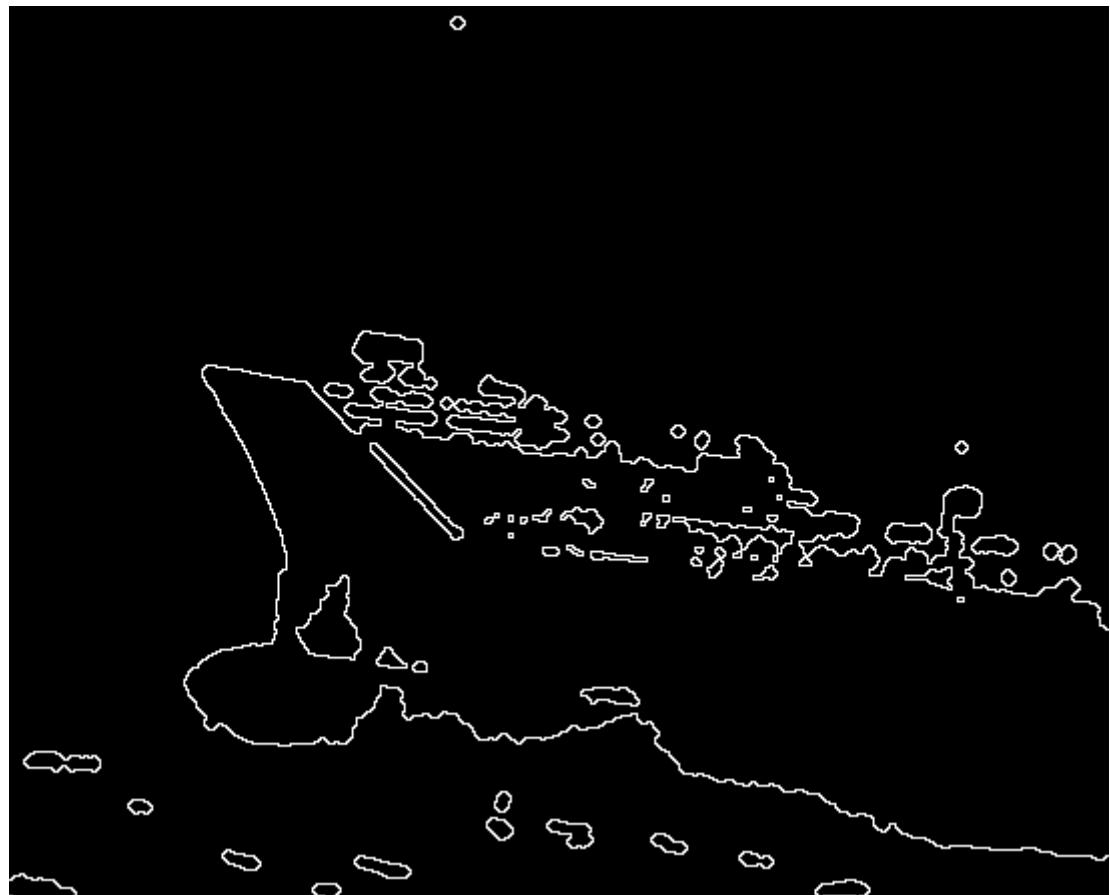
obraz 1. - początkowy



obraz 2. - po erozji



(obraz 1. - obraz 2.) - granica wewnętrzna obiektu





## Operacje złożone

- 2 poprzednie operacje (erozja i dylacja) zmieniały rozmiary obiektu
- miały też interesujące właściwości (usuwanie detali, szumu)
- 2 te operacje wykonane jedna po drugiej zachowują rozmiary obiektu:
  - **otwarcie morfologiczne (*opening*)** jest dylacją erozji

$$I \circ SE = (I \ominus SE) \oplus SE$$

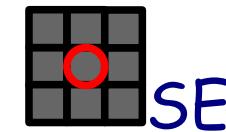
- **zamknięcie morfologiczne (*closing*)** jest erozją dylacji

$$I \bullet SE = (I \oplus SE) \ominus SE$$

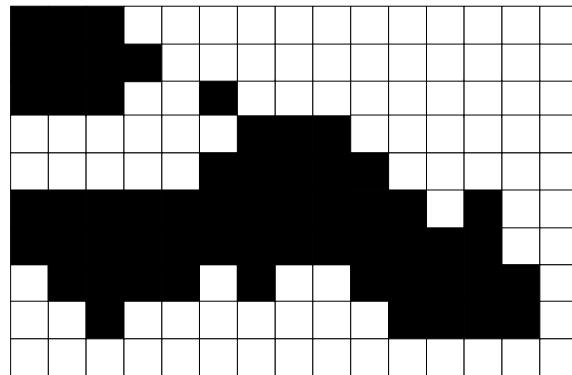
- elementy strukturalne w obu fazach operacji muszą być takie same



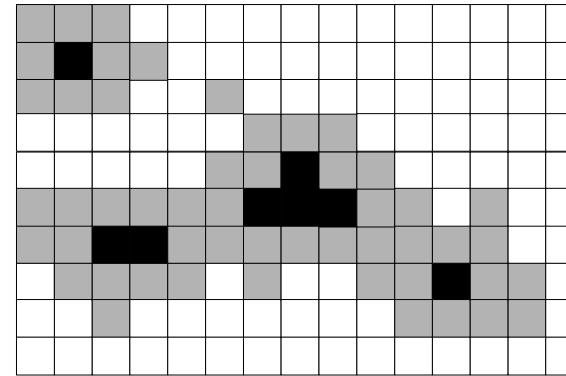
## Otwarcie morfologiczne



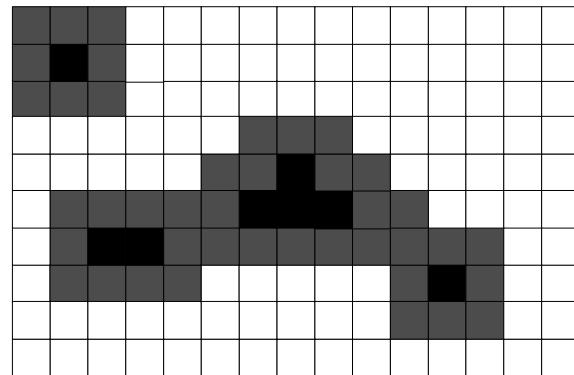
$I \circ SE = ($



$\ominus) \oplus SE =$



$\oplus SE =$

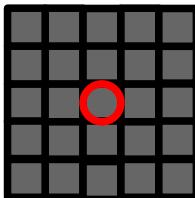
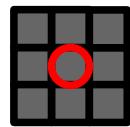


- obiekt zachował rozmiary
- wszystkie wystające fragmenty zostały usunięte
- właściwość idempotencji  
(kolejne otwarcia nie mają znaczenia):

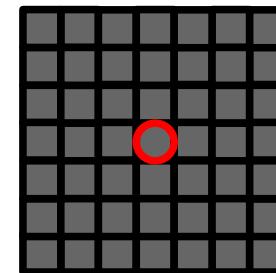
$$(I \circ SE) \circ SE = I \circ SE$$



## Wyniki otwarcia



- im większy rozmiar SE tym efekt usuwania detali silniejszy (znika więcej większych szczegółów)
- tym większy również wpływ kształtu SE (kwadratowe obszary)

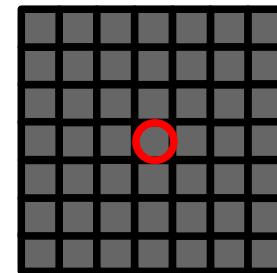
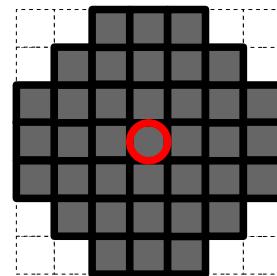




## Otwarcie z „kołowym” SE

- tu również kształt SE zbliżony do koła daje bardziej naturalne wyniki
- szczegóły podobnej wielkości są usuwane z obu wyników,  
ale pozostałe obszary mają inny kształt

•



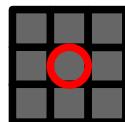


## Otwarcie „wielopoziomowe”

$$(I \ominus SE \ominus SE \ominus SE) \oplus SE \oplus SE \oplus SE$$

- kilkukrotne powtarzanie erozji a następnie dylacji ma ten same efekt co użycie większego SE

trzykrotna erozja, potem trzykrotna dylacja



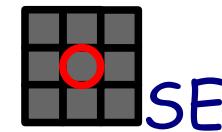
obraz oryginalny

obraz po otwarciu z SE 7x7

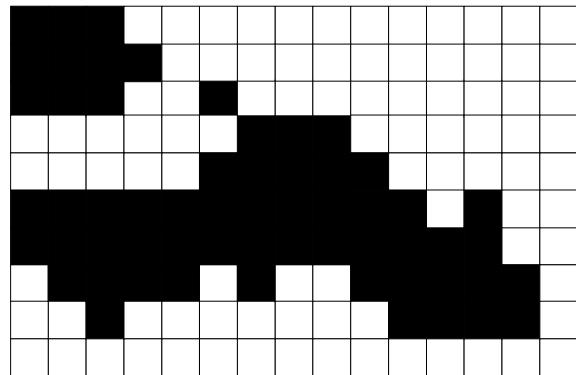




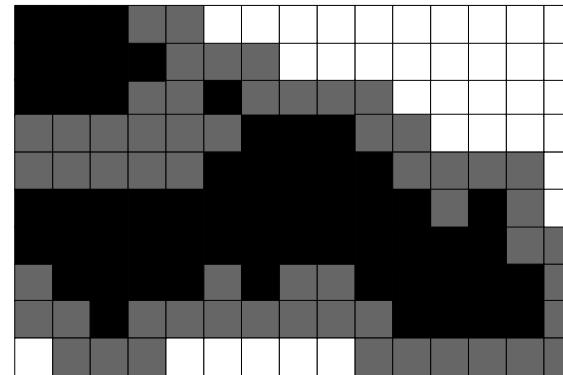
## Zamknięcie morfologiczne



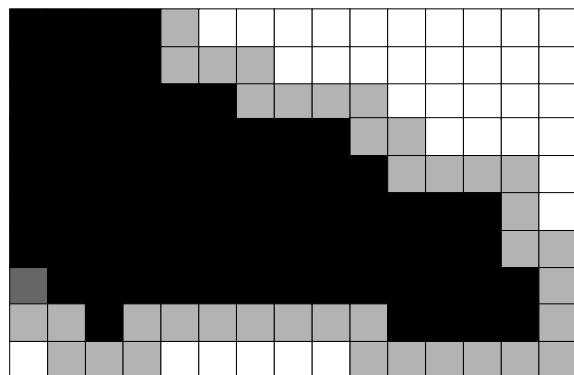
$I \bullet SE = ($



$\oplus) \ominus SE =$



$\ominus SE =$



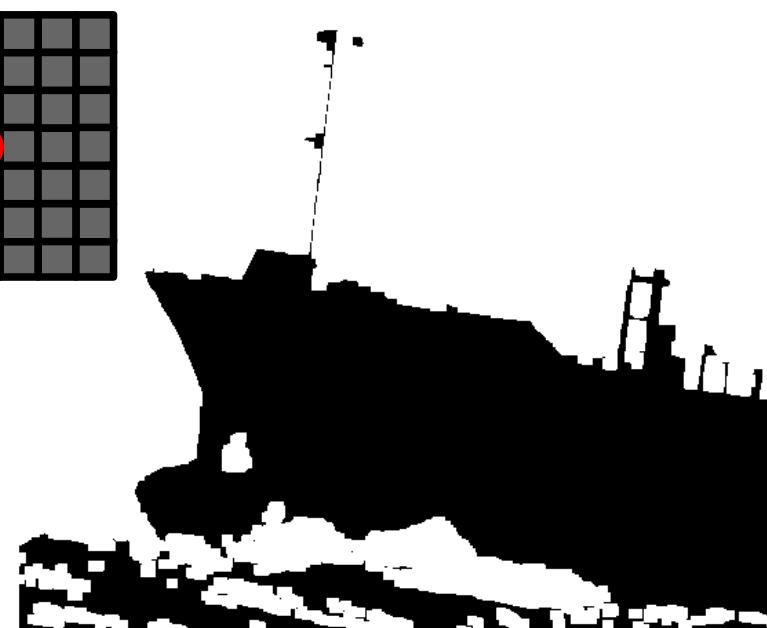
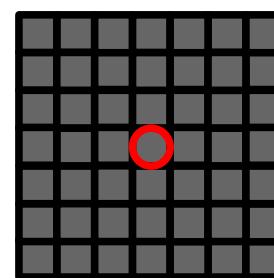
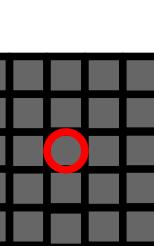
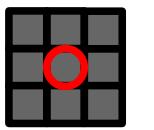
- obiekt zachował rozmiary
- wszystkie wklęsłe obszary zostały włączone do obiektu
- specjalne traktowanie brzegów (tu: obraz jest czasowo powiększany)
- właściwość idempotencji ( kolejne otwarcia nie mają znaczenia):

$$(I \bullet SE) \bullet SE = I \bullet SE$$



## Wyniki zamknięcia

- im większy rozmiar SE tym efekt usuwania detali silniejszy:
- tym większe „dziury” są wypełniane
- tym większy również wpływ kształtu SE (kwadratowe obszary)
- tym bliższe małe detale łączą się

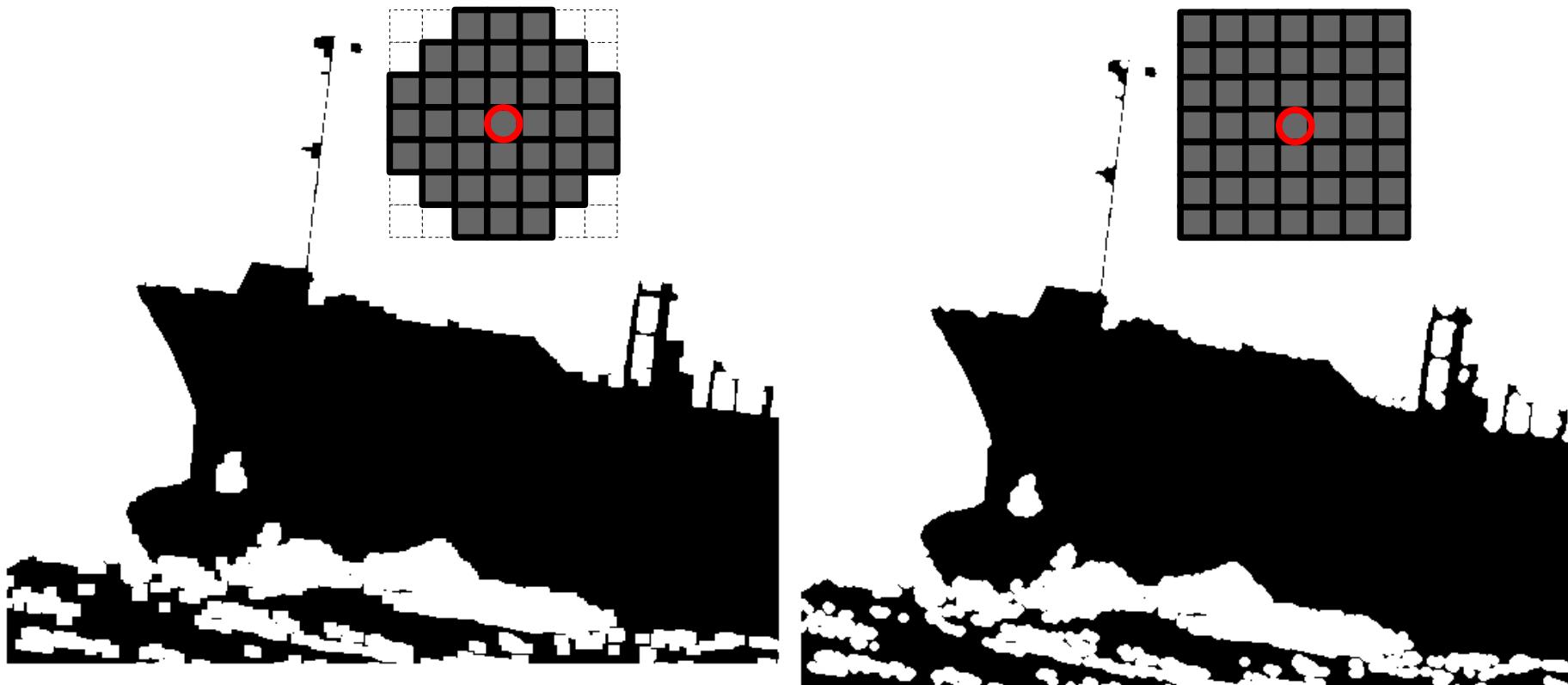


obraz oryginalny



## Zamknięcie z „kołowym” SE

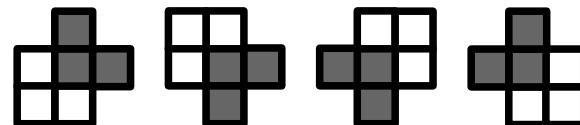
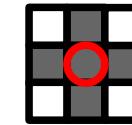
- tu również kształt SE zbliżony do koła daje bardziej naturalne wyniki
- szczegóły podobnej wielkości są usuwane z obu wyników,  
ale niewypełnione obszary mają inny kształt





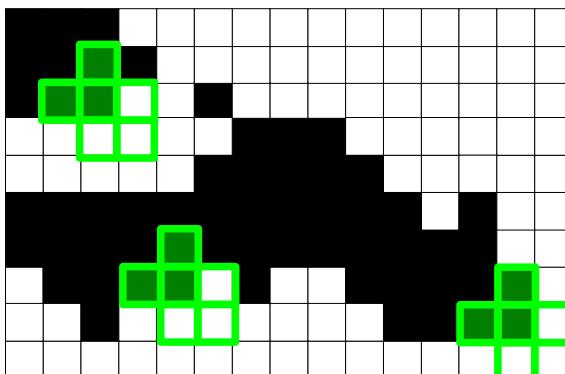
## Operacja „Hit-and-Miss” („Hit-or-Miss”)

- bazowa operacja morfologii matematycznej
- z niej można wywieźć wszystkie inne
- element strukturalny ma tu 3 rodzaje pól/pikseli:
  - piksele obiektu ('1')
  - piksele tła ('0')
  - piksele nieokreślone – niebrane pod uwagę
- obraz wejściowy jest przeglądany piksel po pikselu, jeśli konfiguracja pikseli sąsiednich jest identyczna jak w SE, jest on ustawiany na wartość '1' ('0'), w przeciwnym przypadku na '0' ('1')
- aby wykryć daną cechę wyniki kilku SE mogą się sumować/mnożyć

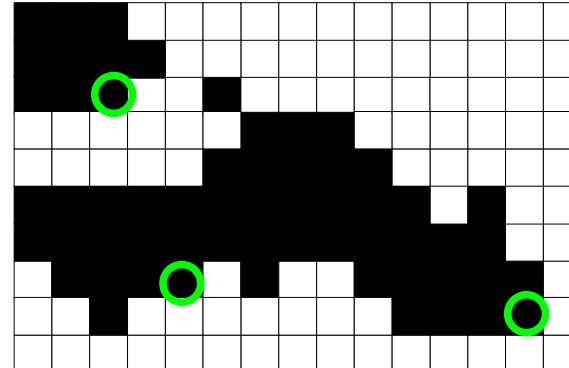


grupa SE wykrywająca rogi

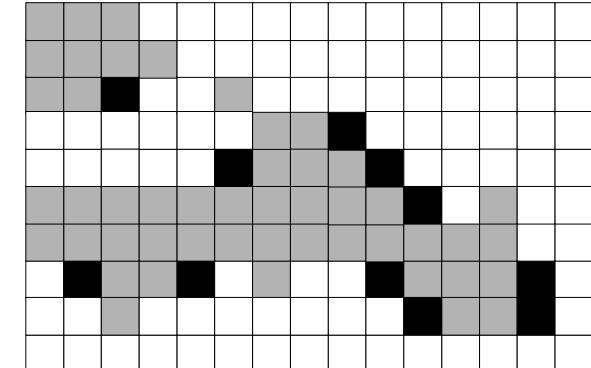
dopasowanie SE rogu dolnego-prawego



wykryte rogi dolne-prawe



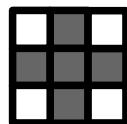
wykryte wszystkie rogi





## Operacja „Hit-and-Miss”

- rezultat operacji może być bardziej złożony, jeśli porównanie się powiedzie, analizowany piksel może:
  - przyjąć wartość '1' – operacje addytywne
  - przyjąć wartość '0' – operacje substraktywne
  - zachować swoją oryginalną wartość
- dla SE  $3 \times 3$  może być  $2^9 = 512$  różnych możliwych SE, dla  $5 \times 5$  –  $2^{25}$
- operacja **erozji**: przyporządkowanie wartości '1' dla jednej kombinacji SE (same piksele '1' w otoczeniu) i wartości '0' dla 511 pozostałych
- operacja **dylacji**: przyporządkowanie wartości '0' dla jednej kombinacji SE (same piksele '0' w otoczeniu) i wartości '1' dla 511 pozostałych
- zwykle jednej logicznej kombinacji/konfiguracji do wykrycia odpowiada wiele konkretnych SE  $\Rightarrow$  buduje się warunki logiczne które je zastępują:



$$X_0 \rightarrow X_0 \wedge X_1 \wedge \neg X_2 \wedge X_3 \wedge \neg X_4 \wedge X_5 \wedge \neg X_6 \wedge X_7 \wedge \neg X_8$$

usunięcie czarnego piksela ('1') otoczonego przez 8 białych ('0')

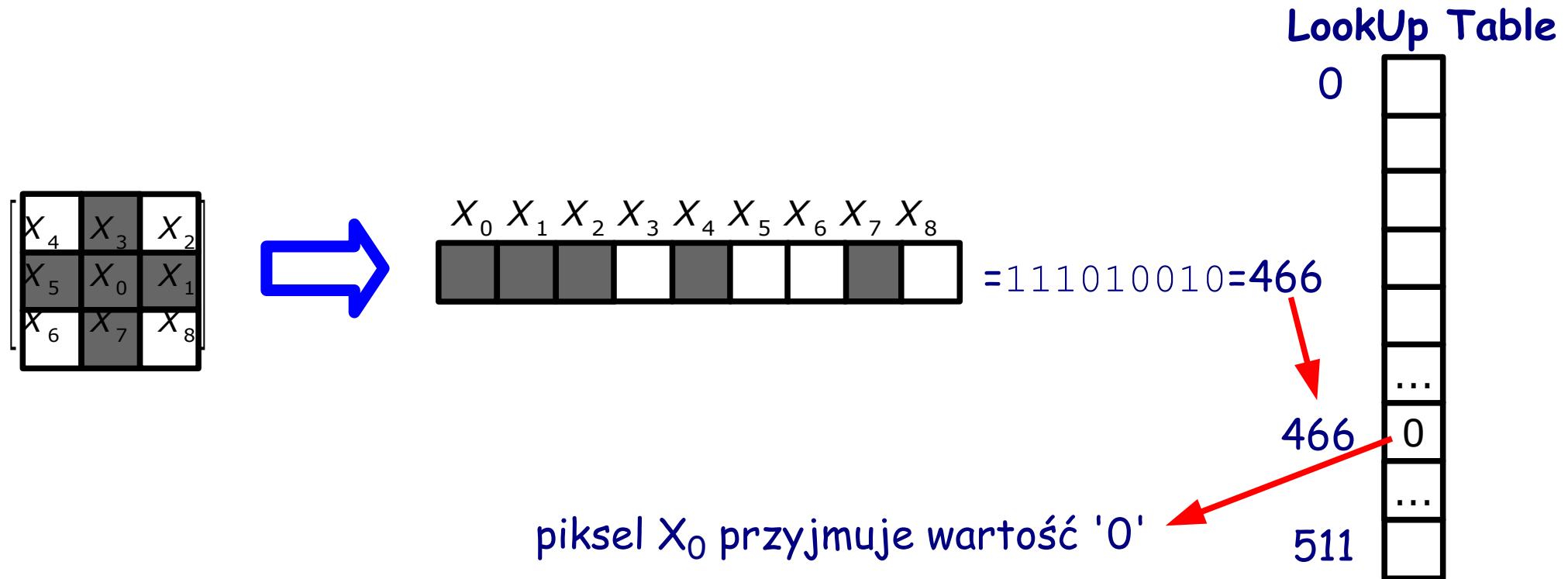
$$X_0 \rightarrow X_0 \wedge (X_1 \text{ lub } X_2 \text{ lub } X_3 \text{ lub } X_4 \text{ lub } X_5 \text{ lub } X_6 \text{ lub } X_7 \text{ lub } X_8)$$

$$\begin{bmatrix} X_4 & X_3 & X_2 \\ X_5 & X_0 & X_1 \\ X_6 & X_7 & X_8 \end{bmatrix}$$



## „Hit-and-Miss” za pomocą LUT

- każde blok obrazu kodowany jest za pomocą liczby od 0 do 511 wg pikseli traktowanych jako pozycje bitowe
- w tablicy LUT zapisany jest wynik operacji „Hit-and-Miss”
- znaczne przyspieszenie czasu operacji





## Operacje „większość białych” i „większość czarnych”

- w blokach 3x3 liczy się, ile jest pikseli czarnych i białych ('1' i '0')
- piksel centralny przyjmuje taką wartość, jaka jest większość w jego sąsiedztwie
- zbliżone w działaniu do erozji i dylacji



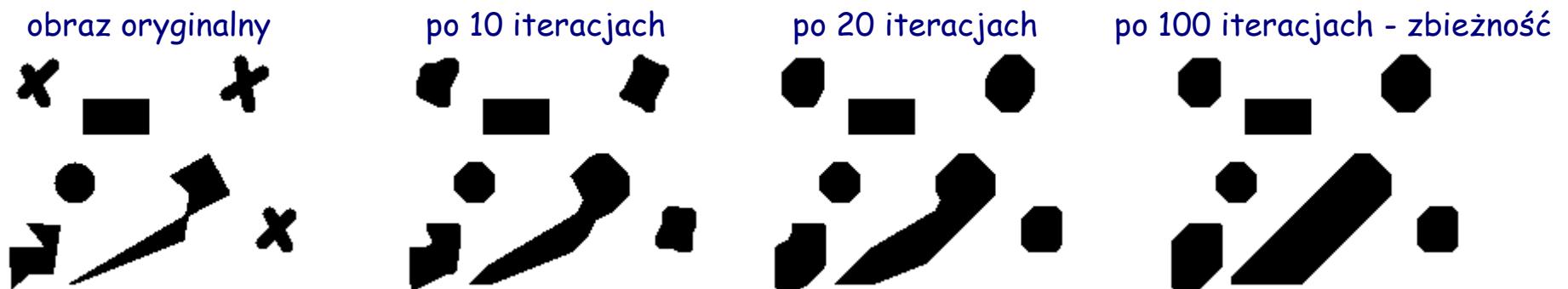
## Operacje pogrubiania (thickening, fattening)

- obiekt jest zwiększany, podobnie jak w dylacji i otwarciu
- nakładanych jest szereg warunków, aby oddzielne obiekty nie połączyły się
- oparty jest na operacji „hit-and-miss”  
$$\text{pogrubienie}(I, SE) = I \cup \text{hit-and-miss}(I, SE)$$
- zamiast jednego SE używana jest ich cała kolekcja
- piksel centralny SE zawsze jest równy zero
- czasami wyróżnia się 2 podoperacje, z angielskiego nazywane:
  - *fattening* – proste pogrubianie, oparte na kilku SE, w pewnych konfiguracjach obiekty mogą się połączyć
  - *thickening* – bardziej zaawansowana technika, pozabawiona wad poprzedniej



## Operacja pogrubiania - wypukły kształt (convex hull)

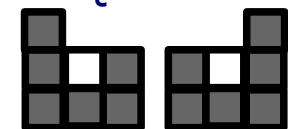
- celem jest znalezienie wypukłej otoczki obiektów
- używanych jest 8 SE:
  - 2 wzorcowe:
  - po 3 SE z każdego wzorcowego obrócone o 90, 180, 270 stopni
- procedura iteracyjna: po osiągnięciu zbieżności kolejne iteracje nie zmieniają wyniku (procedura zatrzymuje się sama)



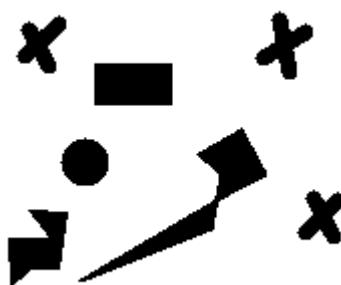


## Operacja pogrubiania - szkielet strefy wpływu (SKIZ - skeleton by zone of influence)

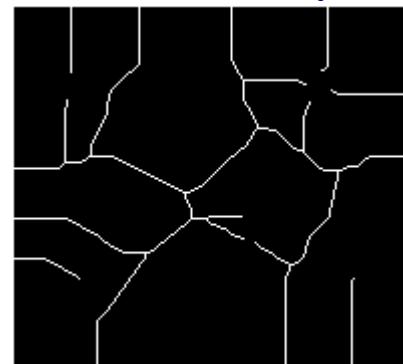
- celem jest znalezienie rozgraniczenia obiektów
- inna nazwa – **diagram Voronoi**
- procedura 2-etapowa:
  - rozrost obiektów – 8 SE powstały z obrotów 2S E:  
(aby uniknąć połączenia 2 obiektów każdy z 8 SE jest analizowany o oddzielnym przejściu przez wszystkie piksele obrazu !!!)
  - pruning – dalszy rozrost aby zlikwidować wszystkie niezamknięte kontury  
(na tej samej zasadzie – 8 SE powstałe z 2 następujących):
- procedura złożona obliczeniowo
- po osiągnięciu zbieżności kolejne iteracje nie zmieniają wyniku  
(procedura zatrzymuje się sama)



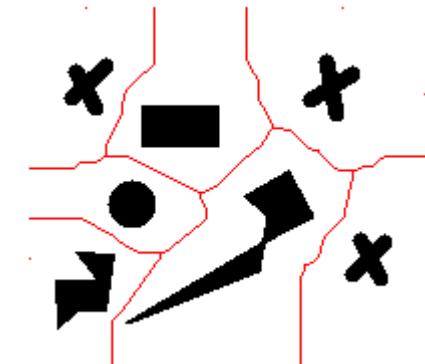
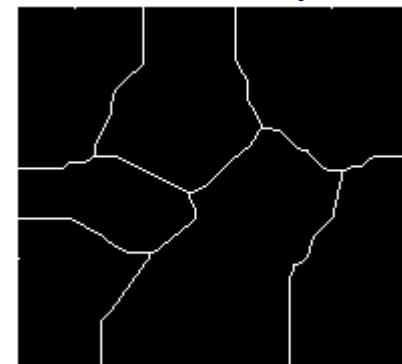
obraz oryginalny



rozrost obiektów  
(ok 130 iteracji)



pruning  
(ok 30 iteracji)



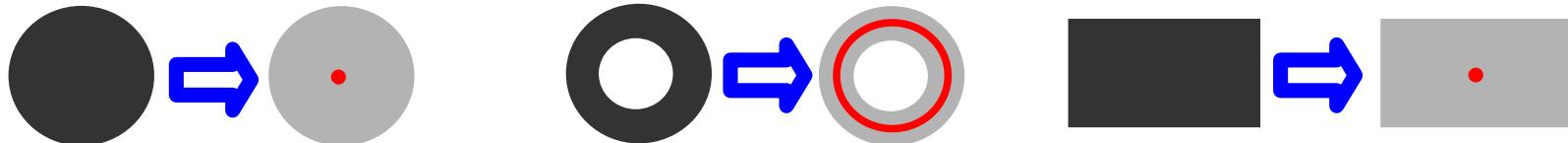


## Operacje pocieniania (thinning)

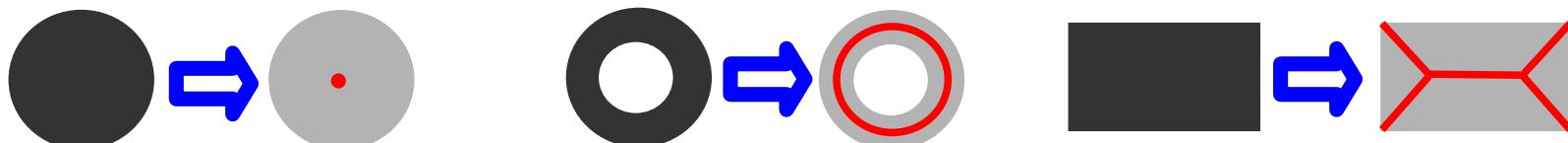
- obiekt jest zmniejszany, podobnie jak w erozji i zamknięciu
- nakładanych jest szereg warunków, aby obiekty nie zanikły
- oparty jest na operacji „hit-and-miss”

$$\text{pocienienie}(I, SE) = I - \text{hit-and-miss}(I, SE)$$

- zamiast jednego SE używana jest ich cała kolekcja
- piksel centralny SE zawsze jest równy jeden
- czasami wyróżnia się 2 podoperacje, z angielskiego nazywane:
  - *shrinking* – obiekty zanikają bez dziur zanikają do pojedynczego punktu obiekty z dziurami do zamkniętych krzywych



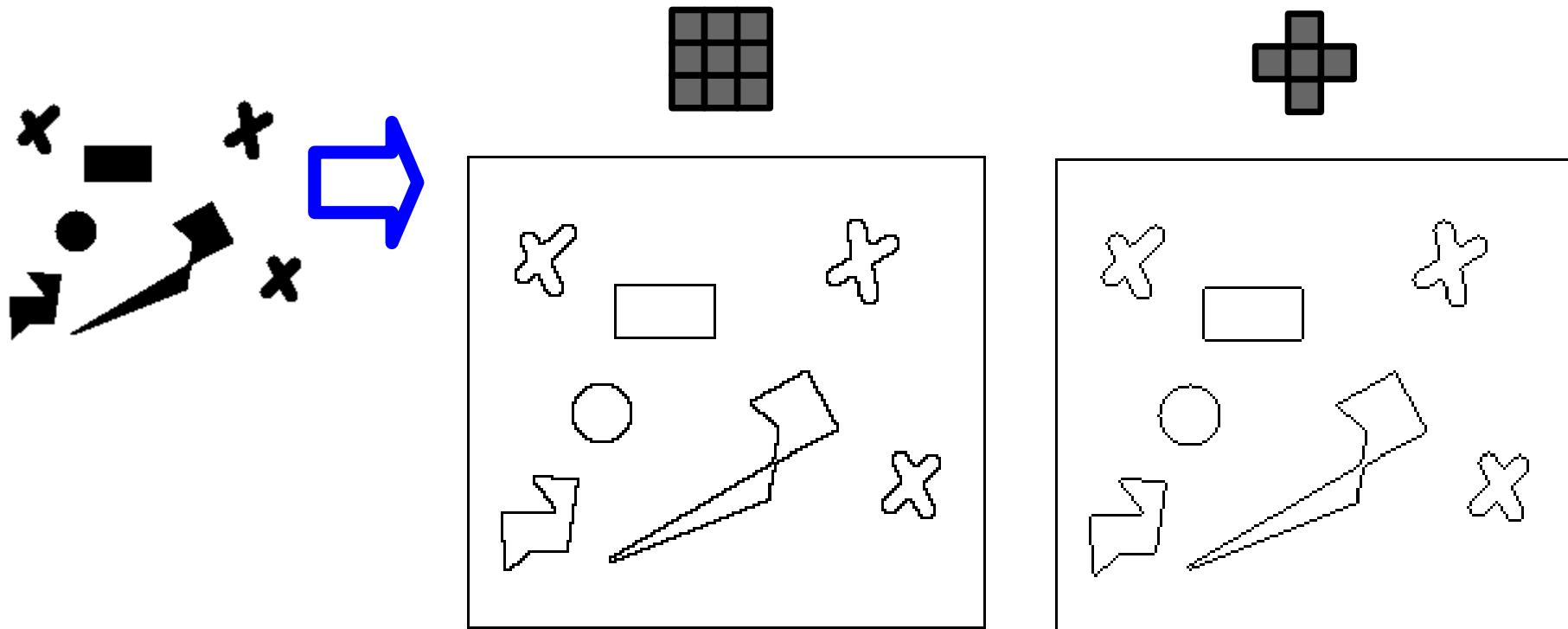
- *thinning* – obiekty zanikają do swoich szkieletów (szkieletyzacja)





## Operacje pocieniania - znajdowanie krawędzi obiektu

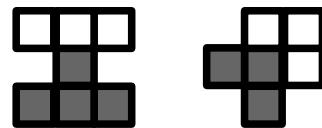
- celem jest znalezienie otoczki obiektu
- iteracyjne zmniejszanie obiektów 2 SE do wyboru (łączność 4-krotna lub 8-krotna)
- po osiągnięciu zbieżności kolejne iteracje nie zmieniają wyniku (procedura zatrzymuje się sama)



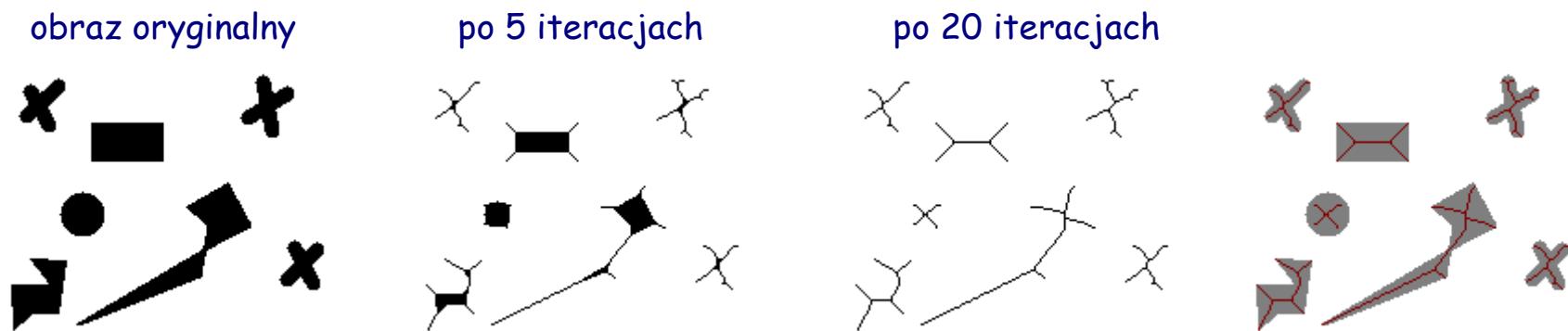


## Operacje pocieniania - znajdowanie szkieletu obiektu

- celem jest znalezienie szkieletu obiektu – linii reprezentujących obiekt
- iteracyjne zmniejszanie obiektów – 8 SE powstałych z obrotów 2 SE:



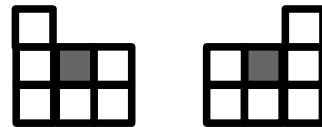
- aby uniknąć jednoczesnego usunięcia 2 ostatnich pikseli przez 2 różne SE, każdy z 8 SE jest stosowany w oddzielnym przejściu przez wszystkie piksele obrazu !!! (każda iteracja kryje w sobie 8 kolejnych iteracji)
- po osiągnięciu zbieżności kolejne iteracje nie zmieniają wyniku (procedura zatrzymuje się sama)



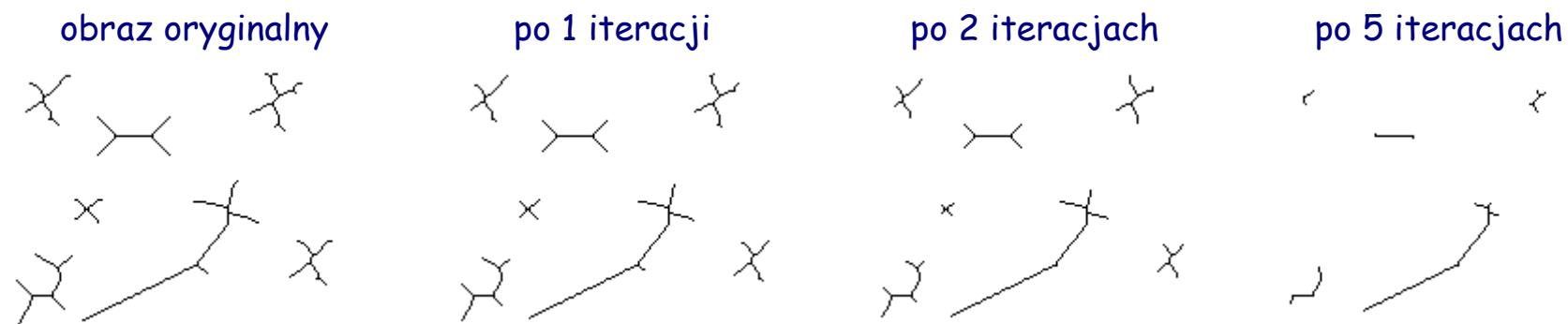


## Operacje pocieniania - pruning (przycięcie)

- celem jest znalezienie oczyszczenie szkieletu z krótkich „odnóg”
- iteracyjne zmniejszanie obiektów – 8 SE powstałych z obrotów 2 SE:



- aby uniknąć jednoczesnego usunięcia 2 ostatnich pikseli przez 2 różne SE, każdy z 8 SE jest stosowany w oddzielnym przejściu przez wszystkie piksele obrazu !!! (każda iteracja kryje w sobie 8 kompletnych iteracji)
- procedura wykonywana określoną liczbę kroków – dostateczna liczba iteracji usunie wszystkie otwarte linie !!!



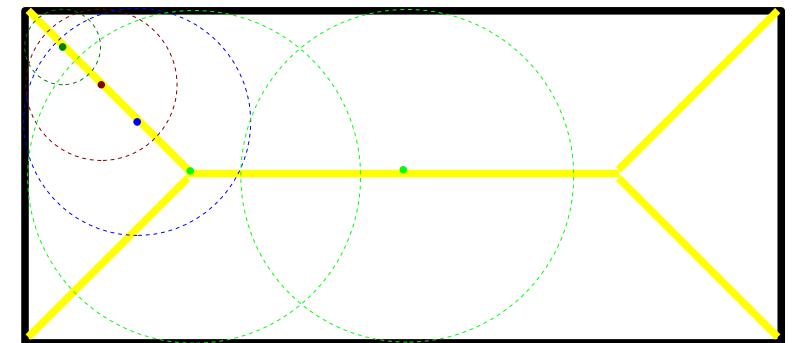


## Szkieletyzacja

- szkielet jest to liniowa reprezentacja obiektu
- linie powinny dobrze obrazować wnętrze obiektu (leżeć w równej odległości od brzegów)
- można sobie wyobrazić: obiekt powoli spalany począwszy od wszystkich swoich brzegów – punkty, do których linia ognia dojdzie z co najmniej 2 kierunków uznaje się za punkty szkieletu:



- inna interpretacja:  
zbiór środków okręgów stycznych  
w co najmniej 2 punktach do krawędzi obiektu:

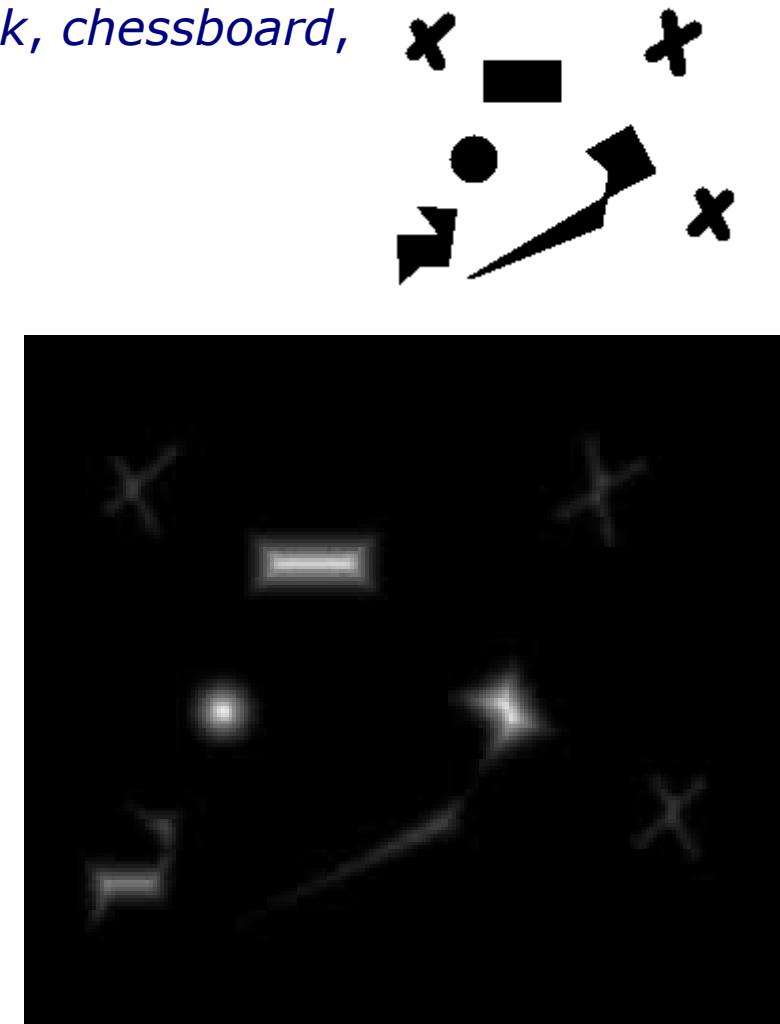


- w zasadzie są 2 rodzaje szkieletu:
  - **szkielet binarny** – obliczany przez pocienianie obiektu
  - szkielet, którego piksele mają wartość proporcjonalną do odległości od krawędzi obiektu – **MAT (Medial Axis Transform)** – obliczany jest przez **transformatę odległościową**



## Transformata odległościowa

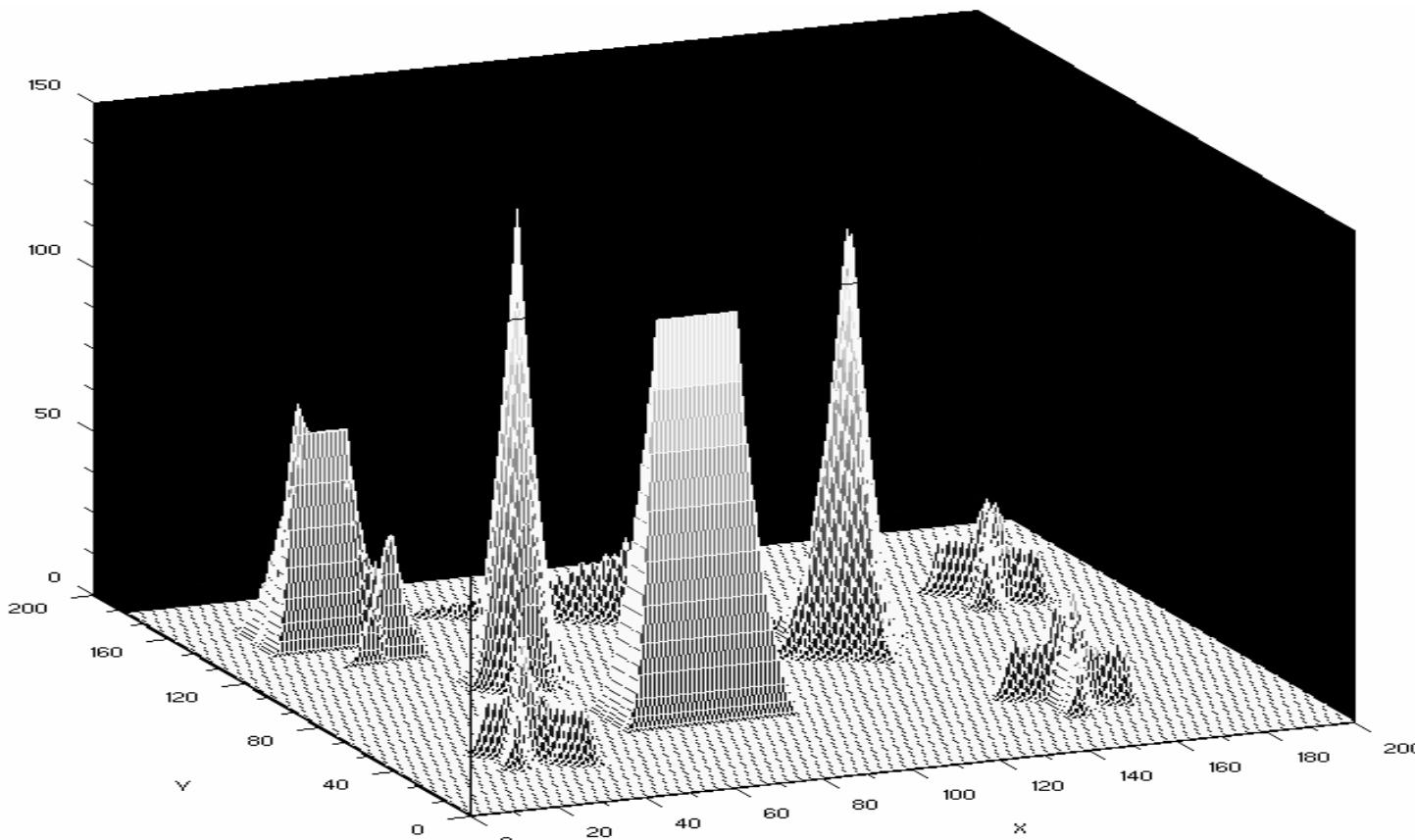
- jej celem jest podanie dla każdego piksela obiektu odległości do najbliższej krawędzi obiektu
- możliwe są różne metryki: euklidesowa, *city block*, *chessboard*,
- różne algorytmy:
  - proste, oparte na erozji:
    - nr iteracji która usuwa dany piksel jest jego odlegością do krawędzi
    - metryka zależy od użytego SE:
      - SE 3x3 złożony z jedynek – metryka *chessboard*
      - SE 3x3 w kształcie krzyża – metryka *city block*
      - SE zbliżony do koła – metryka euklidesowa
  - bardziej zaawansowane algorytmy:
    - *Fast Marching* (z teorii *Level Sets*) – jedna iteracja
    - Rosenfeld and Pfaltz 1968 – 2 iteracje – rekursywna morfologia





## Wyznaczanie MAT z transformaty odległościowej

- transformatę odległościową można przedstawić jako płaszczyznę 3D (współrzędna Z jest wartością transformaty – odlegością do krawędzi)
- krawędzie tej płaszczyzny wyznaczają szkielet binarny (można je zlokalizować licząc krzywiznę transformaty)
- wartości MAT są wartościami transformaty w miejscu krawędzi

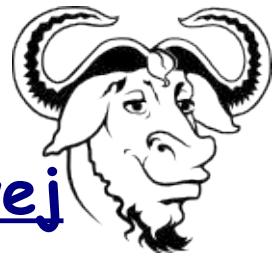


obraz oryginalny



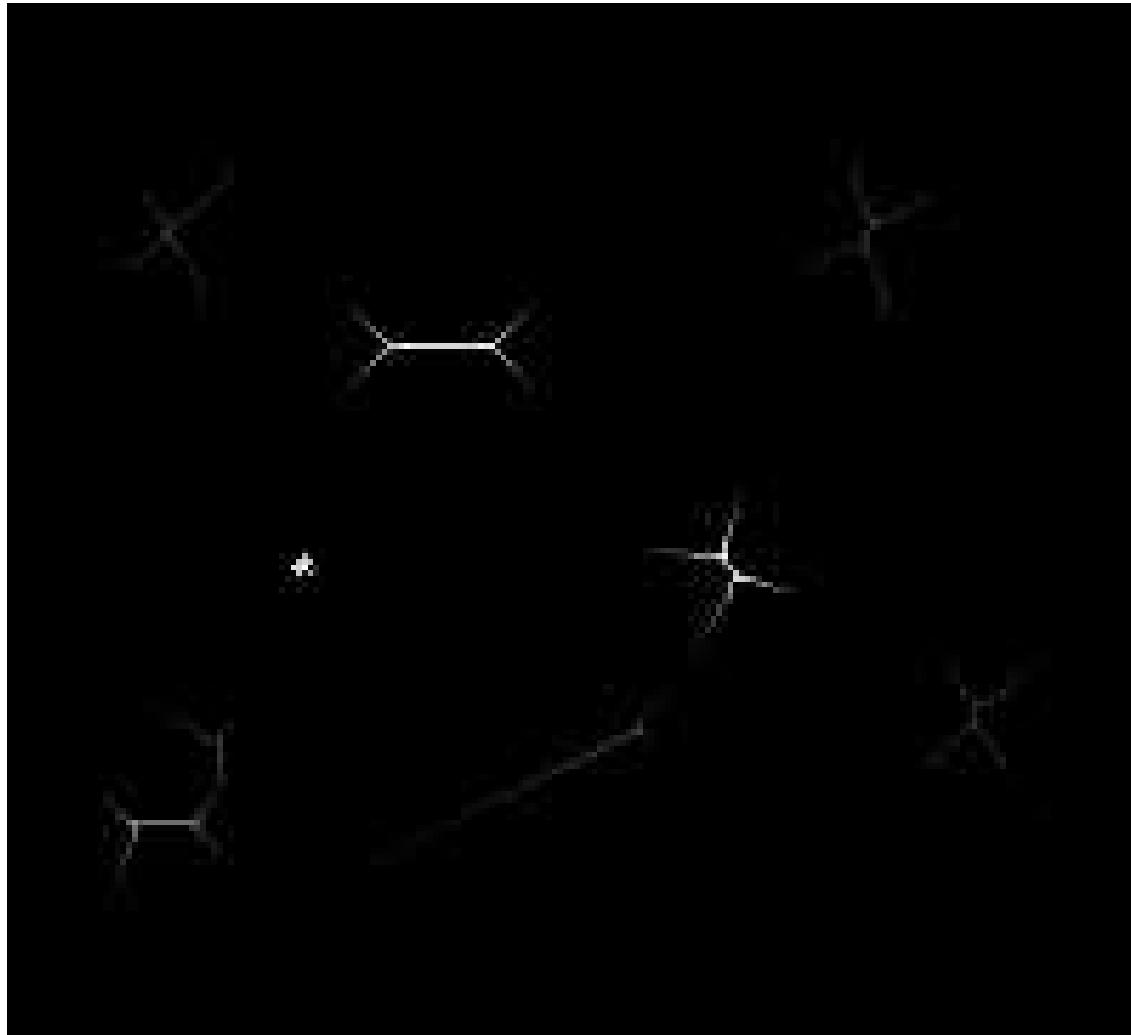
transformata odległościowa



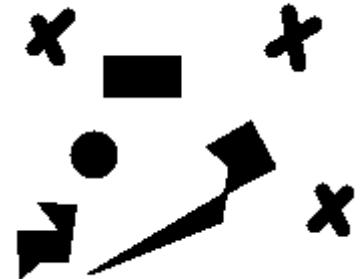


## Wyznaczanie MAT z transformaty odległościowej

szkielet MAT, jako krawędzie transformaty odległościowej



obraz oryginalny



transformata odległościowa





## Szczegółowość szkieletu

- Scilab oblicza szkielet z zaznaczeniem odległości od zakończeń szkieletu:
- poprzez progowanie można dopasować poziom szczegółowości szkieletu



próg = 1% wart. maks

próg = 10% wart. maks

próg = 25% wart. maks

próg = 25% wart. maks





## Rozszerzenie morfologii matematycznej na obrazy w skali szarości

- obraz w skali szarości  $I(x,y)$  może być traktowany jako bryła pod powierzchnia 3D:  $(I, y, I(x,y))$  (podobnie jak w przypadku transformaty odległościowej)
- morfologia w skali szarości jest uważana za rozszerzenie morfologii do przypadku trójwymiarowego:

$$A \oplus B = \{p+q : p \in A \wedge q \in B\}$$

gdzie zbiory A i B (obraz i SE) są trójwymiarowe



## Dylacja w skali szarości

$$I(x, y) \rightarrow \max_{i=-1, \dots, +1, j=-1, \dots, +1} I(x+i, y+j)$$

obraz oryginalny



obraz po 1 dylacji



- „rosną” obszary jasne
- znikają detale

## Dylacja w skali szarości



2 iteracje



3 iteracje



5 iteracji



10 iteracji



## Erozja w skali szarości

$$I(x, y) \rightarrow \min_{i=-1, \dots, +1, j=-1, \dots, +1} I(x+i, y+j)$$

obraz oryginalny



obraz po 1 dylacji



- „rosną” obszary ciemne
- znikają detale

# Erozja w skali szarości



2 iteracje



3 iteracje



5 iteracji



10 iteracji

# Zamknięcie w skali szarości



obraz oryginalny



zamknięcie w skali szarości



zamknięcie z większym SE



- erozja dylacji
- zanikają ciemne szczegóły
- jasne detale pozostają bez zmian

# Otwarcie w skali szarości



obraz oryginalny



otwarcie w skali szarości



otwarcie z większym SE



- dylacja erozji
- zanikają jasne szczegóły
- ciemne detale pozostają bez zmian

# GNU Free Documentation License



Version 1.2, November 2002

Copyright (C) 2000,2001,2002 Free Software Foundation, Inc.  
51 Franklin St, Fifth Floor, Boston, MA 02110-1301 USA  
Everyone is permitted to copy and distribute verbatim copies  
of this license document, but changing it is not allowed.

## 0. PREAMBLE

The purpose of this License is to make a manual, textbook, or other functional and useful document "free" in the sense of freedom: to assure everyone the effective freedom to copy and redistribute it, with or without modifying it, either commercially or noncommercially. Secondly, this License preserves for the author and publisher a way to get credit for their work, while not being considered responsible for modifications made by others.

This License is a kind of "copyleft", which means that derivative works of the document must themselves be free in the same sense. It complements the GNU General Public License, which is a copyleft license designed for free software.

We have designed this License in order to use it for manuals for free software, because free software needs free documentation: a free program should come with manuals providing the same freedoms that the software does. But this License is not limited to software manuals; it can be used for any textual work, regardless of subject matter or whether it is published as a printed book. We recommend this License principally for works whose purpose is instruction or reference.

## 1. APPLICABILITY AND DEFINITIONS

This License applies to any manual or other work, in any medium, that contains a notice placed by the copyright holder saying it can be distributed under the terms of this License. Such a notice grants a world-wide, royalty-free license, unlimited in duration, to use that work under the conditions stated herein. The "Document", below, refers to any such manual or work. Any member of the public is a licensee, and is addressed as "you". You accept the license if you copy, modify or distribute the work in a way requiring permission under copyright law.

A "Modified Version" of the Document means any work containing the Document or a portion of it, either copied verbatim, or with modifications and/or translated into another language.

A "Secondary Section" is a named appendix or a front-matter section of the Document that deals exclusively with the relationship of the publishers or authors of the Document to the Document's overall subject (or to related matters) and contains nothing that could fall directly within that overall subject. (Thus, if the Document is in part a textbook of mathematics, a Secondary Section may not explain any mathematics.) The relationship could be a matter of historical connection with the subject or with related matters, or of legal, commercial, philosophical, ethical or political position regarding them.

The "Invariant Sections" are certain Secondary Sections whose titles are designated, as being those of Invariant Sections, in the notice that says that the Document is released under this License. If a section does not fit the above definition of Secondary then it is not allowed to be designated as Invariant. The Document may contain zero Invariant Sections. If the Document does not identify any Invariant Sections then there are none.

The "Cover Texts" are certain short passages of text that are listed, as Front-Cover Texts or Back-Cover Texts, in the notice that says that the Document is released under this License. A Front-Cover Text may be at most 5 words, and a Back-Cover Text may be at most 25 words.

A "Transparent" copy of the Document means a machine-readable copy, represented in a format whose specification is available to the general public, that is suitable for revising the document straightforwardly with generic text editors or (for images composed of pixels) generic paint programs or (for drawings) some widely available drawing editor, and that is suitable for input to text formatters or for automatic translation to a variety of formats suitable for input to text formatters. A copy made in an otherwise Transparent file format whose markup, or absence of markup, has been arranged to thwart or discourage subsequent modification by readers is not Transparent. An image format is not Transparent if used for any substantial amount of text. A copy that is not "Transparent" is called "Opaque".

Examples of suitable formats for Transparent copies include plain ASCII without markup, Texinfo input format, LaTeX input format, SGML or XML using a publicly available DTD, and standard-conforming simple HTML, PostScript or PDF designed for human modification. Examples of transparent image formats include PNG, XCF and JPG. Opaque formats include proprietary formats that can be read and edited only by proprietary word processors, SGML or XML for which the DTD and/or processing tools are not generally available, and the machine-generated HTML, PostScript or PDF produced by some word processors for output purposes only.

The "Title Page" means, for a printed book, the title page itself, plus such following pages as are needed to hold, legibly, the material this License requires to appear in the title page. For works in formats which do not have any title page as such, "Title Page" means the text near the most prominent appearance of the work's title, preceding the beginning of the body of the text.

A section "Entitled XYZ" means a named subunit of the Document whose title either is precisely XYZ or contains XYZ in parentheses following text that translates XYZ in another language. (Here XYZ stands for a specific section name mentioned below, such as "Acknowledgements", "Dedications", "Endorsements", or "History".) To "Preserve the Title" of such a section when you modify the Document means that it remains a section "Entitled XYZ" according to this definition.

The Document may include Warranty Disclaimers next to the notice which states that this License applies to the Document. These Warranty Disclaimers are considered to be included by reference in this License, but only as regards disclaiming warranties: any other implication that these Warranty Disclaimers may have is void and has no effect on the meaning of this License.

## 2. VERBATIM COPYING

You may copy and distribute the Document in any medium, either commercially or noncommercially, provided that this License, the copyright notices, and the license notice saying this License applies to the Document are reproduced in all copies, and that you add no other conditions whatsoever to those of this License. You may not use technical measures to obstruct or control the reading or further copying of the copies you make or distribute. However, you may accept compensation in exchange for copies. If you distribute a large enough number of copies you must also follow the conditions in section 3.

You may also lend copies, under the same conditions stated above, and you may publicly display copies.



### 3. COPYING IN QUANTITY

If you publish printed copies (or copies in media that commonly have printed covers) of the Document, numbering more than 100, and the Document's license notice requires Cover Texts, you must enclose the copies in covers that carry, clearly and legibly, all these Cover Texts: Front-Cover Texts on the front cover, and Back-Cover Texts on the back cover. Both covers must also clearly and legibly identify you as the publisher of these copies. The front cover must present the full title with all words of the title equally prominent and visible. You may add other material on the covers in addition. Copying with changes limited to the covers, as long as they preserve the title of the Document and satisfy these conditions, can be treated as verbatim copying in other respects.

If the required texts for either cover are too voluminous to fit legibly, you should put the first ones listed (as many as fit reasonably) on the actual cover, and continue the rest onto adjacent pages.

If you publish or distribute Opaque copies of the Document numbering more than 100, you must either include a machine-readable Transparent copy along with each Opaque copy, or state in or with each Opaque copy a computer-network location from which the general network-using public has access to download using public-standard network protocols a complete Transparent copy of the Document, free of added material. If you use the latter option, you must take reasonably prudent steps, when you begin distribution of Opaque copies in quantity, to ensure that this Transparent copy will remain thus accessible at the stated location until at least one year after the last time you distribute an Opaque copy (directly or through your agents or retailers) of that edition to the public.

It is requested, but not required, that you contact the authors of the Document well before redistributing any large number of copies, to give them a chance to provide you with an updated version of the Document.

### 4. MODIFICATIONS

You may copy and distribute a Modified Version of the Document under the conditions of sections 2 and 3 above, provided that you release the Modified Version under precisely this License, with the Modified Version filling the role of the Document, thus licensing distribution and modification of the Modified Version to whoever possesses a copy of it. In addition, you must do these things in the Modified Version:

- \* A. Use in the Title Page (and on the covers, if any) a title distinct from that of the Document, and from those of previous versions (which should, if there were any, be listed in the History section of the Document). You may use the same title as a previous version if the original publisher of that version gives permission.
- \* B. List on the Title Page, as authors, one or more persons or entities responsible for authorship of the modifications in the Modified Version, together with at least five of the principal authors of the Document (all of its principal authors, if it has fewer than five), unless they release you from this requirement.
- \* C. State on the Title page the name of the publisher of the Modified Version, as the publisher.
- \* D. Preserve all the copyright notices of the Document.
- \* E. Add an appropriate copyright notice for your modifications adjacent to the other copyright notices.
- \* F. Include, immediately after the copyright notices, a license notice giving the public permission to use the Modified Version under the terms of this License, in the form shown in the Addendum below.
- \* G. Preserve in that license notice the full lists of Invariant Sections and required Cover Texts given in the Document's license notice.
- \* H. Include an unaltered copy of this License.
- \* I. Preserve the section Entitled "History", Preserve its Title, and add to it an item stating at least the title, year, new authors, and publisher of the Modified Version as given on the Title Page. If there is no section Entitled "History" in the Document, create one stating the title, year, authors, and publisher of the Document as given on its Title Page, then add an item describing the Modified Version as stated in the previous sentence.
- \* J. Preserve the network location, if any, given in the Document for public access to a Transparent copy of the Document, and likewise the network locations given in the Document for previous versions it was based on. These may be placed in the "History" section. You may omit a network location for a work that was published at least four years before the Document itself, or if the original publisher of the version it refers to gives permission.
- \* K. For any section Entitled "Acknowledgements" or "Dedications", Preserve the Title of the section, and preserve in the section all the substance and tone of each of the contributor acknowledgements and/or dedications given therein.
- \* L. Preserve all the Invariant Sections of the Document, unaltered in their text and in their titles. Section numbers or the equivalent are not considered part of the section titles.
- \* M. Delete any section Entitled "Endorsements". Such a section may not be included in the Modified Version.
- \* N. Do not retitle any existing section to be Entitled "Endorsements" or to conflict in title with any Invariant Section.
- \* O. Preserve any Warranty Disclaimers.

If the Modified Version includes new front-matter sections or appendices that qualify as Secondary Sections and contain no material copied from the Document, you may at your option designate some or all of these sections as invariant. To do this, add their titles to the list of Invariant Sections in the Modified Version's license notice. These titles must be distinct from any other section titles.

You may add a section Entitled "Endorsements", provided it contains nothing but endorsements of your Modified Version by various parties--for example, statements of peer review or that the text has been approved by an organization as the authoritative definition of a standard.

You may add a passage of up to five words as a Front-Cover Text, and a passage of up to 25 words as a Back-Cover Text, to the end of the list of Cover Texts in the Modified Version. Only one passage of Front-Cover Text and one of Back-Cover Text may be added by (or through arrangements made by) any one entity. If the Document already includes a cover text for the same cover, previously added by you or by arrangement made by the same entity you are acting on behalf of, you may not add another; but you may replace the old one, on explicit permission from the previous publisher that added the old one.

The author(s) and publisher(s) of the Document do not by this License give permission to use their names for publicity for or to assert or imply endorsement of any Modified Version.



## 6. COLLECTIONS OF DOCUMENTS

You may make a collection consisting of the Document and other documents released under this License, and replace the individual copies of this License in the various documents with a single copy that is included in the collection, provided that you follow the rules of this License for verbatim copying of each of the documents in all other respects.

You may extract a single document from such a collection, and distribute it individually under this License, provided you insert a copy of this License into the extracted document, and follow this License in all other respects regarding verbatim copying of that document.

## 7. AGGREGATION WITH INDEPENDENT WORKS

A compilation of the Document or its derivatives with other separate and independent documents or works, in or on a volume of a storage or distribution medium, is called an "aggregate" if the copyright resulting from the compilation is not used to limit the legal rights of the compilation's users beyond what the individual works permit. When the Document is included in an aggregate, this License does not apply to the other works in the aggregate which are not themselves derivative works of the Document.

If the Cover Text requirement of section 3 is applicable to these copies of the Document, then if the Document is less than one half of the entire aggregate, the Document's Cover Texts may be placed on covers that bracket the Document within the aggregate, or the electronic equivalent of covers if the Document is in electronic form. Otherwise they must appear on printed covers that bracket the whole aggregate.

## 8. TRANSLATION

Translation is considered a kind of modification, so you may distribute translations of the Document under the terms of section 4. Replacing Invariant Sections with translations requires special permission from their copyright holders, but you may include translations of some or all Invariant Sections in addition to the original versions of these Invariant Sections. You may include a translation of this License, and all the license notices in the Document, and any Warranty Disclaimers, provided that you also include the original English version of this License and the original versions of those notices and disclaimers. In case of a disagreement between the translation and the original version of this License or a notice or disclaimer, the original version will prevail.

If a section in the Document is Entitled "Acknowledgements", "Dedications", or "History", the requirement (section 4) to Preserve its Title (section 1) will typically require changing the actual title.

## 9. TERMINATION

You may not copy, modify, sublicense, or distribute the Document except as expressly provided for under this License. Any other attempt to copy, modify, sublicense or distribute the Document is void, and will automatically terminate your rights under this License. However, parties who have received copies, or rights, from you under this License will not have their licenses terminated so long as such parties remain in full compliance.

## 10. FUTURE REVISIONS OF THIS LICENSE

The Free Software Foundation may publish new, revised versions of the GNU Free Documentation License from time to time. Such new versions will be similar in spirit to the present version, but may differ in detail to address new problems or concerns. See <http://www.gnu.org/copyleft/>.

Each version of the License is given a distinguishing version number. If the Document specifies that a particular numbered version of this License "or any later version" applies to it, you have the option of following the terms and conditions either of that specified version or of any later version that has been published (not as a draft) by the Free Software Foundation. If the Document does not specify a version number of this License, you may choose any version ever published (not as a draft) by the Free Software Foundation.

## 5. COMBINING DOCUMENTS

You may combine the Document with other documents released under this License, under the terms defined in section 4 above for modified versions, provided that you include in the combination all of the Invariant Sections of all of the original documents, unmodified, and list them all as Invariant Sections of your combined work in its license notice, and that you preserve all their Warranty Disclaimers.

The combined work need only contain one copy of this License, and multiple identical Invariant Sections may be replaced with a single copy. If there are multiple Invariant Sections with the same name but different contents, make the title of each such section unique by adding at the end of it, in parentheses, the name of the original author or publisher of that section if known, or else a unique number. Make the same adjustment to the section titles in the list of Invariant Sections in the license notice of the combined work.

In the combination, you must combine any sections Entitled "History" in the various original documents, forming one section Entitled "History"; likewise combine any sections Entitled "Acknowledgements", and any sections Entitled "Dedications". You must delete all sections Entitled "Endorsements."

(not as a draft) by the Free Software Foundation.

# Cyfrowe Przetwarzanie Obrazów

## Digital Image Processing



Aurelio A. Heckert

Copyright (c) 2008 Cezary Bołdak

Permission is granted to copy, distribute and/or modify this document under the terms of the GNU Free Documentation License, Version 1.2 or any later version published by the Free Software Foundation; with no Invariant Sections, no Front-Cover Texts, and no Back-Cover Texts. A copy of the license is included in the section entitled "GNU Free Documentation License"

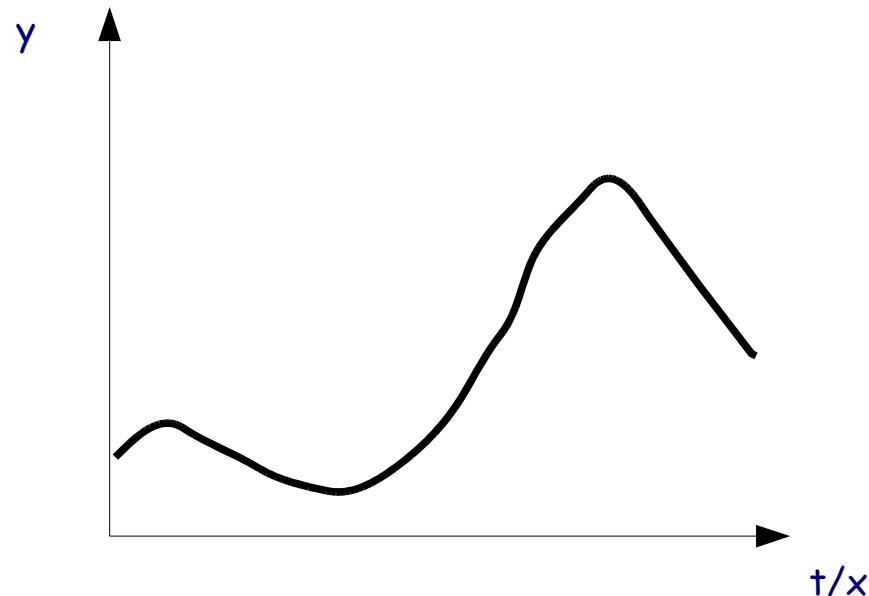


## Wykład 6

# Przetwarzanie obrazów w dziedzinie częstotliwości (frequency-space image processing)



## Sygnal analogowy



- sygnał jest zapisem zmienności pewnej wielkości  $y$  w zależności od upływającego czasu

$$y(t)=f(t)$$

- funkcja  $f$  może być dana analitycznie (równanie) lub nie
- zmienna niezależna  $t$  (czas) może też oznaczać dystans, położenie  $x$
- na sygnał taki można narzucić szereg warunków (ciągłość, różniczkowalność, ...)

## Rozkład sygnału na funkcje bazowe

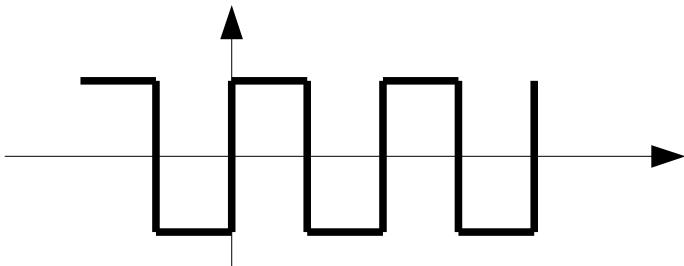


- każdy sygnał można zastąpić kombinacją liniową innych funkcji nazywanych **funkcjami bazowymi**:

$$y(t) = a_1 f_1(t) + a_2 f_2(t) + \dots + a_n f_n(t) \quad (+\dots)$$

- czasami jest to suma skończona, czasami nie
- założmy sygnał dany wzorem:

$$y(t) = \begin{cases} 1 & \text{jeżeli } k \cdot 2\pi \leq t < k \cdot 2\pi + \pi \\ 0 & \text{jeżeli } k \cdot 2\pi + \pi \leq t \leq (k+1) \cdot (2\pi) + \pi \end{cases}$$



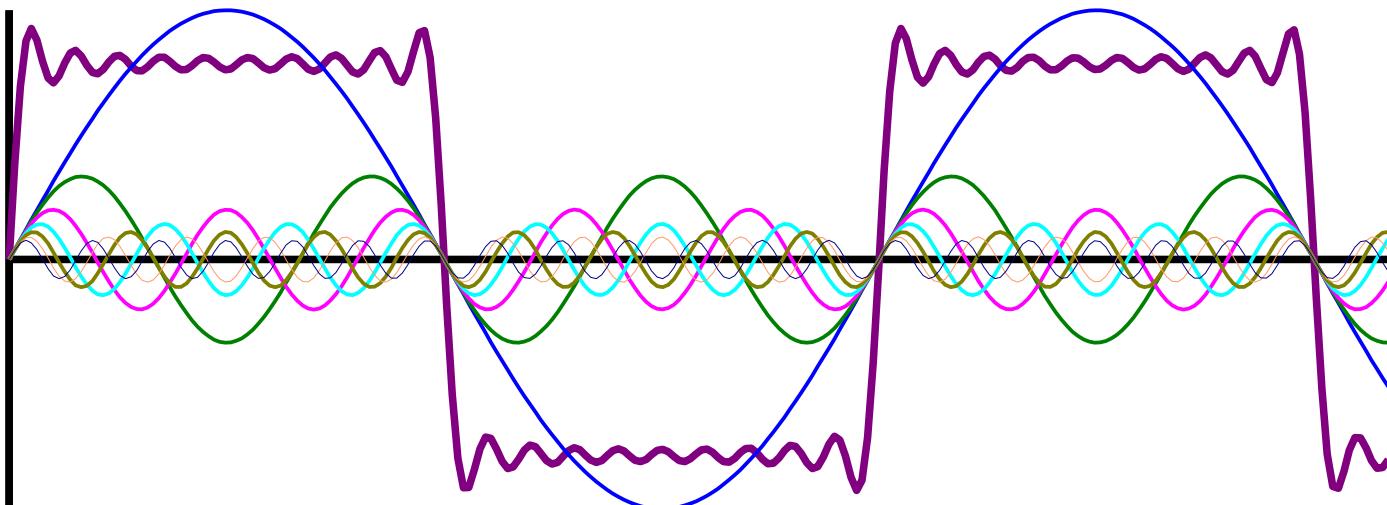
- sygnał taki można przedstawić za pomocą nieskończonej sumy sinusów o wzrastającej częstotliwości:

$$\sin(x) + \frac{1}{3}\sin(3x) + \frac{1}{5}\sin(5x) + \frac{1}{7}\sin(7x) + \dots = \sum_{i=0, \dots, \infty} \frac{1}{2i+1} \sin((2i+1)x)$$

- jeśli weźmiemy skończoną liczbę składników, będzie to tylko przybliżenie sygnału (tym dokładniejsze im więcej składników)



## Rozkład sygnału na funkcje bazowe



- na wykresie przedstawiona jest suma 10 składników ( $i=0, \dots, 9$ ) (nie wszystkie składowe sinusy są narysowane)
- animacja pokazująca wpływ liczby składników na jakość przybliżenia

składowe harmoniczne: 1

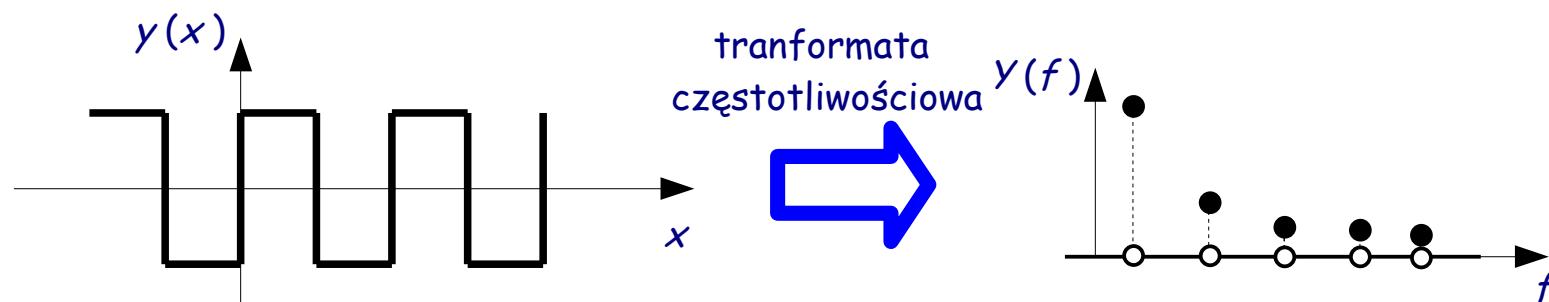


Fullofstars, Wikipedia.org



## Transformata częstotliwościowa

- jeśli przyjmiemy że funkcjami bazowymi są w tym przypadku funkcje sinus o wzrastającej częstotliwości (**funkcje harmoniczne**), to można uznać że dokonaliśmy przejścia z **dziedziny czasu na dziedzinę częstotliwości**
- wartościami funkcji w dziedzinie częstotliwości są same zera z wyjątkiem częstotliwości  $f=1$  (wartość 1),  $f=3$  (wartość 1/3),  $f=5$  (wartość 1/5), ...



- użycie innych funkcji bazowych da inne wartości
- transformacja ta (transformata) jest odwracalna – znowu możemy przejść z dziedziny częstotliwości do dziedziny czasowej (przestrzennej)

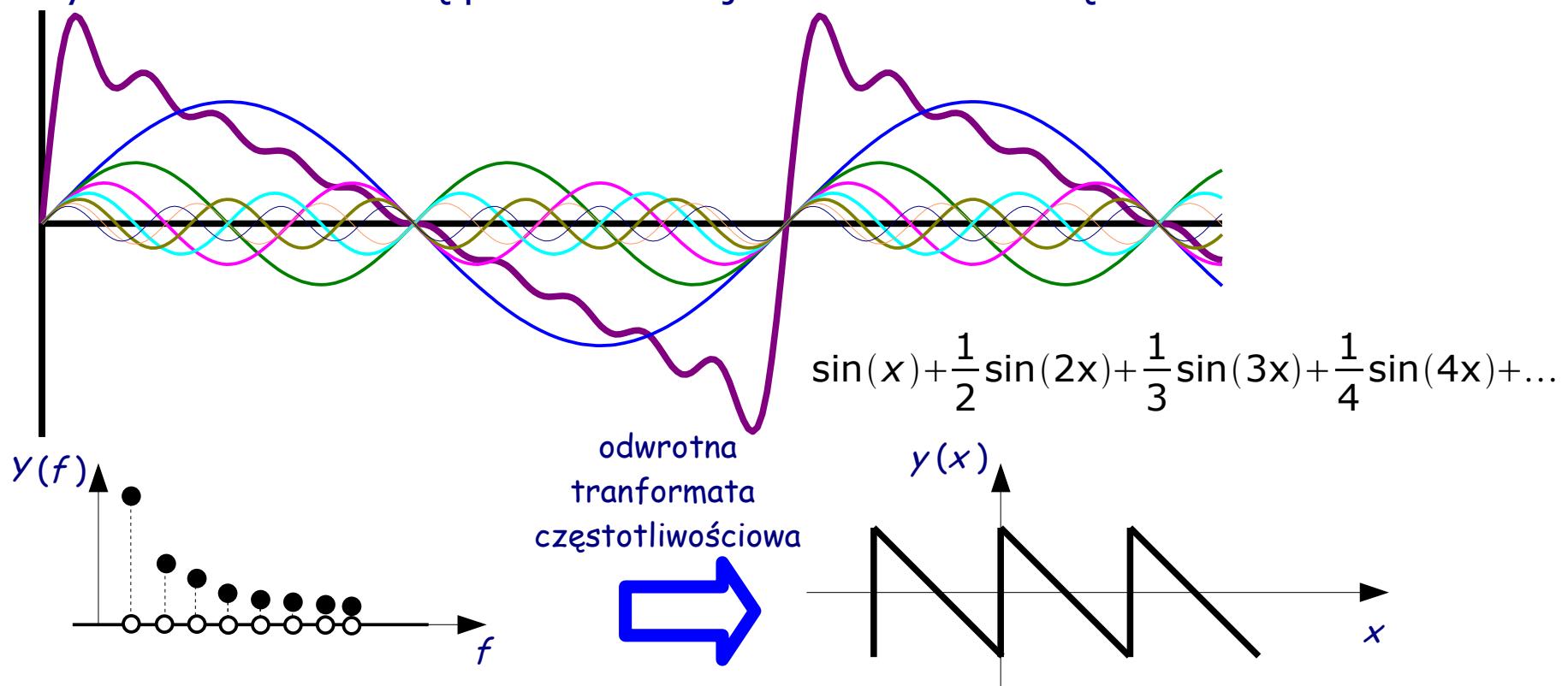


## Odwrotna transformata częstotliwościowa

- transformacja ta jest odwracalna



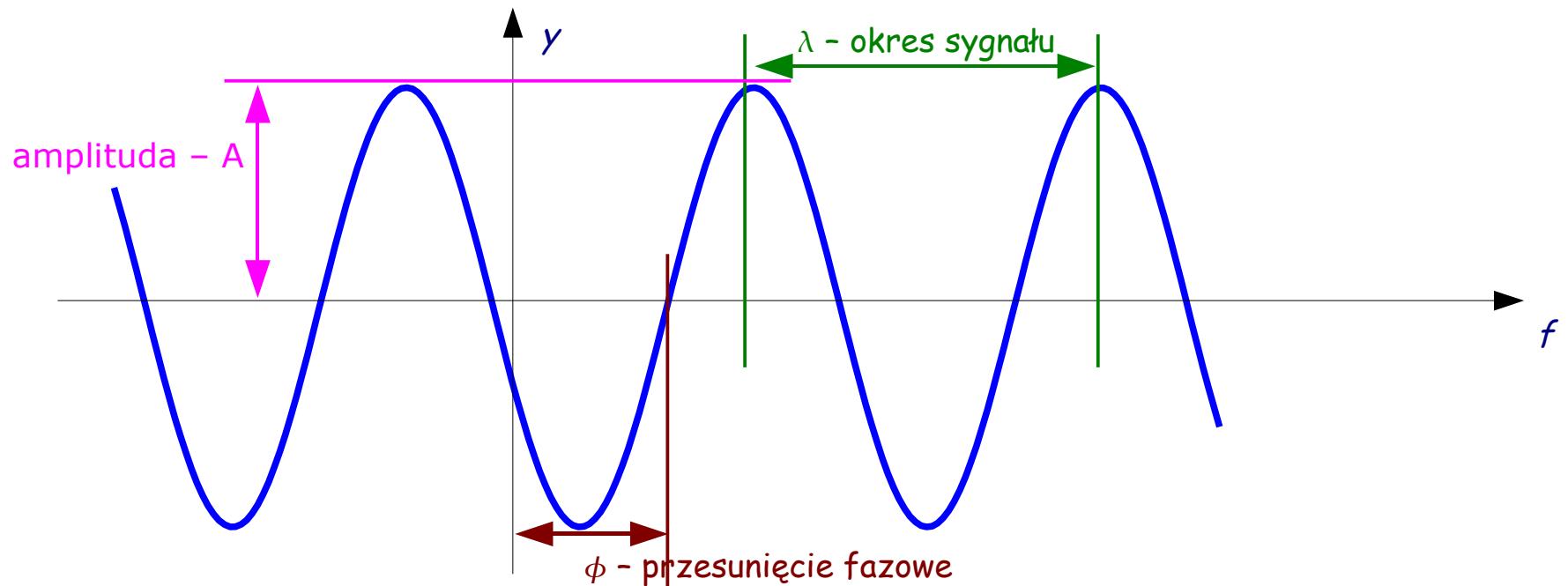
- zmiana wartości w dziedzinie częstotliwości przy tych samych funkcjach bazowych może dać inną postać funkcji w dziedzinie częstotliwości:





## Charakterystyka sinusoidy

$$y(x) = A \sin\left(\frac{2\pi}{\lambda} x - \phi\right)$$



$$\omega = \frac{2\pi}{\lambda} - \text{częstotliwość kątowa [radiany]}$$

$$f = \frac{1}{\lambda} - \text{częstotliwość [Hz]}$$



## Transformata Fouriera

- transformata Fouriera jest przykładem takiej transformaty częstotliwościowej
- funkcjami bazowymi są tu funkcje sinusów i cosinusów o zwiększającej się częstotliwości (funkcje harmoniczne)

$$y(x) = A_0 + \sum_{k=1}^{\infty} A_k \cos\left(\frac{2\pi k}{\lambda}x\right) + B_k \sin\left(\frac{2\pi k}{\lambda}x\right)$$

gdzie:  $\lambda$  – okres sygnału  $y$

- współczynniki  $A$  i  $B$  są **zespolone !!!**
- zespolony współczynnik koduje amplitudę i fazę funkcji bazowej

$$y(x) = A_0 + \sum_{k=1}^{\infty} |A_k| \cos\left(\frac{2\pi k}{\lambda}x + \phi_{A_k}\right) + |B_k| \sin\left(\frac{2\pi k}{\lambda}x + \phi_{B_k}\right)$$

$$|A_k| = \sqrt{\operatorname{re}(A_k)^2 + \operatorname{im}(A_k)}, \quad \phi_{A_k} = \operatorname{tg}^{-1} \frac{\operatorname{im}(A_k)}{\operatorname{re}(A_k)}, \quad B_k \text{ i } \phi_{B_k} \text{ analogicznie.}$$

- równoważna forma:

$$y(x) = \sum_{k=-\infty}^{\infty} C_k e^{i \frac{2\pi k}{\lambda} x} = \sum_{k=-\infty}^{\infty} |C_k| \{ \cos\left(\frac{2\pi}{\lambda}x + \phi_k\right) + i \sin\left(\frac{2\pi}{\lambda}x + \phi_k\right) \}$$

gdzie współczynniki  $C$  są zespolone

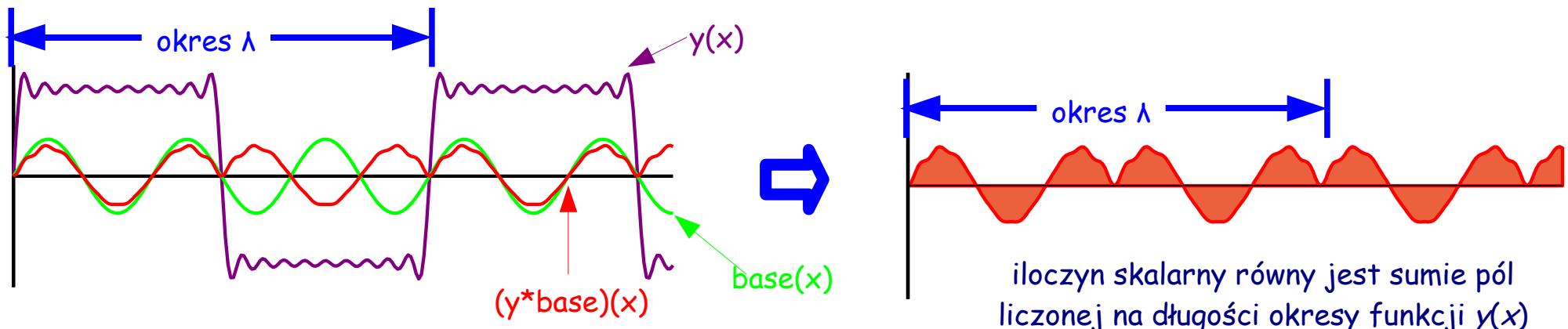
$$e^{\pm i \cdot x} = \cos(x) \pm i \sin(x)$$



## Wyliczanie współczynników

$$C_k = |C_k| e^{+i\phi_k} = \frac{1}{\lambda} \int_{-\lambda}^{\lambda} y(x) e^{-i \frac{2\pi k}{\lambda} x} dx$$

- współczynniki  $C$  mierzą podobieństwo funkcji  $y(x)$  do każdej z funkcji bazowych
- miara podobieństwa wyliczana jest przez całkę z iloczynu skalarnego 2 funkcji



- funkcja  $y(x)$  jest uważana za periodyczną

$$y(x) = y(x + k \cdot \lambda)$$

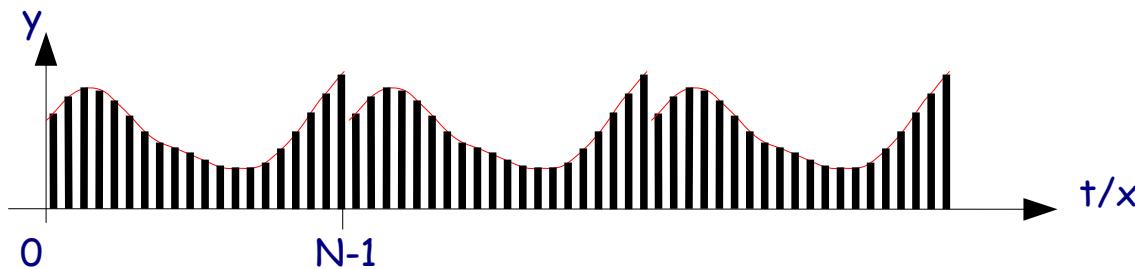
- transformata Fouriera jest odwracalna:

$$Y(f) = \int_{-\infty}^{\infty} y(x) e^{-i 2\pi f x} dx \quad y(x) = \int_{-\infty}^{\infty} Y(f) e^{i 2\pi f x} df$$



## Dyskretna transformata Fouriera

- w przypadku dyskretnym (sygnał cyfrowy, obraz cyfrowy) całki zamieniają się na sumy
- funkcje bazowe również mają postać dyskretną
- sygnał cyfrowy traktowany jest jako jeden okres i uzupełniany z przodu i z tyłu swoimi powtórzeniami – może to wprowadzać nieciągłości



- liczba próbek określana jest na  $N$  – jest to też okres funkcji

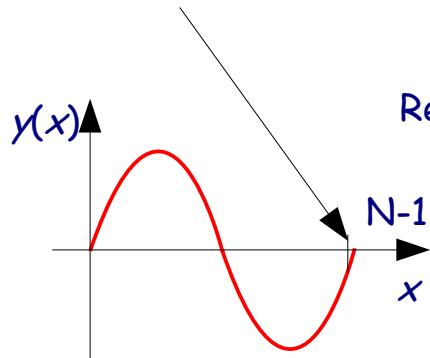
$$y(x) = \sum_{k=0}^{N-1} C_k e^{i \frac{2\pi k}{N} x} \quad Y(f) = \sum_{k=0}^{N-1} C_k e^{-i \frac{2\pi k}{N} f}$$

$$C_k = \frac{1}{N} \sum_{x=0}^{N-1} y(x) e^{-i \frac{2\pi k}{N} x}$$

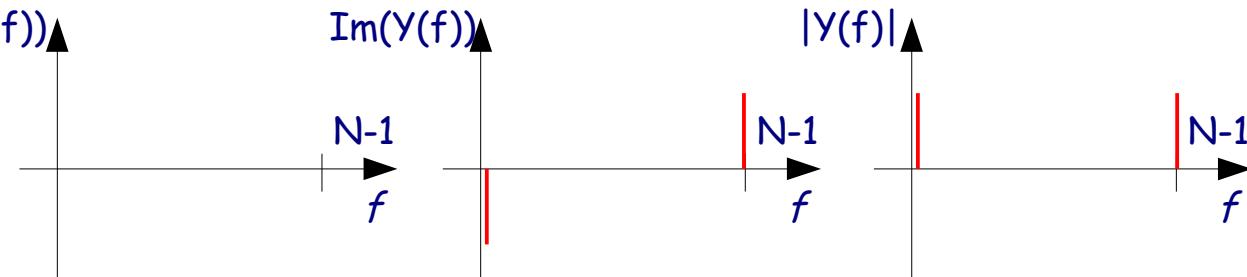


## Transformata Fouriera sinusa

$$\begin{aligned}y[N-1] &\neq 0 \\y[N] = y[0] &= 0\end{aligned}$$



$$y[x] = \sin\left(\frac{2\pi}{N}x\right)$$



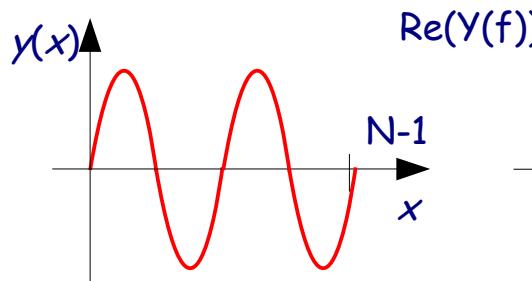
$$\begin{aligned}Y[0] &= 0 \\Y[1] &= 0 - C \cdot i \\Y[N-1] &= 0 + C \cdot i\end{aligned}$$

$$\sqrt{\Re(Y(t))^2 + \Im(Y(t))^2}$$

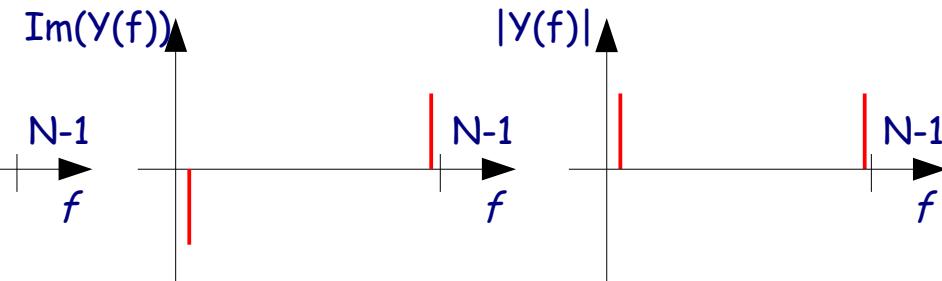
- sinus ma częstotliwość  $f=1$  (jeden przebieg na N próbek)
- powoduje powstanie 2 „pików” w DFT: komponentu drugiego ( $Y[1]$ ) i ostatniego ( $Y[N-1]$ ), są one symetryczne w części urojonej względem punktu  $(N/2, 0)$



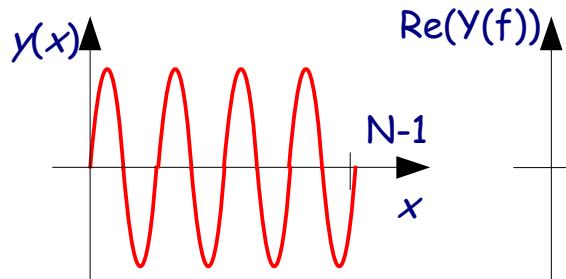
## Transformata Fouriera sinusa



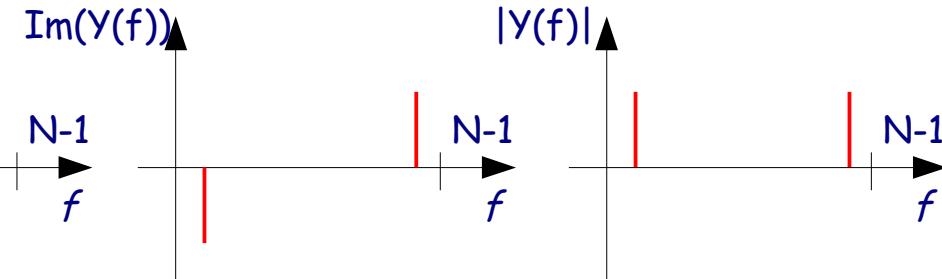
$$y[x] = \sin\left(\frac{2\pi}{N} 2 \cdot x\right)$$



$$\begin{aligned} Y[2] &= 0 - C \cdot i \\ Y[N-2] &= 0 + C \cdot i \end{aligned}$$



$$y[x] = \sin\left(\frac{2\pi}{N} 2 \cdot x\right)$$

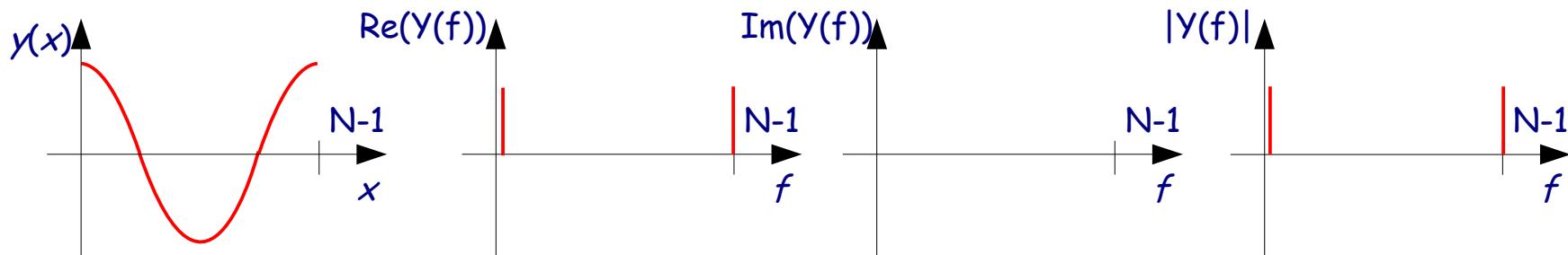


$$\begin{aligned} Y[4] &= 0 - C \cdot i \\ Y[N-4] &= 0 + C \cdot i \end{aligned}$$

- wzrastająca częstotliwość powoduje przesunięcie pików części urojonej w stronę punktu  $(N/2, 0)$
- moduł transformaty również się przesuwa

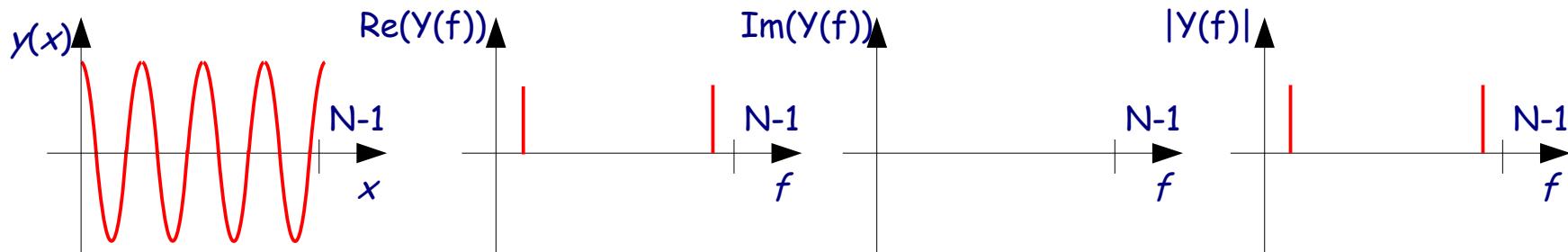


## Transformata Fouriera cosinusa



$$y[x] = \cos\left(\frac{2\pi}{N}x\right)$$

$$\begin{aligned} Y[1] &= C + 0 \cdot i \\ Y[N-1] &= C + 0 \cdot i \end{aligned}$$



$$y[x] = \cos\left(\frac{2\pi}{N}4 \cdot x\right)$$

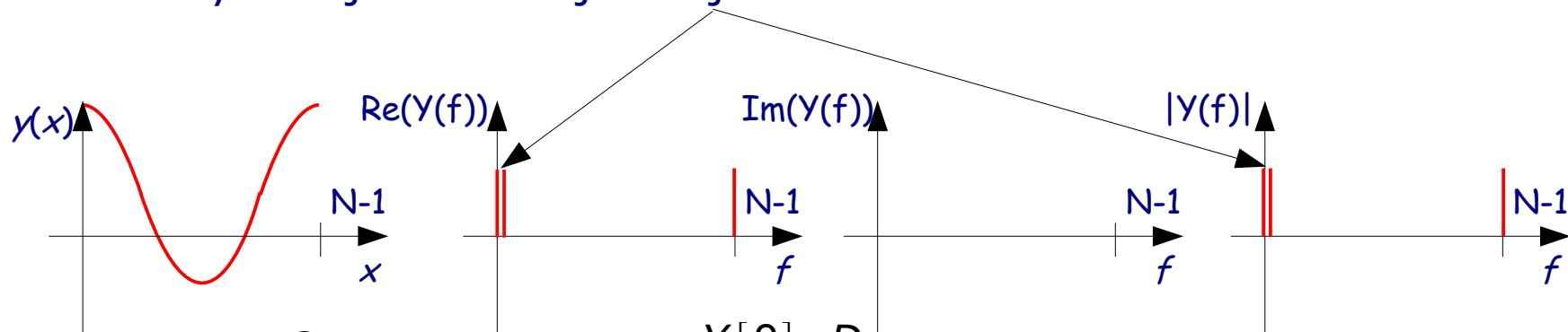
$$\begin{aligned} Y[4] &= C + 0 \cdot i \\ Y[N-4] &= C + 0 \cdot i \end{aligned}$$

- tu również położenie „pików” reprezentuje częstotliwość
- pikи w części rzeczywistej są symetryczne względem linii prostopadłej do osi OX i przechodzącej przez punkt  $(N/2, 0)$
- amplituda transformaty pozostaje cały czas taka sama

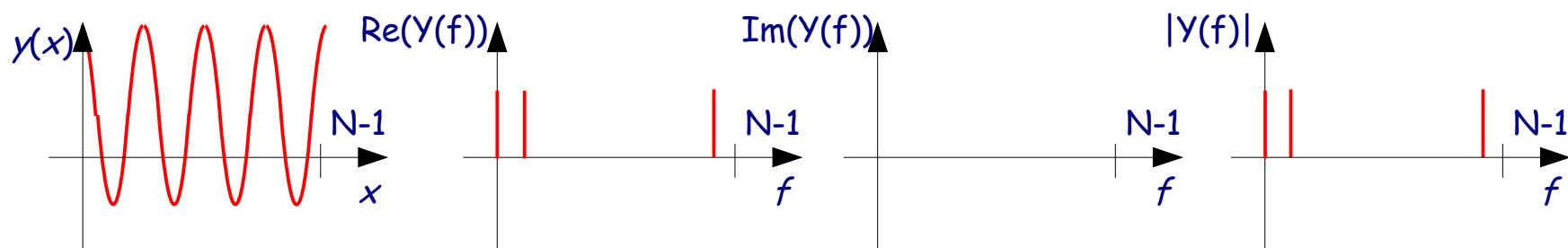


## Składowa stała sygnału

- $Y[0] = 0 \Rightarrow$  całka pod wykresem jest równa 0
- jest to częstotliwość 0
- przesunięcie funkcji do góry powoduje pojawienie się rzeczywistej składowej stałej



$$\begin{aligned} Y[0] &= D \\ Y[1] &= C + 0 \cdot i \\ Y[N-1] &= C + 0 \cdot i \end{aligned}$$

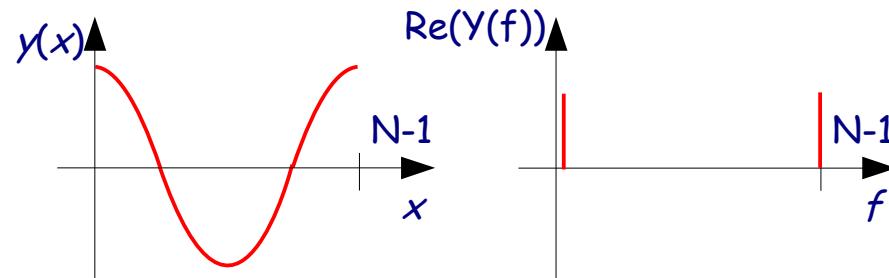


$$\begin{aligned} Y[0] &= D \\ Y[4] &= C + 0 \cdot i \\ Y[N-4] &= C + 0 \cdot i \end{aligned}$$

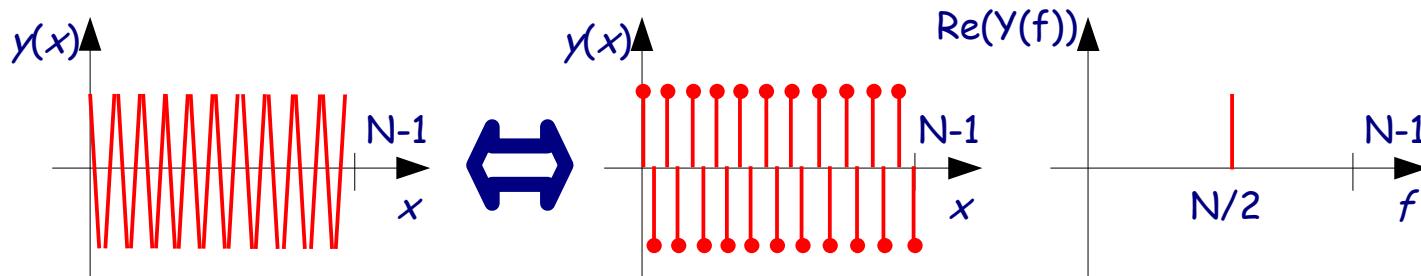


## Zakres częstotliwości

- najmniejszą niezerową częstotliwością brana pod uwagę jest  $f_{\min}=1$  – jedno powtórzenie sygnału w oknie/obrazie



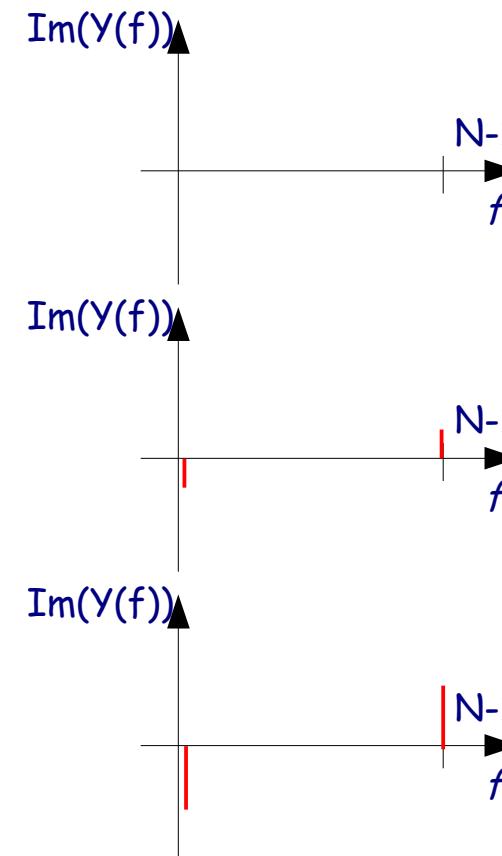
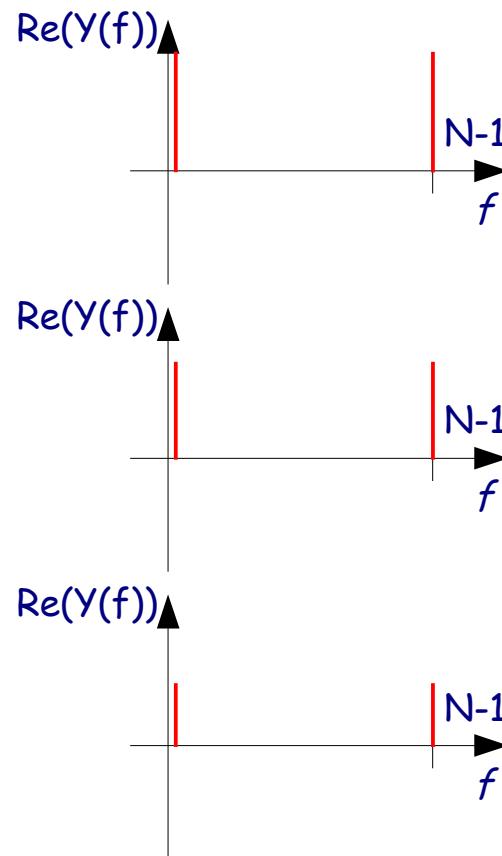
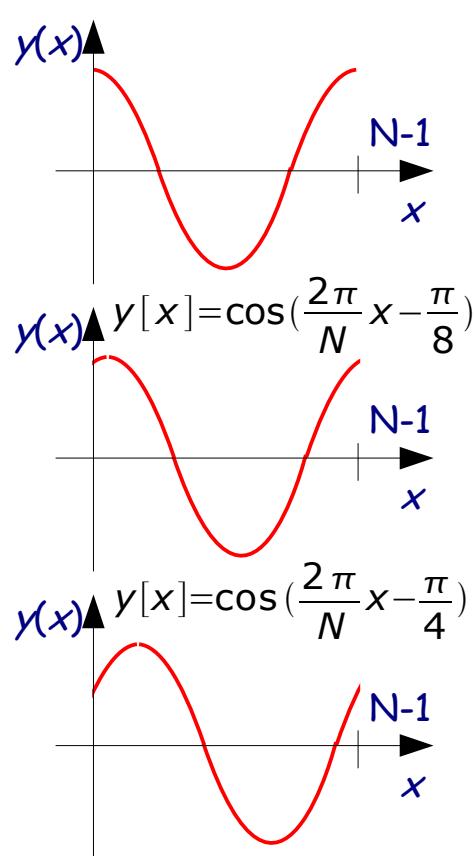
- największą częstotliwością, zgodnie z teorią próbkowania Nyquista-Shannona, jest  $f=N/2$



- częstotliwość  $f$  daje w transformacji 2 „piki”:  $Y[f]$  i  $Y[N-F]$

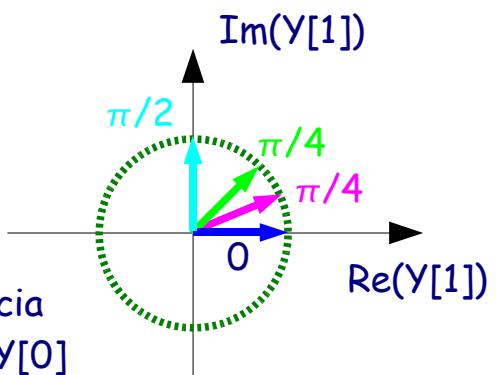


## Przesunięcie w fazie sygnału



- przesunięcie w fazie o  $\pi/2$  da w wyniku sinus
- w miarę przesuwania maleje część rzeczywista i rośnie urojona
- moduł pozostaje cały czas ten sam !

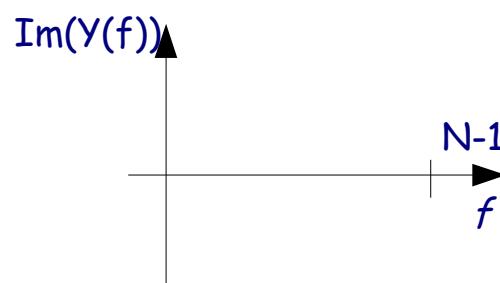
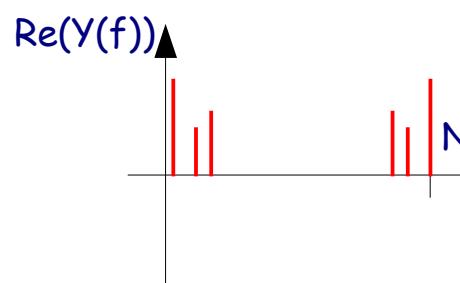
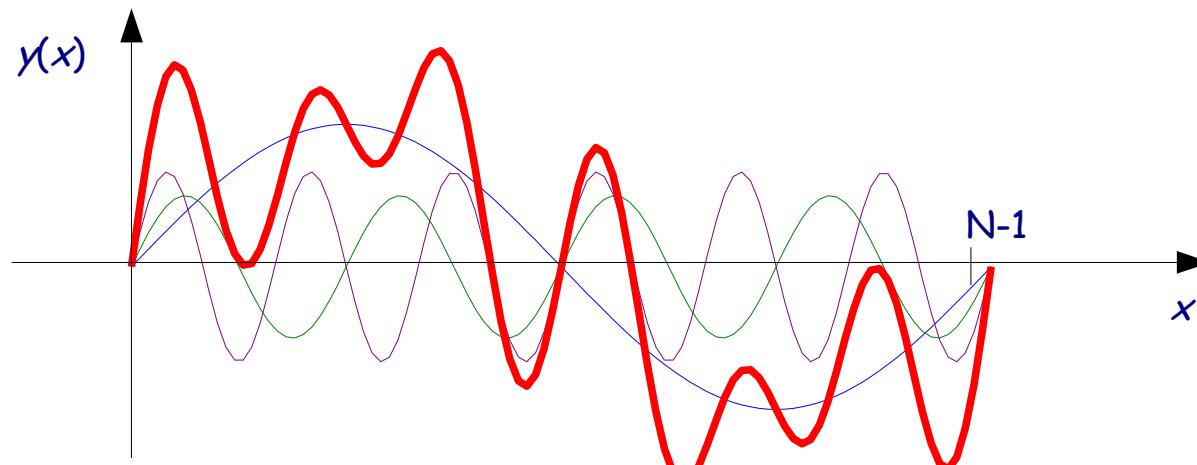
wpływ przesunięcia  
fazy na wartość  $Y[0]$





## Sygnal złożony

$$y[x] = \cos\left(\frac{2\pi}{N}x\right) + \frac{1}{2}\cos\left(\frac{2\pi}{N}4 \cdot x\right) + \frac{2}{3}\cos\left(\frac{2\pi}{N}6 \cdot x\right); \quad x=0, \dots, N-1$$



$$Y[1]=Y[N-1]=50$$

$$Y[4]=Y[N-4]=25$$

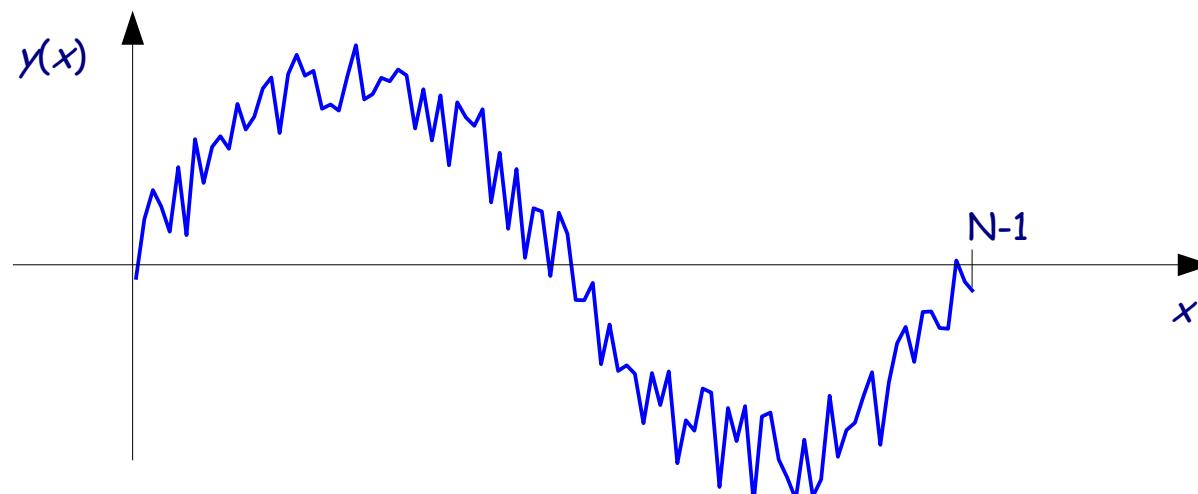
$$Y[6]=Y[N-6]=33.33$$

- sygnał złożony łatwo się separuje w dziedzinie częstotliwości

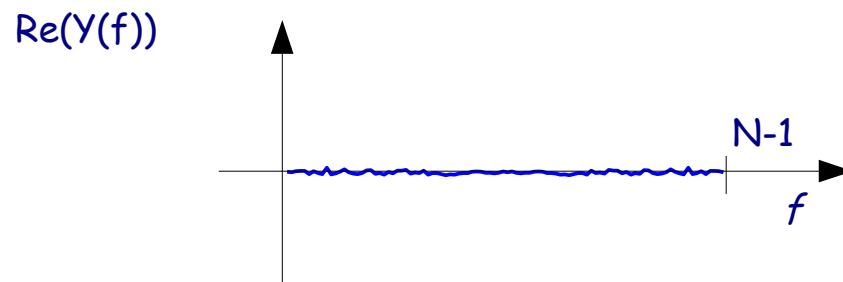
## Separowanie sygnału od szumu



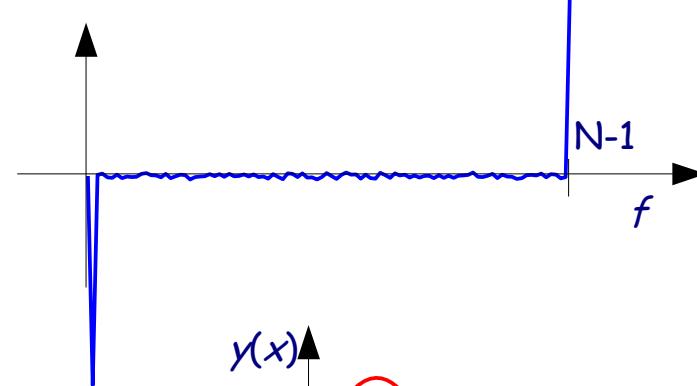
- sygnał sinusa z szumem



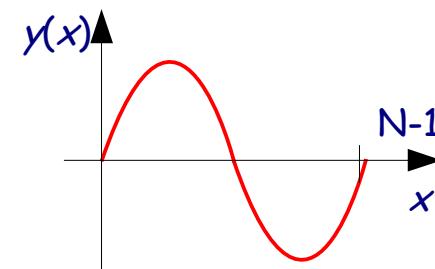
- po transformacie Fouriera:



$\text{Im}(Y(f))$



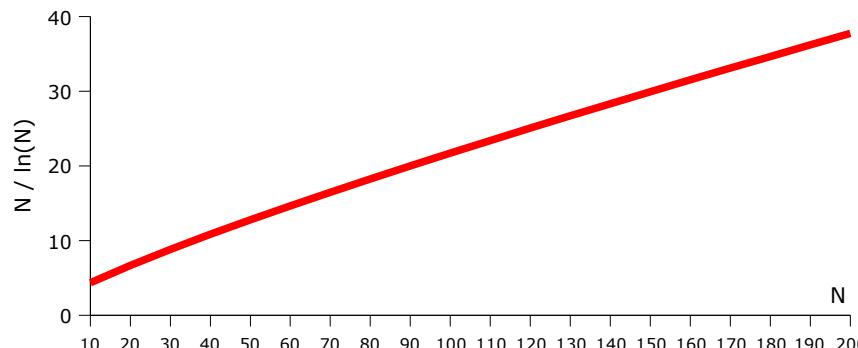
- po progowaniu i odwrotnej transformacie odzyskujemy sygnał bez szumu:
- sygnał szumu to różne częstotliwości o mniejszej częstotliwości





## Złożoność obliczeniowa DFT Szybka Transformata Fouriera (FFT)

- mamy do policzenia  $\sim N$  współczynników
- każdy współczynnik to suma  $\sim N$  wyrażeń
- złożoność obliczeniowa **DFT:  $O(N^2)$**
- złożoność odwrotnej transformaty jest taka sama
- szybka transformata Fouriera (*Fast Fourier Transform*) – **FFT** – pozwala policzyć współczynniki ze złożonością  **$O(N \ln N)$  !!!**
- dla małych  $N$  nie ma to wielkiego znaczenia
- dla dużych  $N$  różnice są bardzo duże !!!



- FFT jest dość skomplikowana
- zwykle są funkcje biblioteczne które liczą FFT
- oryginalna FFT liczy dla  $N=2^k$  (potęgi 2)
- procedury biblioteczne radzą sobie z tym



## Dyskretna transformata Fouriera 2D

- zakładamy rozmiar obrazu NxN

$$F(a, b) = \frac{1}{N^2} \sum_{x=0}^{N-1} \sum_{y=0}^{N-1} I(x, y) e^{-i2\pi \left( \frac{a \cdot x}{N} + \frac{b \cdot y}{N} \right)}$$

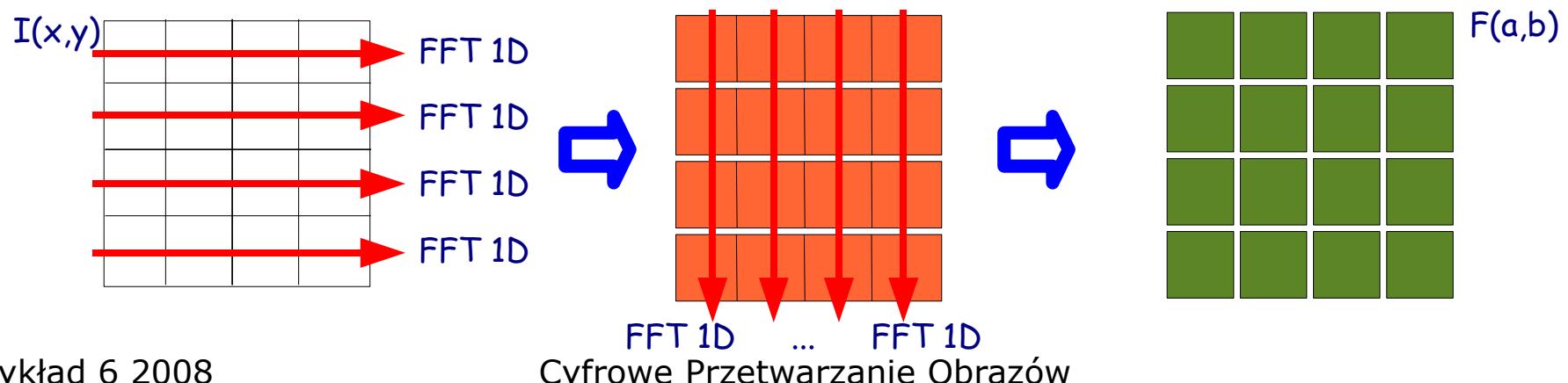
- IDFT 2D (odwrotna transformata Fouriera 2D):

$$I(x, y) = \frac{1}{N^2} \sum_{a=0}^{N-1} \sum_{b=0}^{N-1} F(a, b) e^{+i2\pi \left( \frac{a \cdot x}{N} + \frac{b \cdot y}{N} \right)}$$

- transformata Fouriera jest separowalna względem wymiarów:

$$F(a, b) = \frac{1}{N} \sum_{y=0}^{N-1} X(a, y) e^{-i2\pi \frac{b \cdot y}{N}} \quad \text{gdzie} \quad X(a, y) = \frac{1}{N} \sum_{x=0}^{N-1} I(x, y) e^{-i2\pi \frac{a \cdot x}{N}}$$

- tzn. najpierw można policzyć N transformat 1D w N wierszach a następnie N transformat 1D w kolumnach (z wynikami pierwszych N transformat) co znacznie przyspiesza obliczenia –  $O(N^2)$  zamiast  $O(N^4)$
- dalsze przyspieszenie obliczeń: FFT –  $O(N \ln N)$



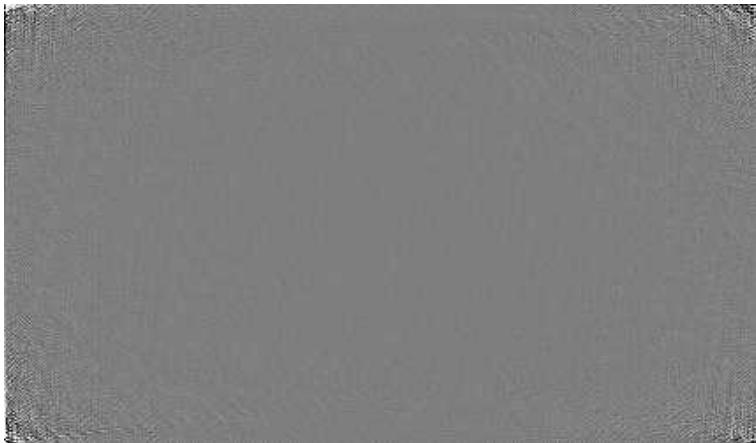


## Transformata obrazu

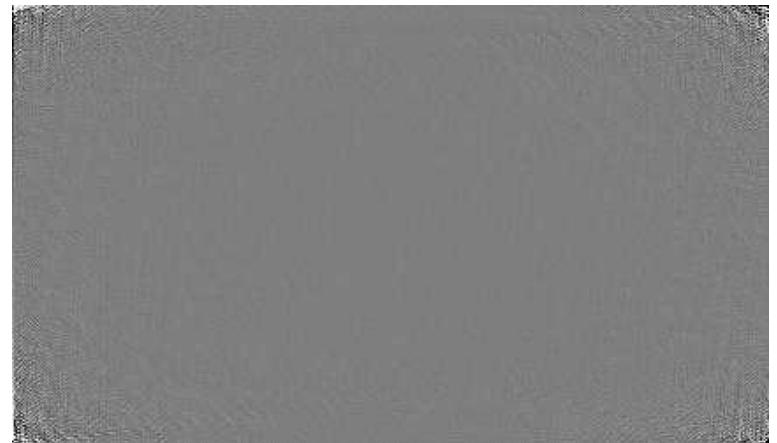


Image

Re(FFT(Image))



Im(FFT(Image))



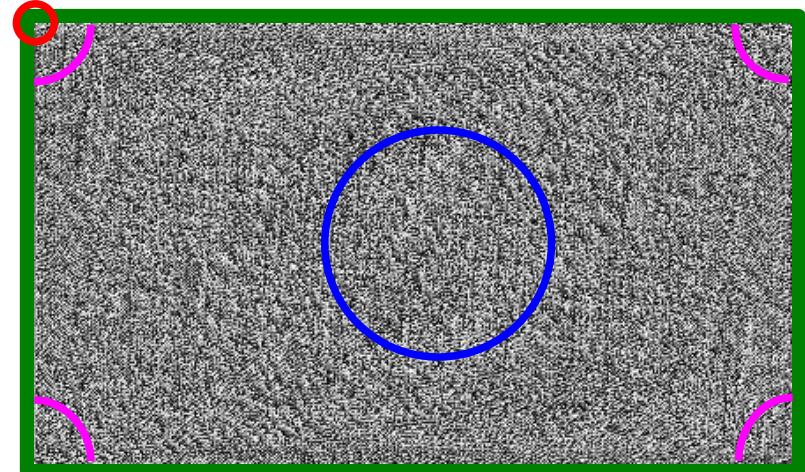
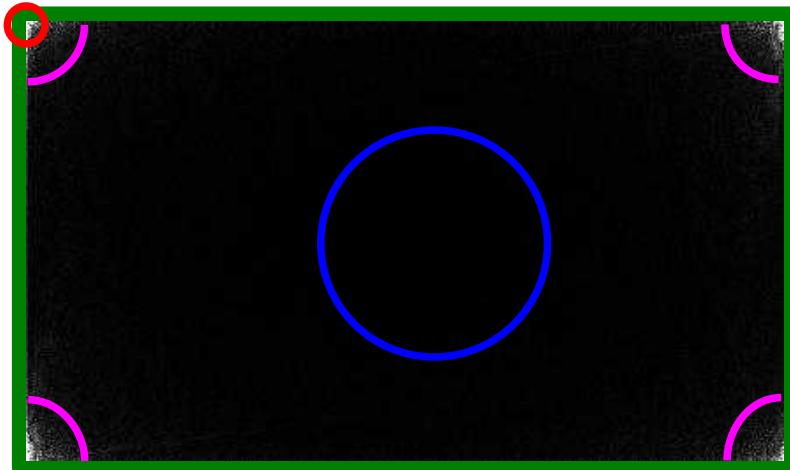
- kolor biały – wartości dodatnie, czarny – ujemne, szary – wartości zerowe
- niewiele tu widać



## Moduł + faza transformaty

$$|FFT( Image)| = \sqrt{\Re(FFT( Image))^2 + \Im(FFT( Image))^2}$$

$$\text{faza } FFT( Image) = \operatorname{tg}^{-1} \frac{\Im(FFT( Image))}{\Re(FFT( Image))}$$



kolor czarny: faza  $-\pi$ , kolor biały: faza  $+\pi$

składowa stała

niskie częstotliwości

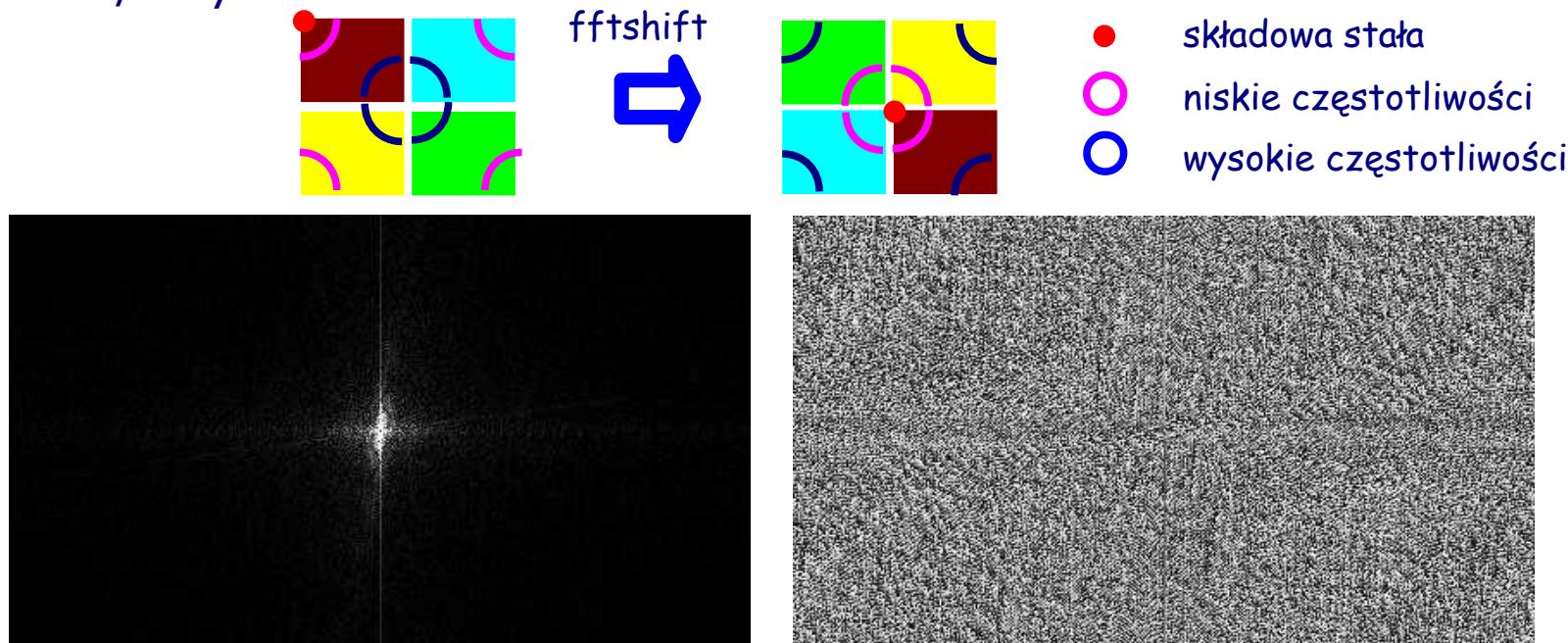
wysokie częstotliwości

- $\text{FFT}(\text{image})[0] = 34721,451 + 0i$  – składowa stała jest dużą liczbą całkowitą (moduł = 34721,451), faza=0°
- składowa stała znacznie przewyższa inne częstotliwości (max = 11221)
- małe częstotliwości są silniej reprezentowane niż wysokie
- obraz nie ma żadnej zauważalnej prawidłowości
- większość informacji ważnej w interpretacji niesie ze sobą moduł – jest on najczęściej używany



## Przesunięcie obrazu transformaty

- na poprzednim rysunku częstotliwości niskie (niosące najwięcej użytecznej informacji) są rozproszone po rogach transformaty
- zwykle gromadzi się je w środku obrazu przez przesunięcie ćwiartek obrazu transformaty – dotyczy to wartości rzeczywistej/zespolonej oraz modułu/fazy



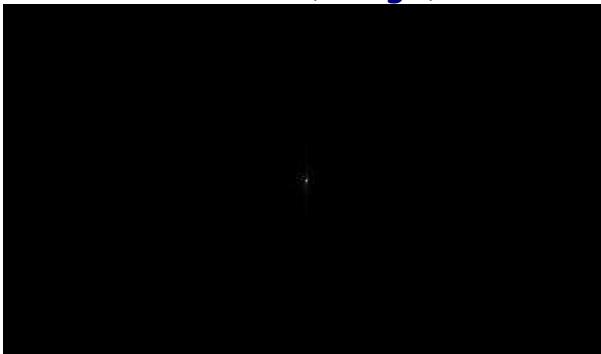
- obraz modułu jest uporządkowany, w obrazie fazy nic to wizualnie nie zmienia
- przed transformata odwrotną trzeba powrócić do stanu początkowego



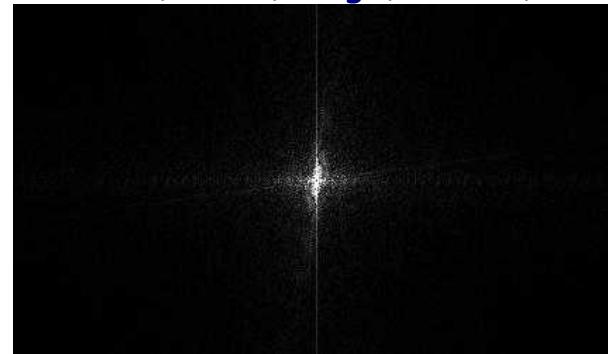
## Logarytmiczna skala modułu

- składowa stała ma dużo większą wartość od pozostałych częstotliwości (na środkowym obrazie została ona obcięta z 34721 do 1000 aby inne częstotliwości nie były czarne – jak na lewym obrazie)
- dla tego w większości wizualizacji przedstawia się logarytm modułu (obraz prawy)

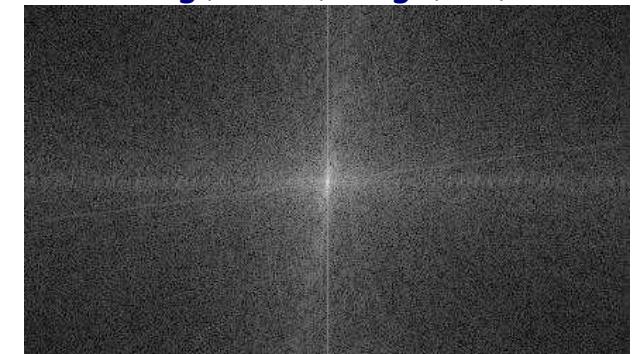
$| \text{FFT}(\text{Image})|$



$\min(| \text{FFT}(\text{Image})|, 1000)$



$\log(| \text{FFT}(\text{Image})|+1)$



- w obrazie fazy nie ma to znaczenia – zakres jest  $(-\pi, +\pi)$
- operacja tylko w celach wizualizacyjnych

# Transformata odwrotna



Image

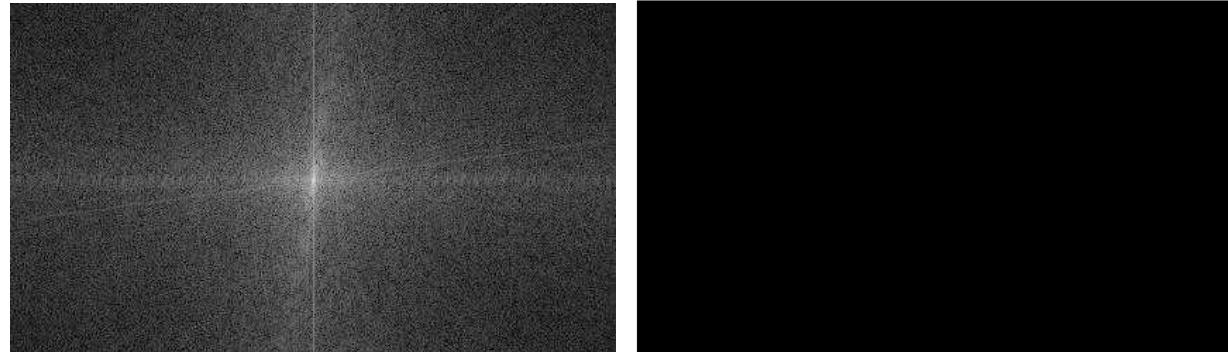


IFFT(FFT(Image))

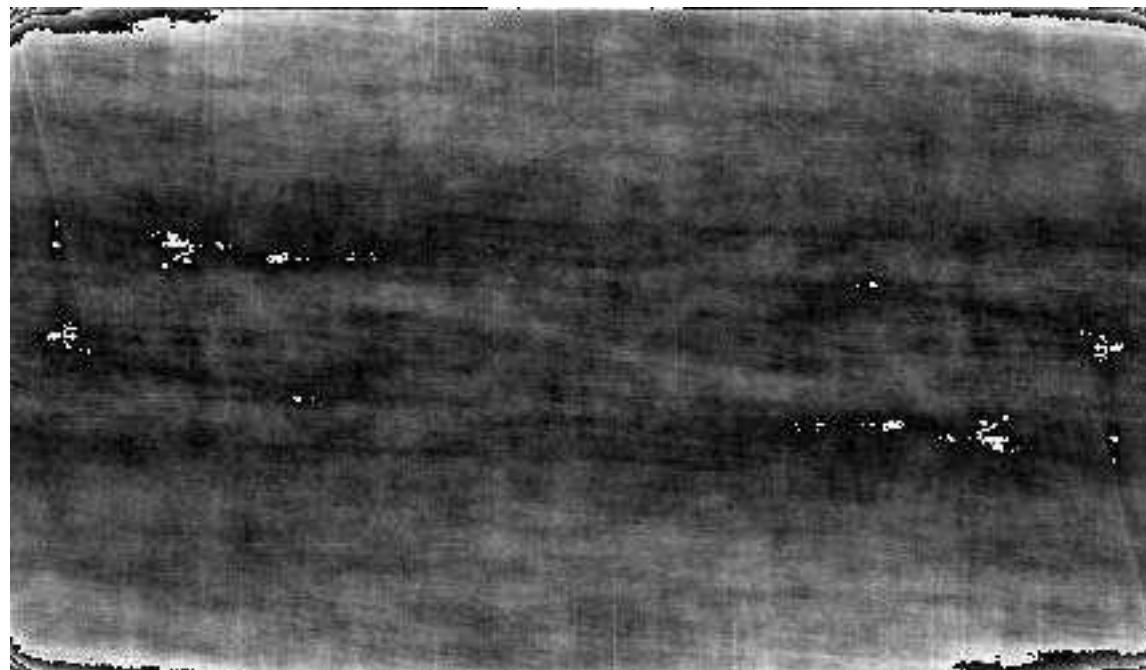
- transformata Fouriera jest odwracalna:  $X = \text{IFT}(\text{FT}(X))$
- DFT jest tylko przybliżeniem transformaty Fouriera
- stąd w DFT pojawia się strata jakości (efekt przybliżeń)



## Transformata odwrotna bez fazy

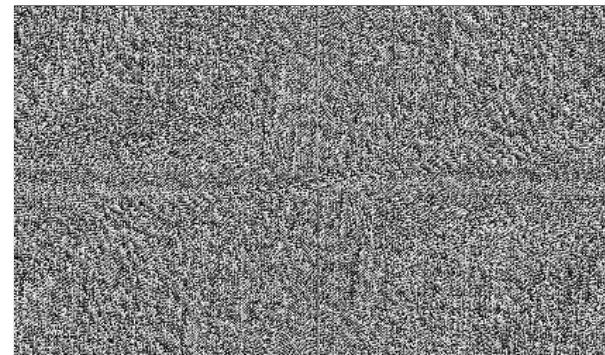
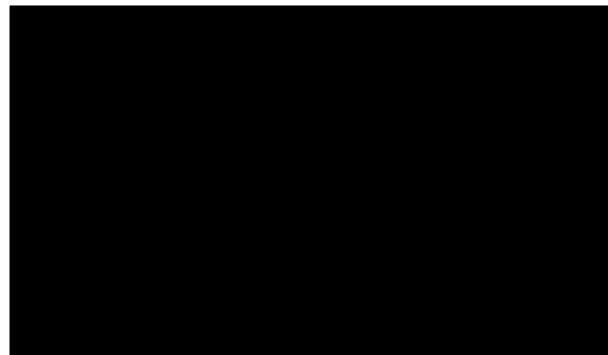


IFFT

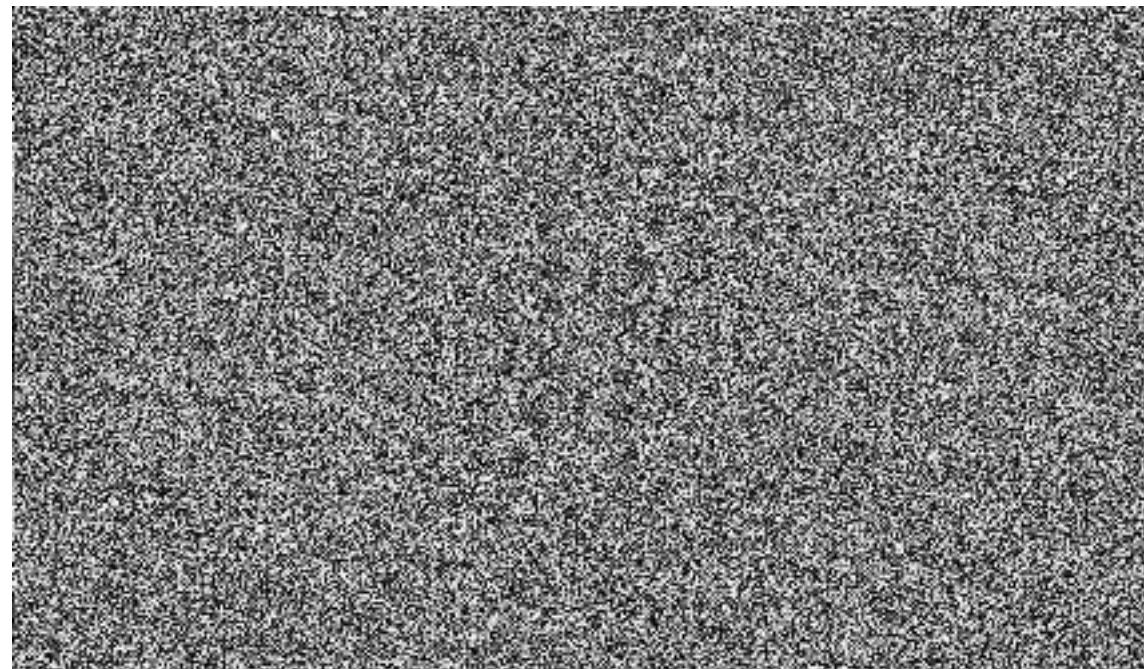


- pomimo pozornego chaosu faza jest niezbędna do odwrócenia transformaty

## Transformata odwrotna bez modułu



IFFT

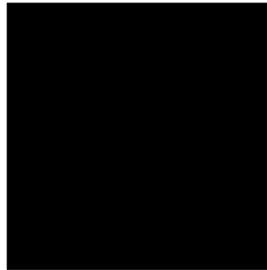
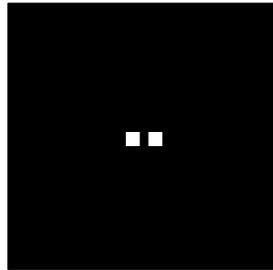


- rekonstrukcja bez modułu też jest niemożliwa



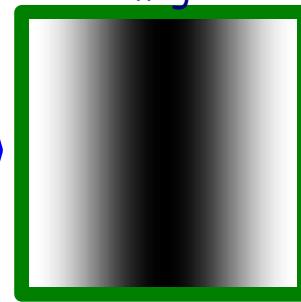
## Funkcje bazowe 2D

Re(FFT(Image)) Im(FFT(Image))



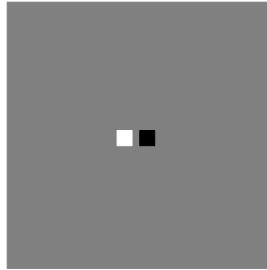
IFFT

Image



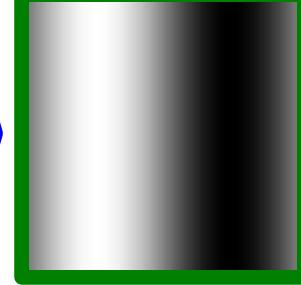
- funkcjami bazowymi są sinusoidy i cosinusoidy 2D

Re(FFT(Image)) Im(FFT(Image))



IFFT

Image



$$\sin\left(2\pi \cdot \frac{x}{N}\right)$$

$$\cos\left(2\pi \cdot \frac{x}{N}\right)$$

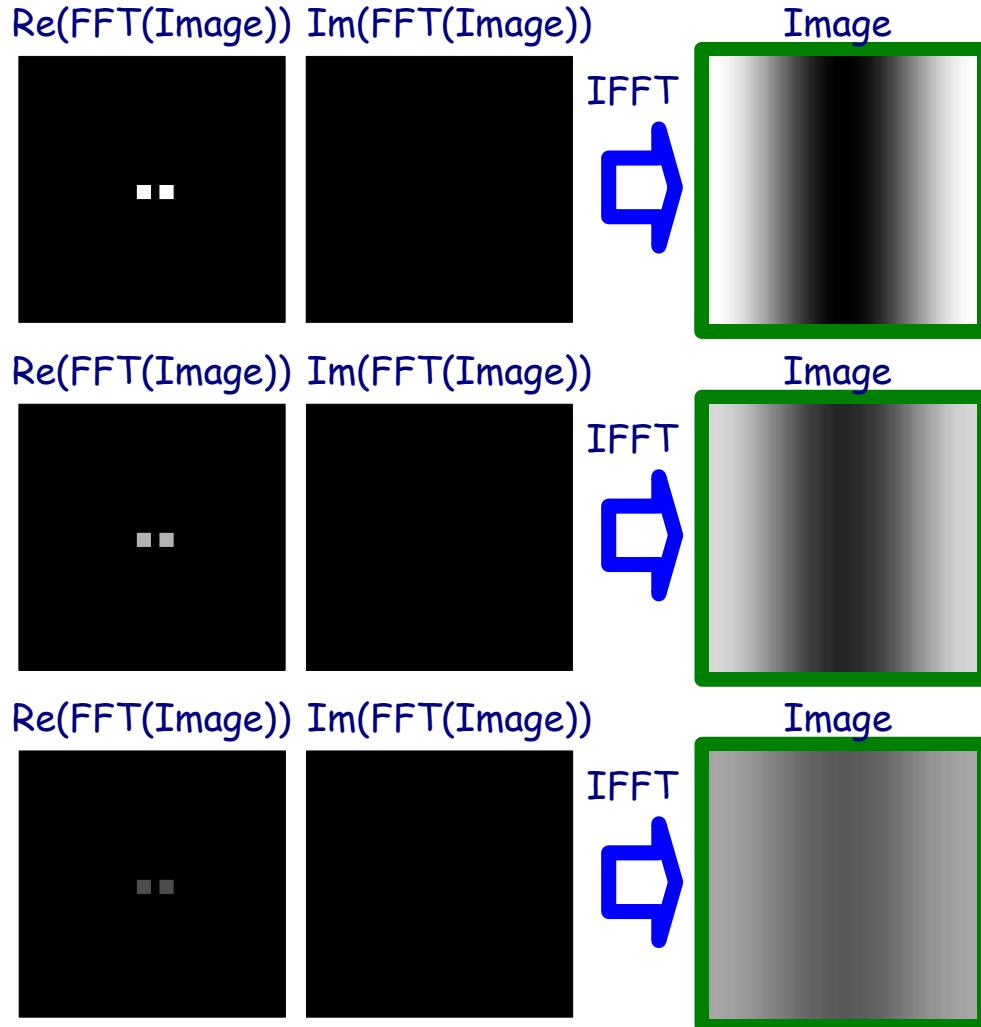
kolor szary: 0

kolor biały: >0

kolor czarny: <0



## Amplituda funkcji bazowych



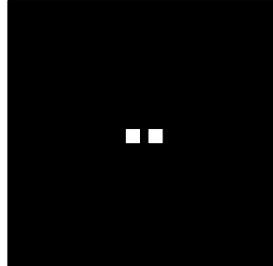
- wysokość „pików” steruje amplitudą (jasnością) funkcji bazowych
- sinusoida staje się mniej kontrastowa – bardziej szara (zblżona do wartości 0 – kolor szary)

kolor szary: 0  
kolor biały: >0  
kolor czarny: <0

## Faza funkcji bazowych



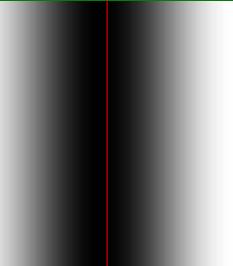
Re(FFT(Image)) Im(FFT(Image))



IFFT



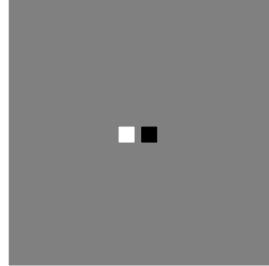
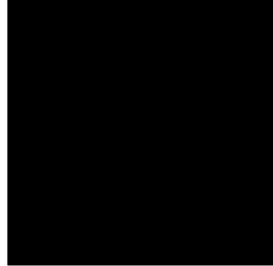
Image



faza(FFT(Image))



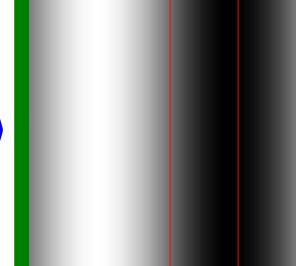
Re(FFT(Image)) Im(FFT(Image))



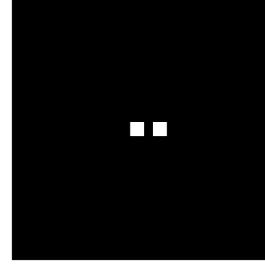
IFFT



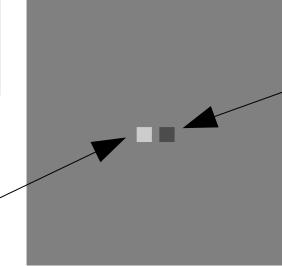
Image



| FFT(Image) |



faza(FFT(Image))



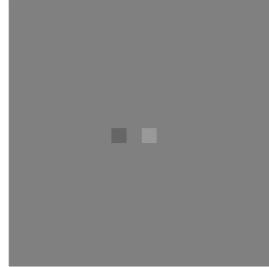
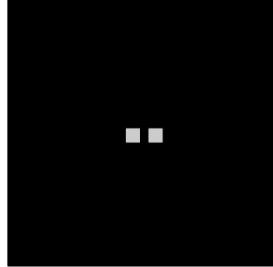
-π/2

+π/2

π/2

- faza steruje przesunięciem f. bazowej (sinus to przesunięty w fazie cosinus)

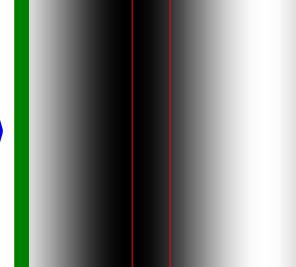
Re(FFT(Image)) Im(FFT(Image))



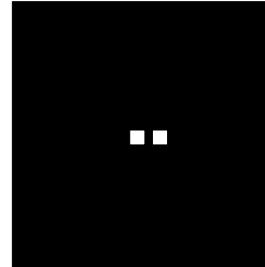
IFFT



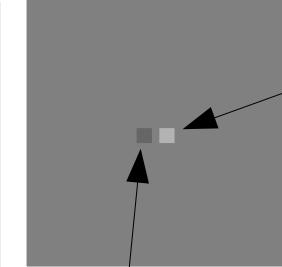
Image



| FFT(Image) |



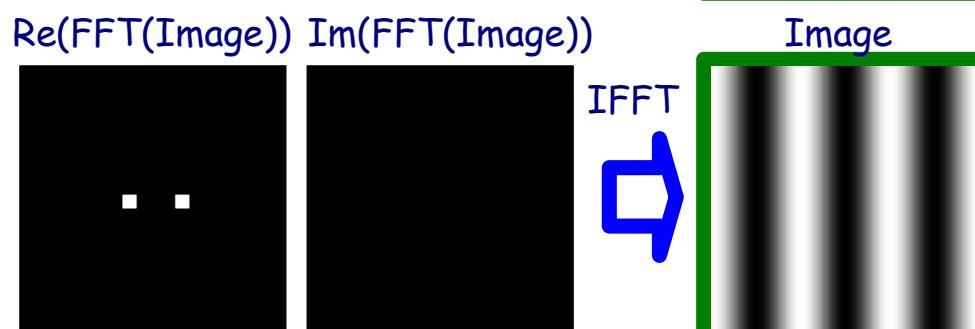
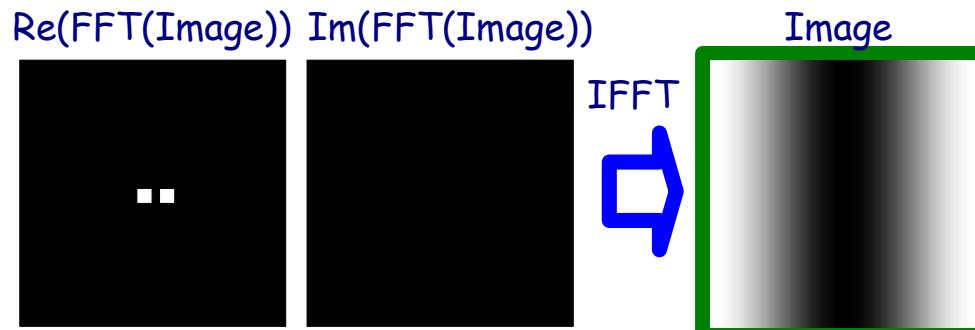
faza(FFT(Image))



+π/4

-π/4

π/4



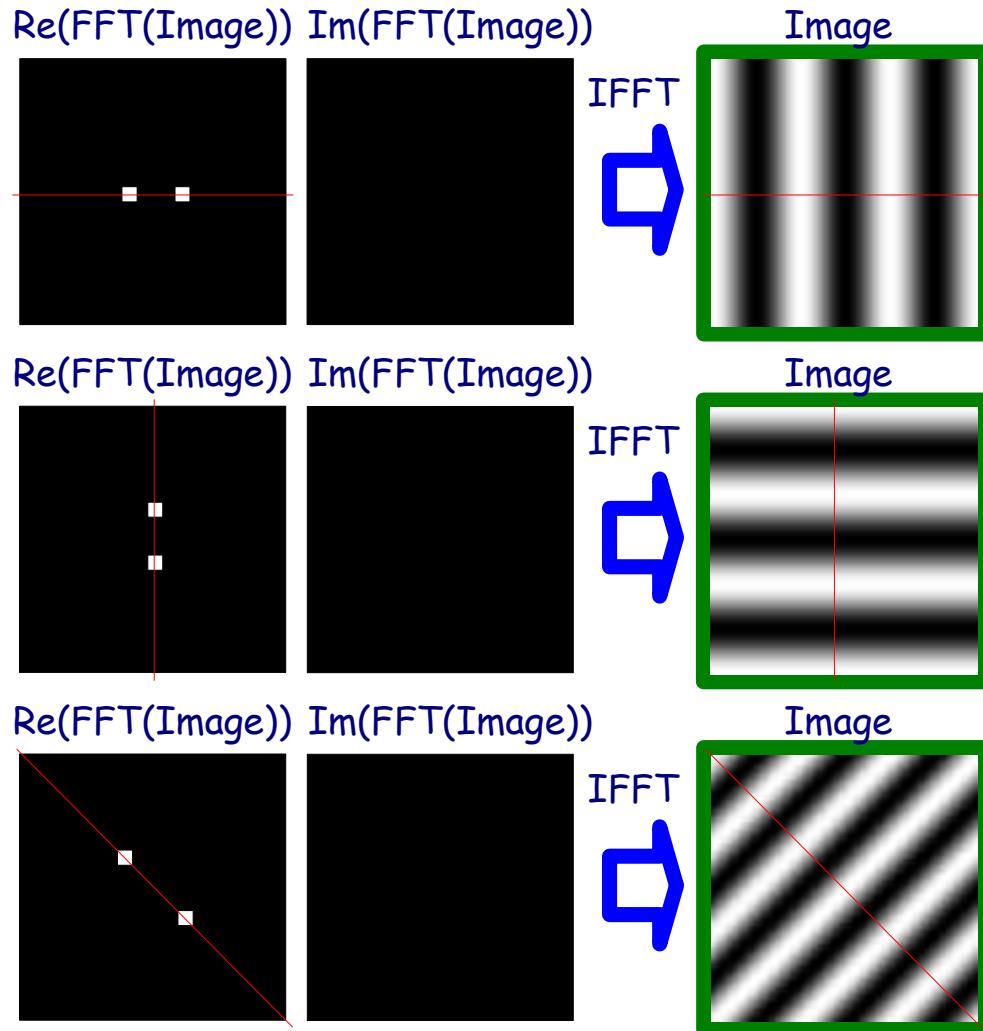
## Częstotliwość funkcji bazowych

- położenie „pików” steruje częstotliwością f. bazowych
- bliżej środka – mniejsze częstot.
- dalej od środka – większe częstot.
- częstotliwość minimalna = 1
- częstotliwość maksymalna =  $N/2$  (tylko 1 „pik” na FFT)

paski biało-czarne o grubości 1



## Orientacja funkcji bazowych

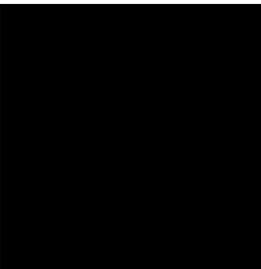
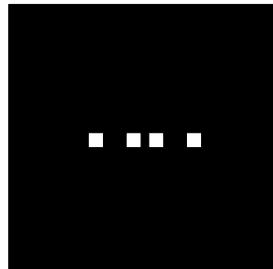


- orientacja funkcji bazowej określona jest przez stosunek składowych  $x$  i  $y$  „pików” w dziedzinie częstotliwości

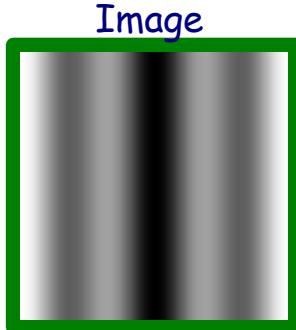


## Składanie funkcji bazowych

Re(FFT(Image)) Im(FFT(Image))

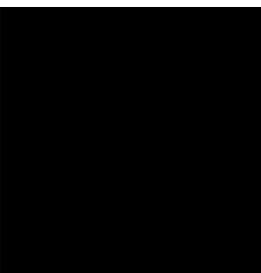
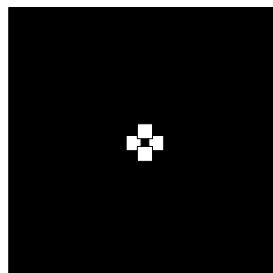


IFFT

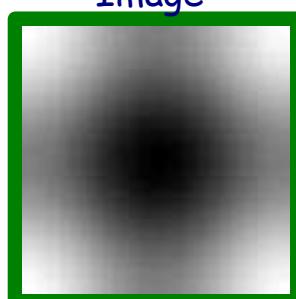


$$\cos\left(2\pi \cdot \frac{x}{N}\right) + \cos\left(2\pi \cdot \frac{3x}{N}\right)$$

Re(FFT(Image)) Im(FFT(Image))

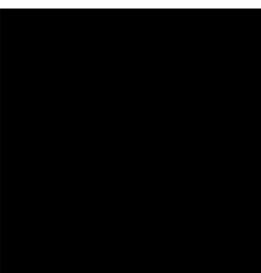
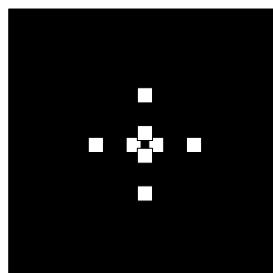


IFFT

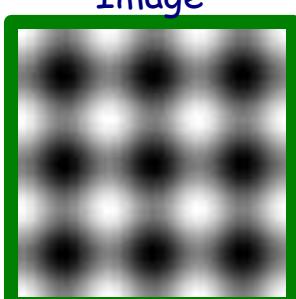


$$\cos\left(2\pi \cdot \frac{x}{N}\right) + \cos\left(2\pi \cdot \frac{y}{N}\right)$$

Re(FFT(Image)) Im(FFT(Image))



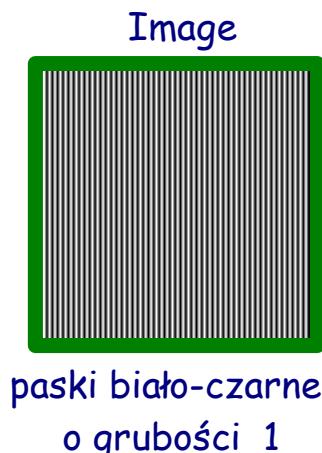
IFFT



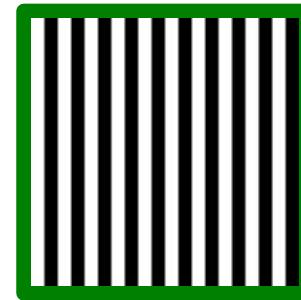
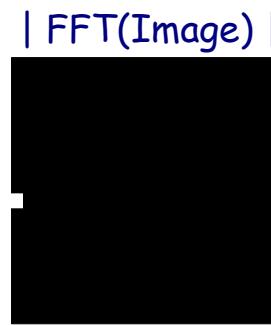
$$\cos\left(2\pi \cdot \frac{3x}{N}\right) + \cos\left(2\pi \cdot \frac{3y}{N}\right)$$



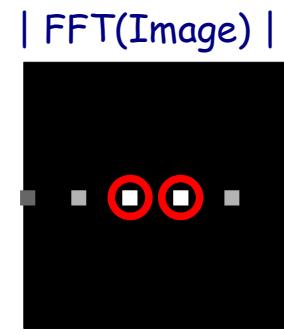
## Transformata wysokich częstotliwości



FFT



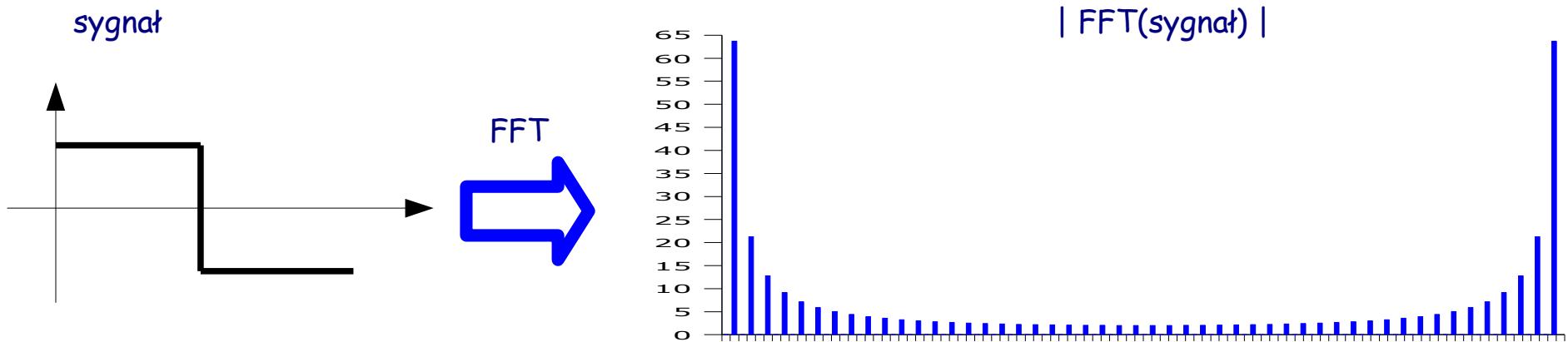
FFT



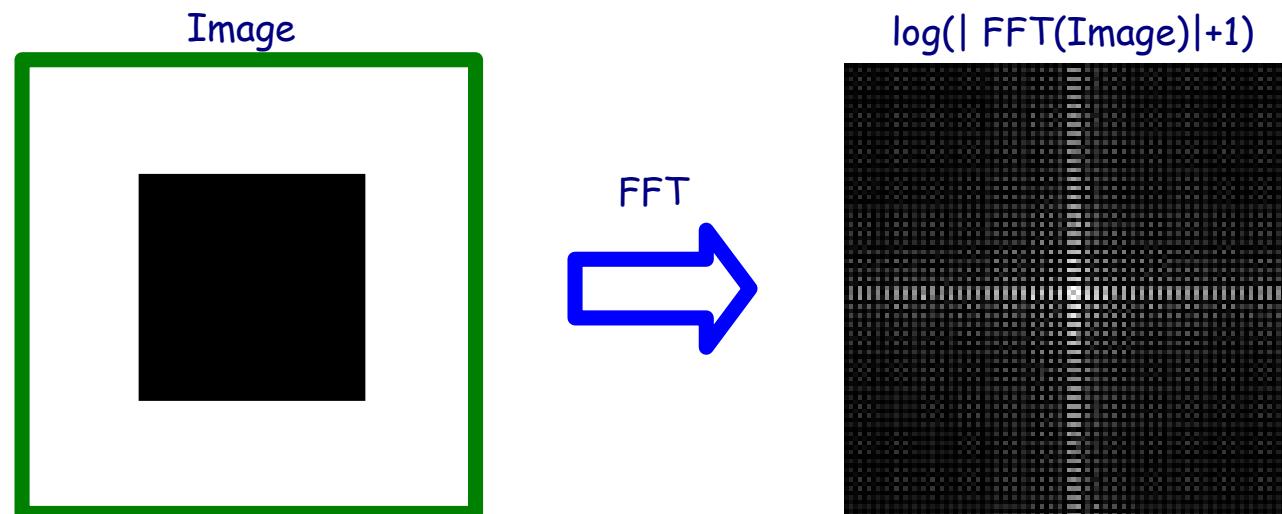
- lewy obraz i jego transformata jest zrozumiała, to maksymalna częstotliwość cosinusa –  $N/2$
- ale prawa ?
- pojawiają się dodatkowe częstotliwości: oprócz głównej ( $N/10$ , w kółkach) również jej wielokrotności:  $N/5$  i  $\sim N/2$  o niższych amplitudach !
- są to składowe harmoniczne – dzieje się tak, ponieważ próbkowany sygnał nie jest już sinusoidą tylko wykresem 0101010101



## Transformata sygnałów nieciągłych

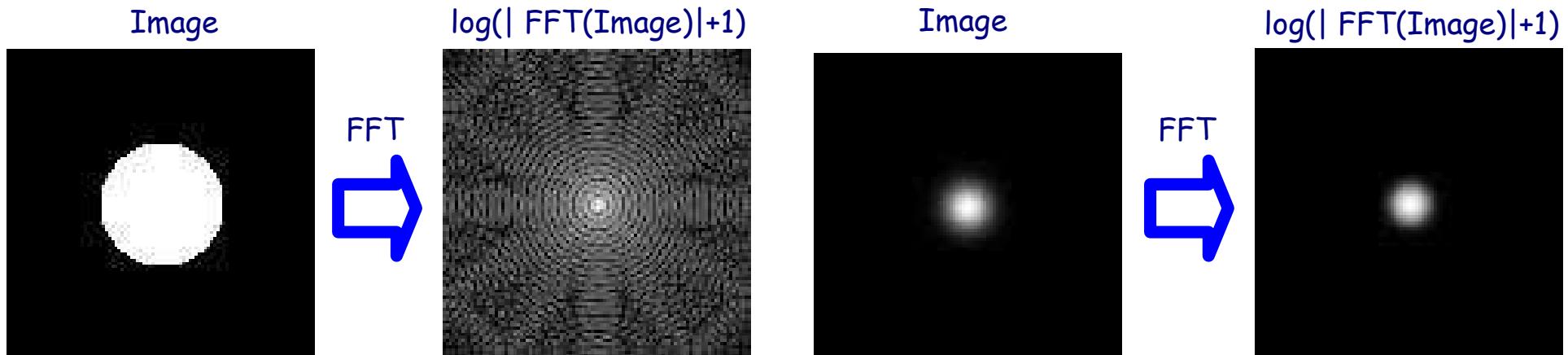


- transformata sygnału ze skokiem wartości jest złożeniem wielu częstotliwości
- tak w 2D jak i w 3D:

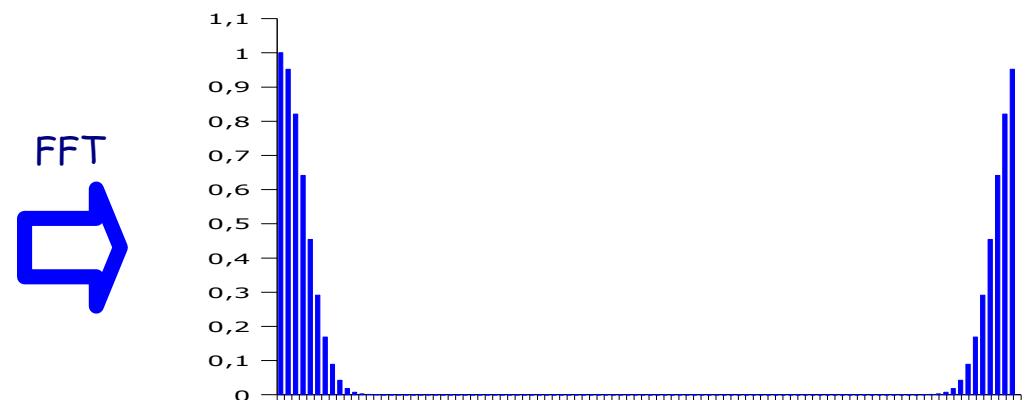
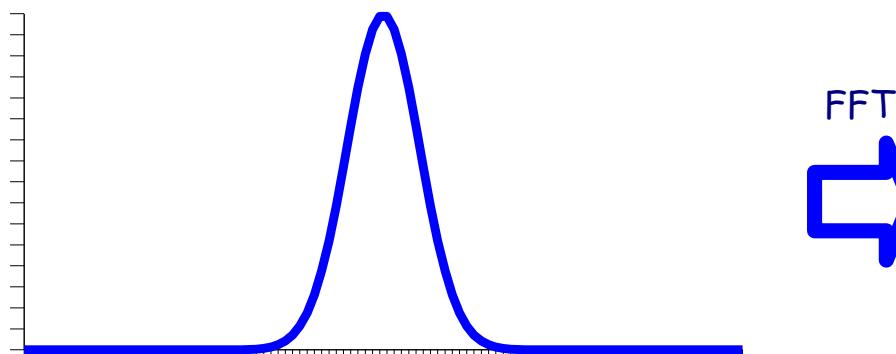




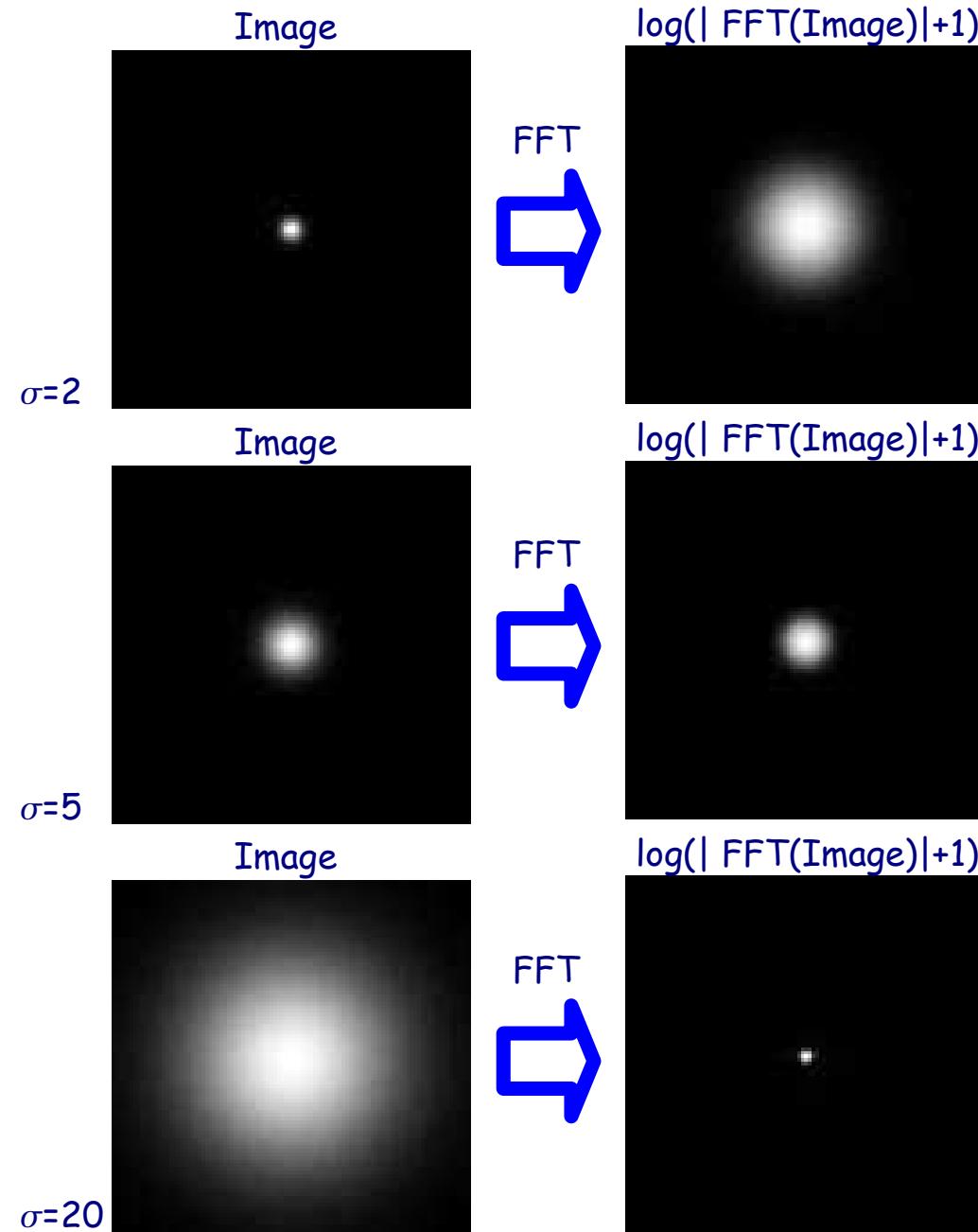
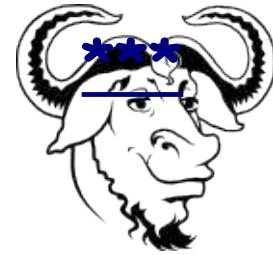
## Transformata rozkładu Gausa



- transformata Fouriera krzywej Gausa też jest krzywą Gausa !!!
- jest to jedna z niewielu form które mają skończony, kompaktowy kształt w dziedzinie częstotliwości i w dziedzinie przestrzennej
- tak w 3D jak i w 2D:



# Zasada nieoznaczoności Heisenberga



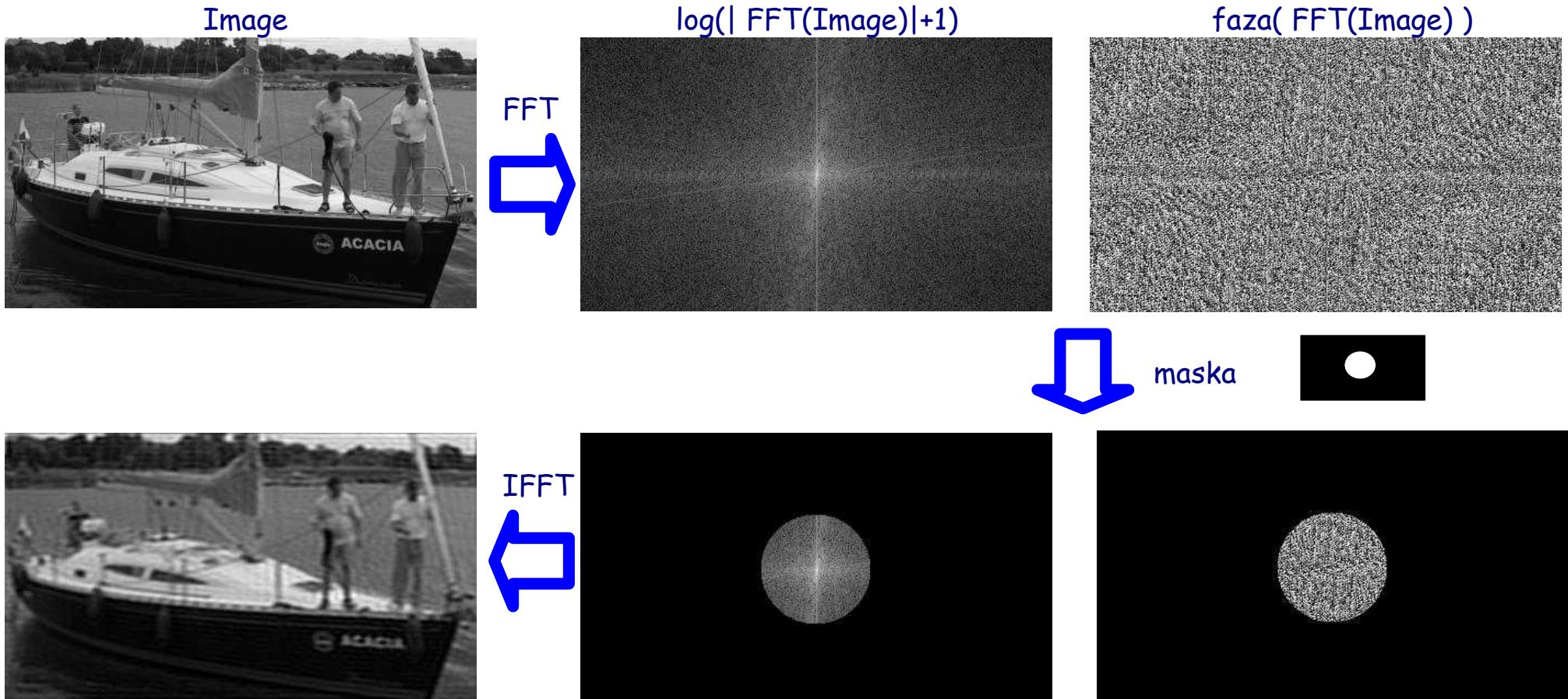
- nie można mieć małego obrazu funkcji Gausa w dziedzinie częstotliwości i przestrzennej
- zwiększając szerokość w dziedzinie przestrzennej zmniejsza się ją w dziedzinie czasu
- iloczyn szerokości (odchylenia standardowego) w dziedzinie częstotliwości i przestrzennej jest stały:

$$\sigma_{xy} \cdot \sigma_f = const$$



## Filtrowanie obrazu: filtr dolnoprzepustowy

- skoro mamy widmo częstotliwości obrazu, można wybrać tylko te, które nas interesują i usunąć pozostałe
- na przykład filtr wygładzający jest filtrem dolnoprzepustowym – zatrzymuje w obrazie tylko niższe częstotliwości i usuwa wyższe:





## Filtrowanie obrazu: maska z krawędziami

- obraz co prawda jest wygładzony, ale pojawiają się zniekształcenia !!!
- jest to wpływ „ostrych” krawędzi maski obcinającej wysokie częstotliwości (które wnoszą do widma wiele składowych harmonicznych)

Image



IFFT(FFT(Image))

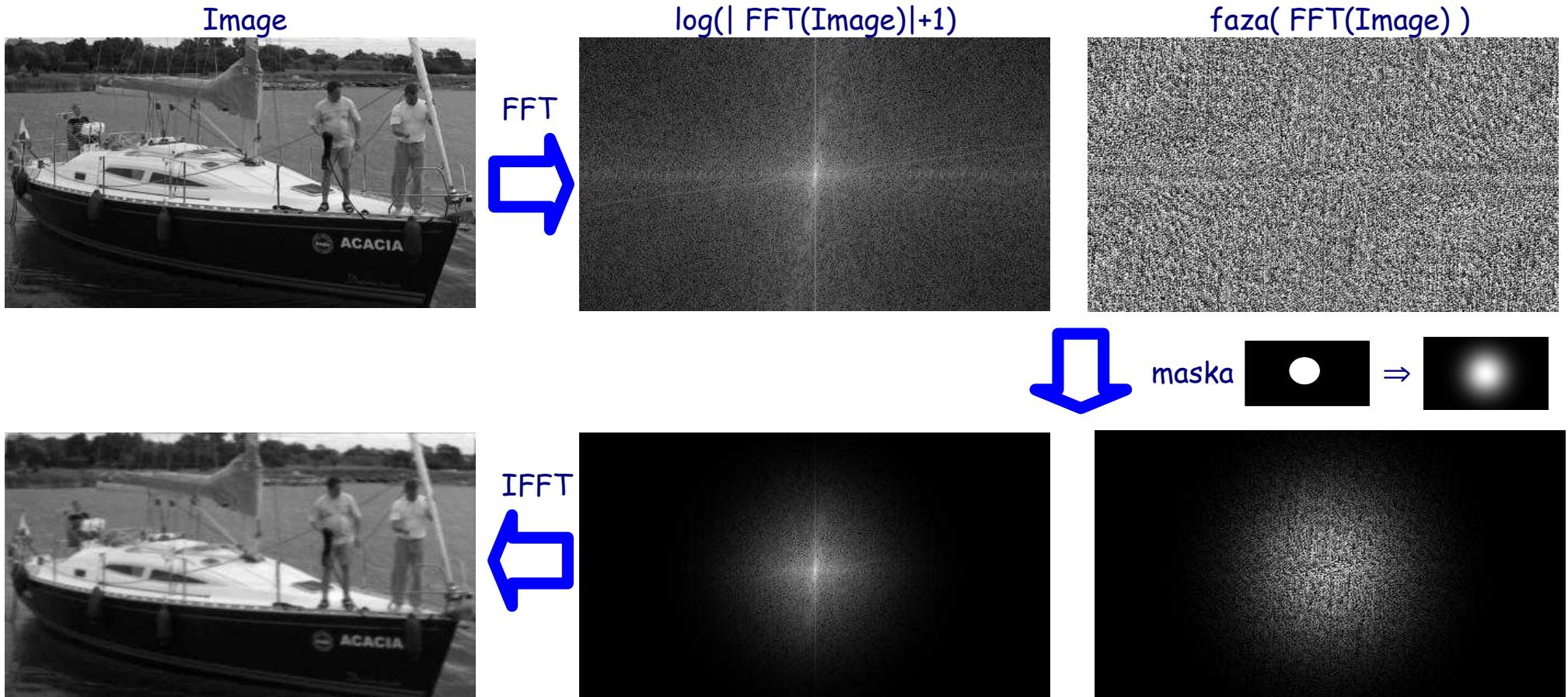


- lepsze efekty da zastosowanie maski w postaci funkcji Gausa 2D ⇒ powstaje filtr **Butterworth'a**



## Filtrowanie obrazu: filtr Butterworth'a

- zastosowana maska jest funkcją Gausa 2D o takim samym odchyleniu standardowym co promień koła poprzedniej maski
- znieksztalcenia znikają !!! a obraz jest wygładzony



IFFT(FFT(Image)) - filtr z maską kołową



IFFT(FFT(Image)) - filtr Butterwortha



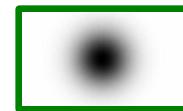
Porównanie:  
filtr Butterworth'a  
vs maska kołowa



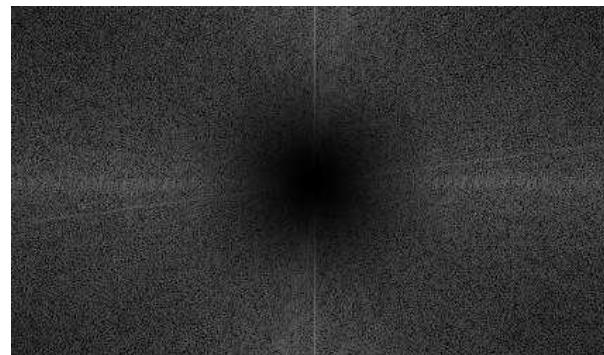
## Filtr Butterworth'a górnoprzepustowy



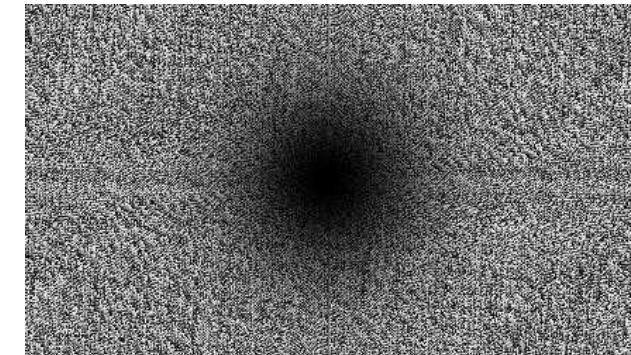
maska



$\text{maska} \cdot * | \text{FFT}(\text{Image}) |$



$\text{maska} \cdot * \text{faza}(\text{FFT}(\text{Image}))$



$\text{IFFT}(\text{maska} \cdot * \text{FFT}(\text{Image}))$



- usunięte są niskie częstotliwości
- wysokie częstotliwości zostają
- wysokie częstotliwości to głównie krawędzie
- uzyskujemy detektor krawędzi
- na podobnej zasadzie filtry typu *band-pass*, *band-stop*

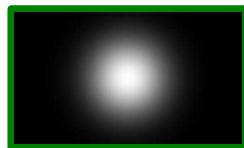
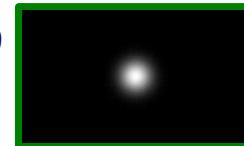


## Czułość filtrów dolnoprzepustowych

- czułością (stopniem rozmycia) można sterować przez zmianę szerokości maski (odchylenia standardowego)
- czym mniejsza, tym rozmycie większe !!!
- odwrotnie niż z maskami konwolucji



maska  $\sigma=20$



maska  $\sigma=50$



maska  $\sigma=100$



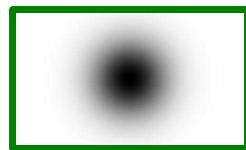
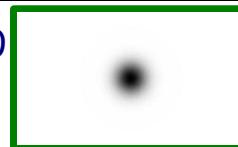


## Czułość filtrów górnoprzepustowych

- czułością na cienkie linie można sterować przez zmianę szerokości maski (odchylenia standardowego)
- usunięcie niewielu niskich częstotliwości zachowuje detale



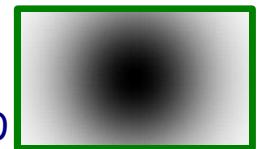
maska-negatyw  $\sigma=20$



maska-negatyw  $\sigma=50$



maska-negatyw  $\sigma=100$





## Wykorzystanie FFT w konwolucji

- konwolucja w dziedzinie przestrzennej odpowiada zwykłemu pomnożeniu elementów macierzy w dziedzinie częstotliwości:

$$FFT( \text{Image} * \text{Mask} ) = FFT(\text{Image}) \cdot * FFT(\text{Mask})$$

- ponieważ FFT jest szybką operacją  $O(N \ln N)$  pojawia się możliwość szybkiego filtrowania w dziedzinie przestrzennej przez tymczasowe przejście do dziedziny częstotliwości:

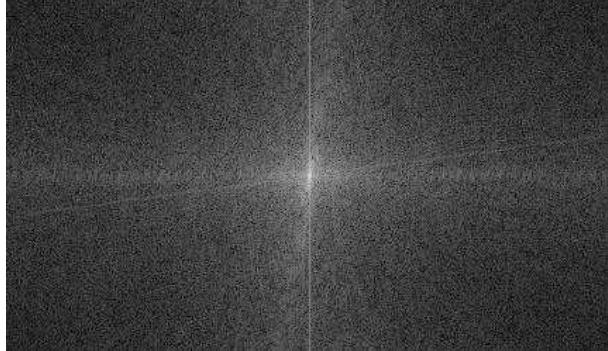
$$\text{Image} * \text{Mask} = IFFT(FFT(\text{Image}) \cdot * FFT(\text{Mask}))$$

- znany już przykład: filtracja z maską Gausa (poprzednie slajdy)

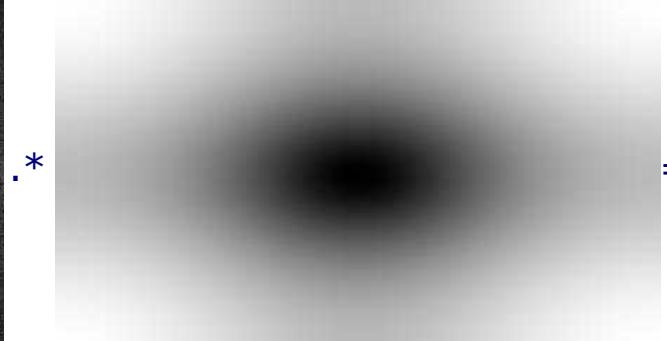
## Schemat konwolucji przez FFT

1. weź maskę konwolucji i obraz *Image*
2. umieść ją w centrum obrazu o wart. 0 o rozmiarach takich jak *Image*
3. oblicz transformaty FFT maski i obrazu i wymnóż ich elementy

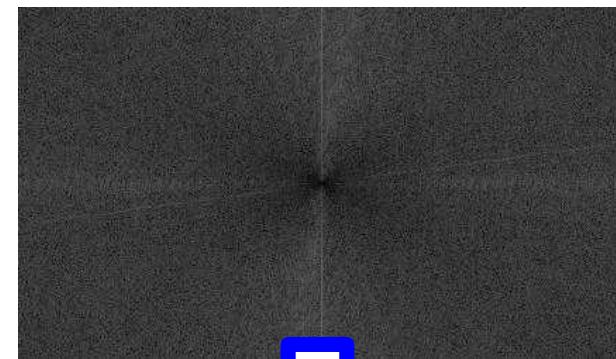
|FFT (Image)|



|FFT (mask) |



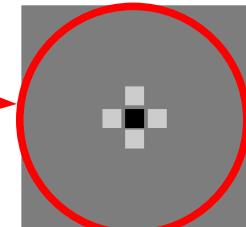
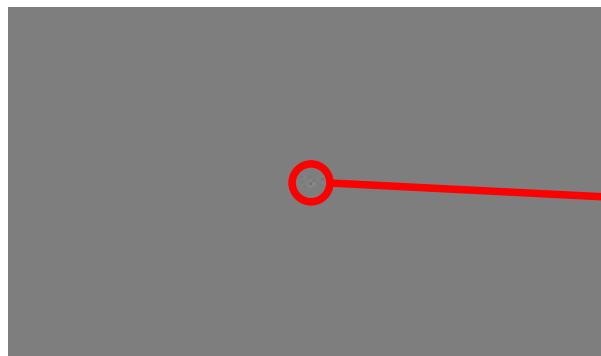
=



IFFT



$$\begin{bmatrix} 0 & +1 & 0 \\ +1 & -4 & +1 \\ 0 & +1 & 0 \end{bmatrix}$$



4. do wyniku zastosuj odwrotną FFT

\* Uwaga: Należy pamiętać o wykonywanych operacjach *fftshift* !



## Porównanie konwolucji przez FFT i zwykłej

- brak znaczących różnic
- FFT zapewnia stały czas, niezależnie od wielkości maski
- złożoność konwolucji w dziedzinie przestrzennej silnie zależy od wielkości maski
- FFT wnosi skomplikowanie
- praktyka:
  - małe maski: zwykła konw.
  - duże maski: przez FFT



## Selektywne filtrowanie częstotliwości: filtry Gabora

- pojedynczy filtr Gabora pozwala przefiltrować obraz zachowując tylko precyzyjnie dobrany zakres częstotliwości
- pozwala to zrobić zarówno w dziedzinie częstotliwości za pomocą konwolucji z wyliczoną maską, jak i w dziedzinie częstotliwości
- maska konwolucji ma zespolone współczynniki (choć istnieją też wersje ze współczynnikami rzeczywistymi)
- maska w dziedzinie częstotliwości jest funkcją Gausa (*koperta, envelop*) modulowaną (zespoloną) sinusoidą (*nośnik, carrier*)



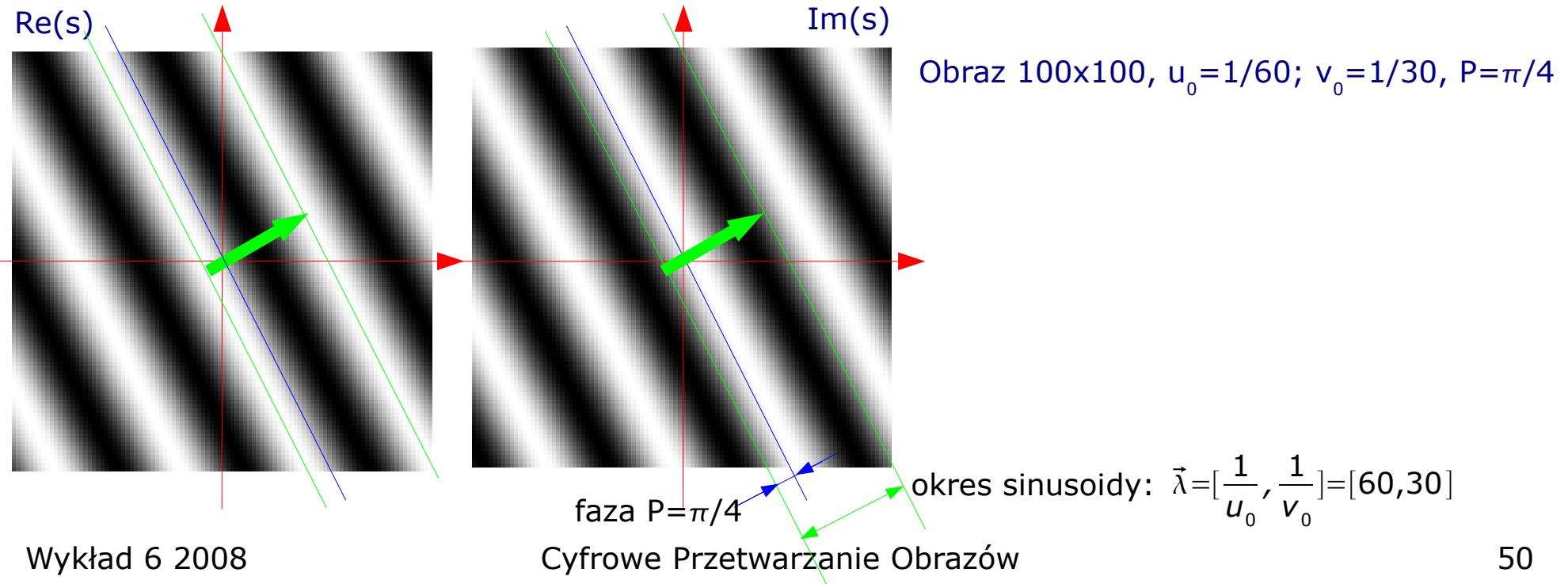
## Nośnik maski filtru Gabora

- nośnik jest zespoloną sinusoidą:

$$s(x, y) = e^{i \cdot (2\pi(u_0 \cdot x + v_0 \cdot y) + P)} = \cos(2\pi(u_0 \cdot x + v_0 \cdot y) + P) + i \cdot \sin(2\pi(u_0 \cdot x + v_0 \cdot y) + P)$$

gdzie:  $(u_0, v_0)$  – określa częstotliwość zespolonej sinusoidy  
 $P$  – fazę tej sinusoidy

- część rzeczywista – cosinusoida, część urojona – sinusoida,  
obie przesunięte w fazie  $P$
- częstotliwość w kierunku  $x$ :  $N \cdot u_0$ , w kierunku  $y$ :  $N \cdot v_0$





## Koperta maski filtru Gabora

- kopertą jest funkcja Gaussa:

$$w_r(x, y) = K \cdot e^{-\pi(a^2(x-x_0)^2 + b^2(y-y_0)^2)}$$

gdzie:  $K$  – czynnik skalujący

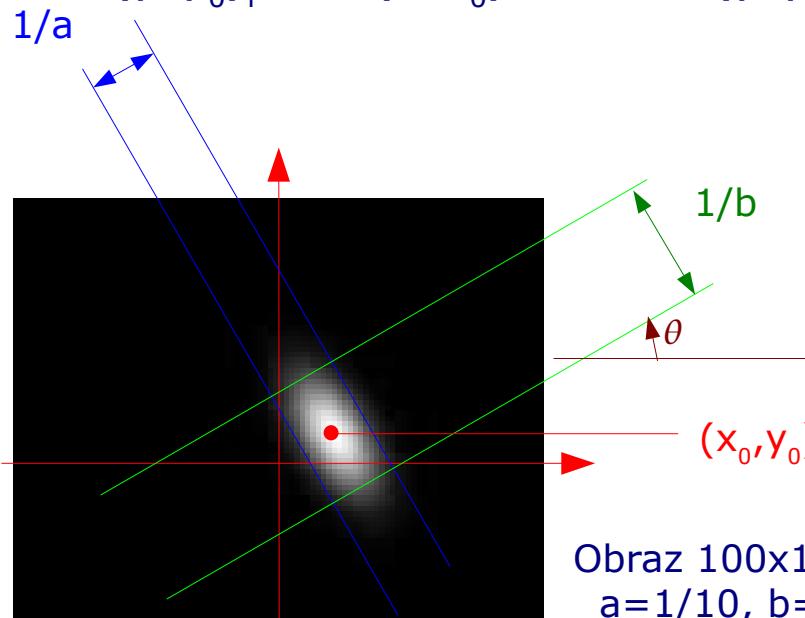
$a, b$  – skalują 2 osie funkcji Gausa

$(x_0, y_0)$  – środek funkcji Gausa

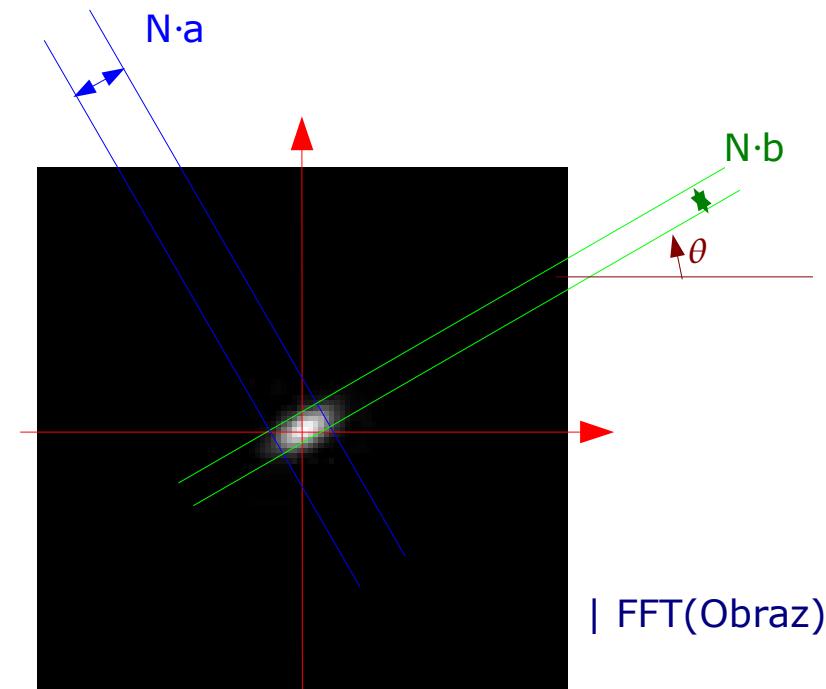
$\theta$  – kąt obrotu funkcji Gausa wokół  $(x_0, y_0)$  – nie ma znaczenia jeśli  $a=b$

$$(x-x_0)_r = (x-x_0) \cos \theta + (y-y_0) \sin \theta$$

$$(y-y_0)_r = - (x-x_0) \sin \theta + (y-y_0) \cos \theta$$



Obraz  $100 \times 100$ ,  
 $a=1/10$ ,  $b=1/20$ ,  
 $(x_0, y_0)=(10,5)$ ,  
 $K=5$ ,  $\theta=\pi/6$

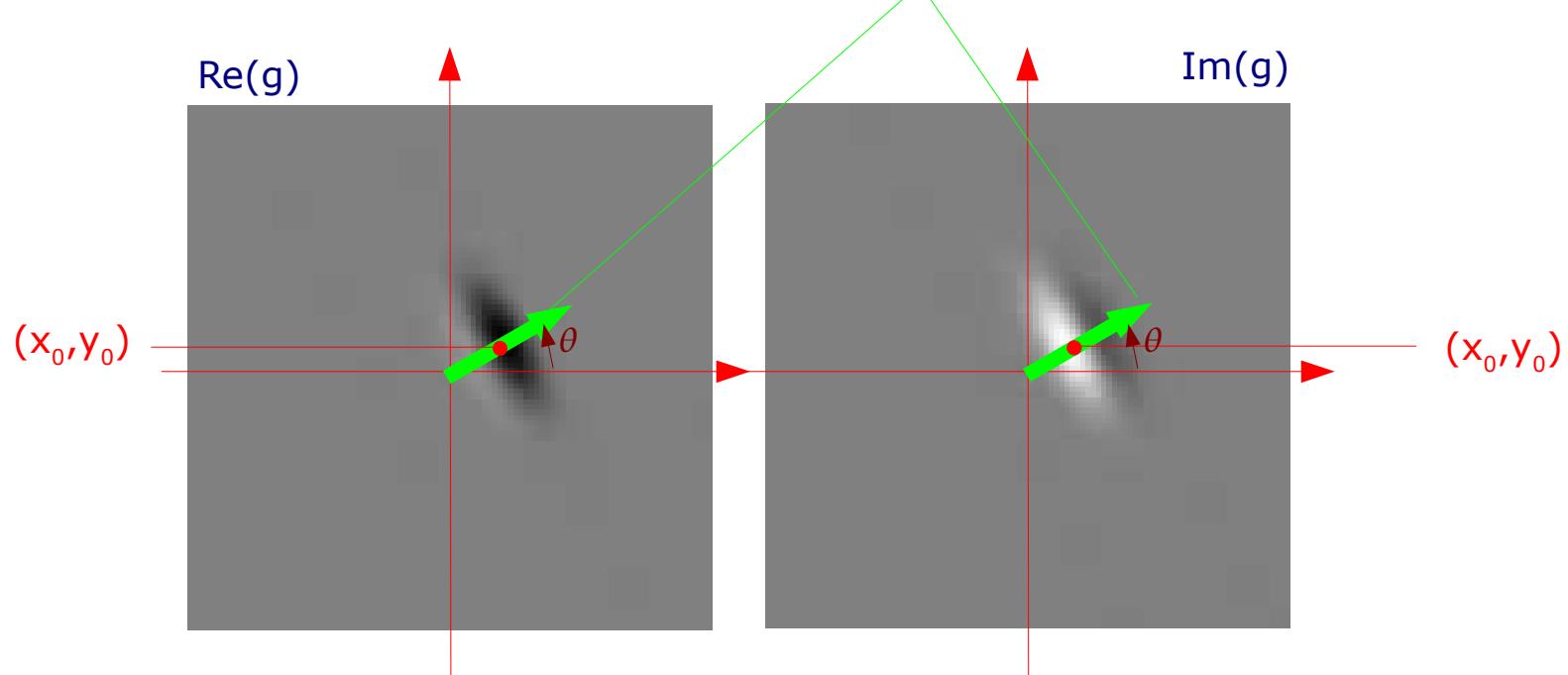




## Maska filtru Gabora

$$g(x, y) = K \cdot e^{-\pi(a^2(x-x_0)^2 + b^2(y-y_0)^2)} \cdot e^{i \cdot (2\pi(u_0 \cdot x + v_0 \cdot y) + P)}$$

$$\text{okres sinusoidy: } \vec{\lambda} = \left[ \frac{1}{u_0}, \frac{1}{v_0} \right] = [60, 30]$$

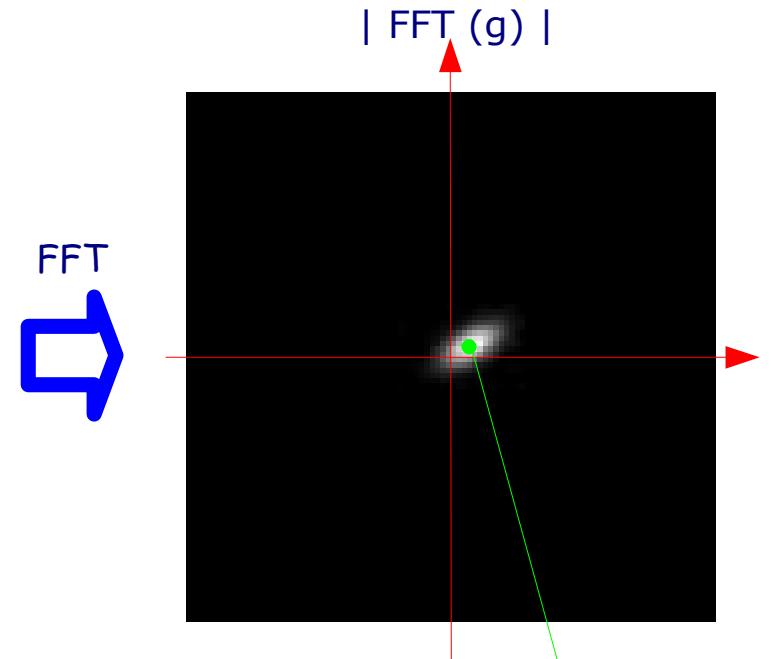
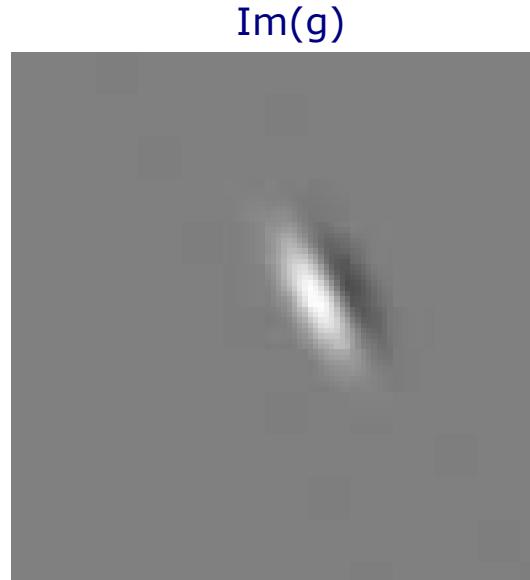
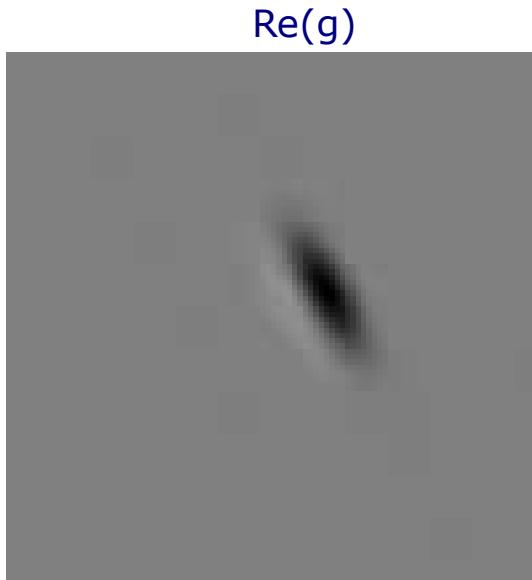


Obraz 100x100,  $a=1/10$ ,  $b=1/20$ ,  $(x_0, y_0)=(10,5)$ ,  $K=5$ ,  $\theta=\pi/6$ ,  $u_0=1/30$   $v_0=1/60$ ,  $P=0$

- maska jest iloczynem sygnału-nośnika z ograniczającą ją kopertą



## Maska filtru Gabora w dziedzinie częstotliwości



$$(N \cdot u_0, N \cdot v_0) = (3.333, 1.666)$$

$$FFT(g)(u, v) = \frac{K}{ab} \cdot e^{-i \cdot (2\pi(x_0(u-u_0) + y_0(v-v_0)) - P)} \cdot e^{-\pi \left( \frac{(u-u_0)_r^2}{a^2} + \frac{(v-v_0)_r^2}{b^2} \right)}$$

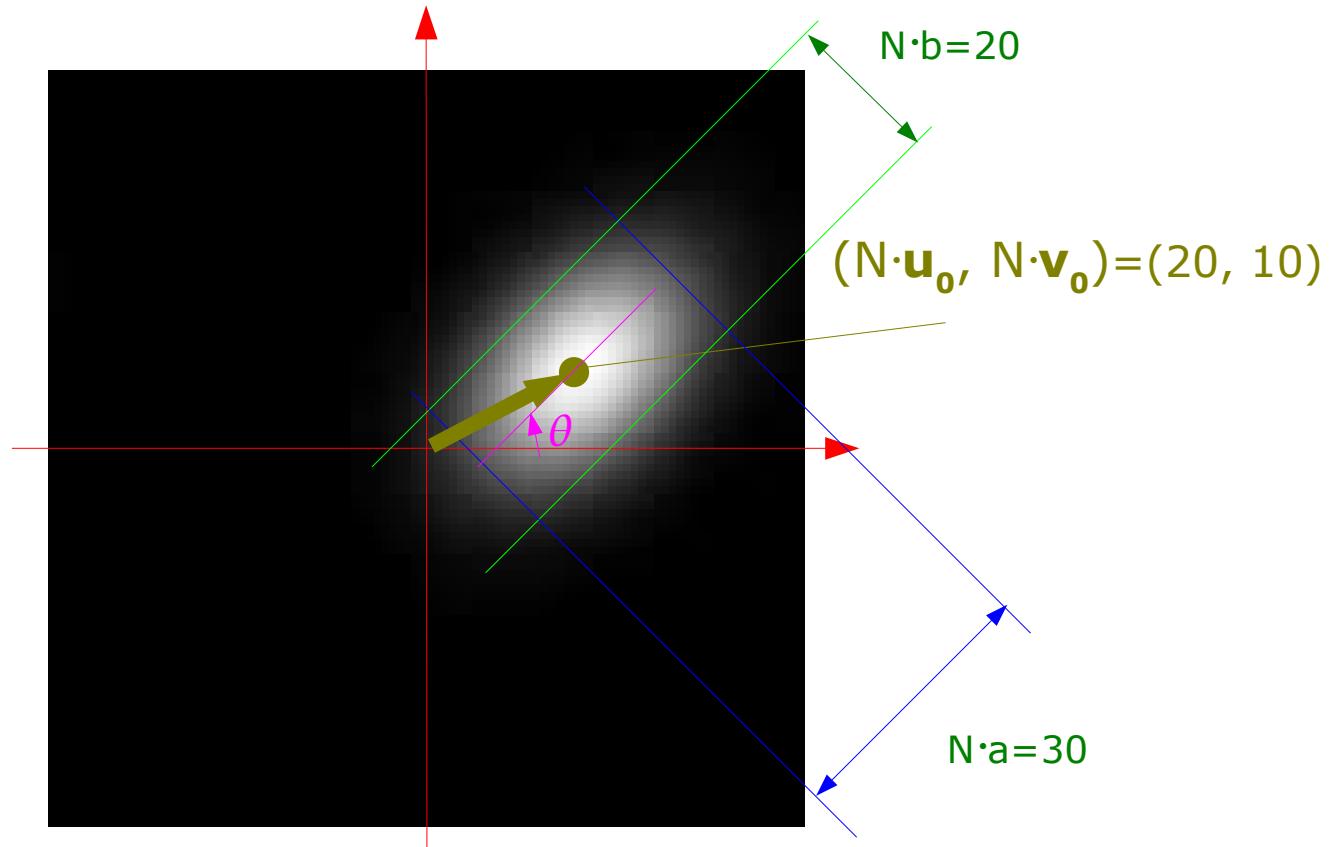
$$|FFT(g)(u, v)| = \frac{K}{ab} \cdot e^{-\pi \left( \frac{(u-u_0)_r^2}{a^2} + \frac{(v-v_0)_r^2}{b^2} \right)}$$

$$\text{faza}(u, v) = -2\pi(x_0(u-u_0) + y_0(v-v_0)) + P$$



## Zakres częstotliwości filtru Gabora

Obraz 100x100,  $\mathbf{a}=0.3$ ,  $\mathbf{b}=0.2$ ,  $(\mathbf{x}_0, \mathbf{y}_0)=(10, 5)$ ,  $\mathbf{K}=5$ ,  $\theta=\pi/4$ ,  $\mathbf{u}_0=1/5$   $\mathbf{v}_0=1/10$ ,  $\mathbf{P}=0$



- parametry  $P$  oraz  $(x_0, y_0)$  nie mają wpływu na zakres częstotliwości tylko na jej fazę



## Filtrowanie

- filtrowanie z tak obliczoną maską w dziedzinie przestrzennej to zwykła konwolucja, ale wynik jest zespolony !!!
- do filtrowania używa się tylko środka maski – jej brzegi zawierają wartości 0
- filtrować można też w dziedzinie częstotliwości – po przejściu przez FFT – wynik jest ten sam
- ma to zwykle zdarzenie, bo aby wybrać precyzyjnie częstotliwości  $a$  i  $b$  muszą być małe a zatem koperta Gausa jest duża, więc maski w dziedzinie przestrzennej są duże



## Filtrowanie częścią rzeczywistą i urojoną maski

- filtrowanie w dziedzinie częstotliwości
- maska  $41 \times 41$  – długi czas obliczeń
- Maska  $100 \times 100$ ,  $\mathbf{a}=0.1$ ,  $\mathbf{b}=0.05$ ,  $(\mathbf{x}_0, \mathbf{y}_0)=(10,5)$ ,  $\mathbf{K}=5$ ,  $\theta=\pi/4$ ,  $\mathbf{u}_0=1/5$   $\mathbf{v}_0=1/10$ ,  $\mathbf{P}=0$

mr=Re(g(30:70,30:70))

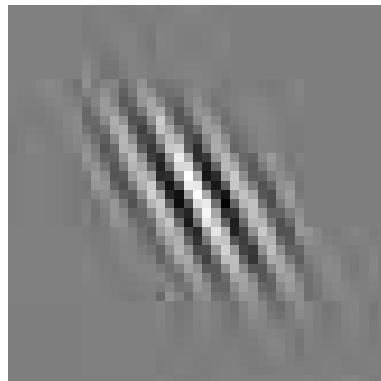
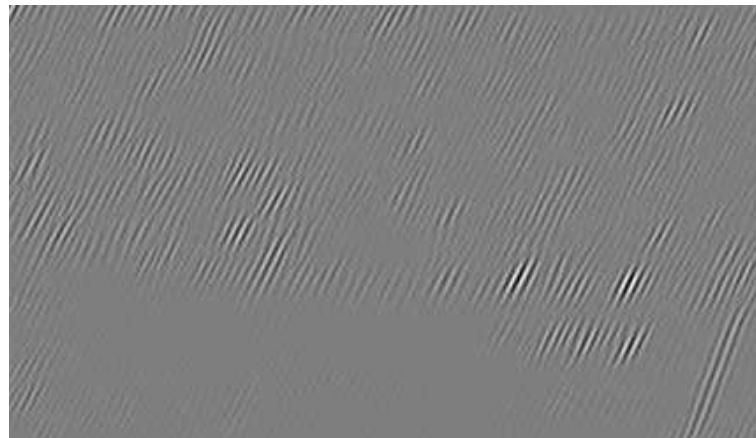


Image \* mr



| Image \* mr |



mi=Im(g(30:70,30:70))

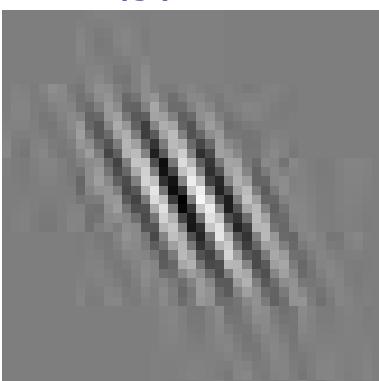
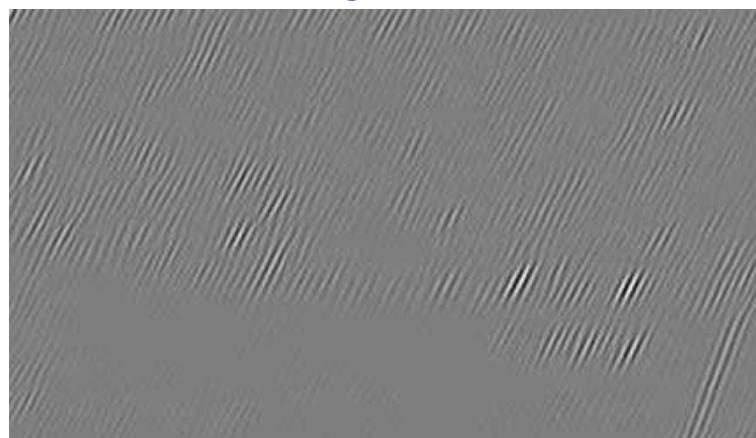


Image \* mi



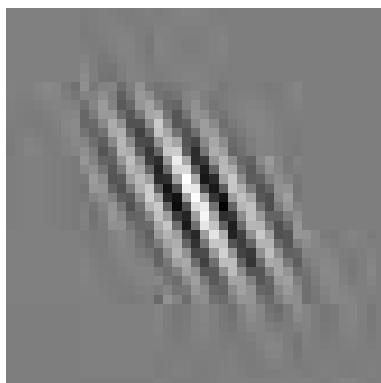
| Image \* mi |





## Filtrowanie całością maski

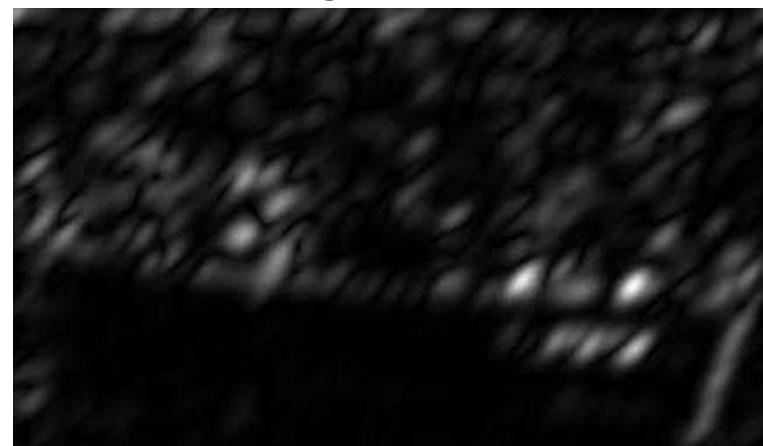
- jedynie filtrowanie maską zespoloną daje ciągłe wyniki
- wzięcie tylko części rzeczywistej lub urojonej daje wyniki „poszatkowane”



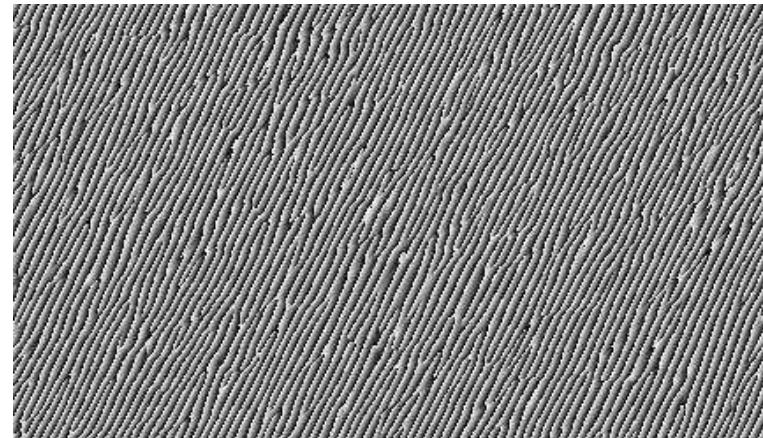
**maska 41x41  
zespolona !!!**



| Image \* maska |



faza (Image \* maska)

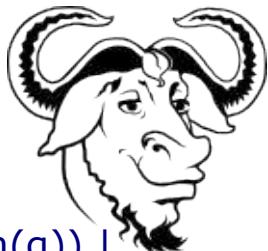


| Re (Image \* maska) |

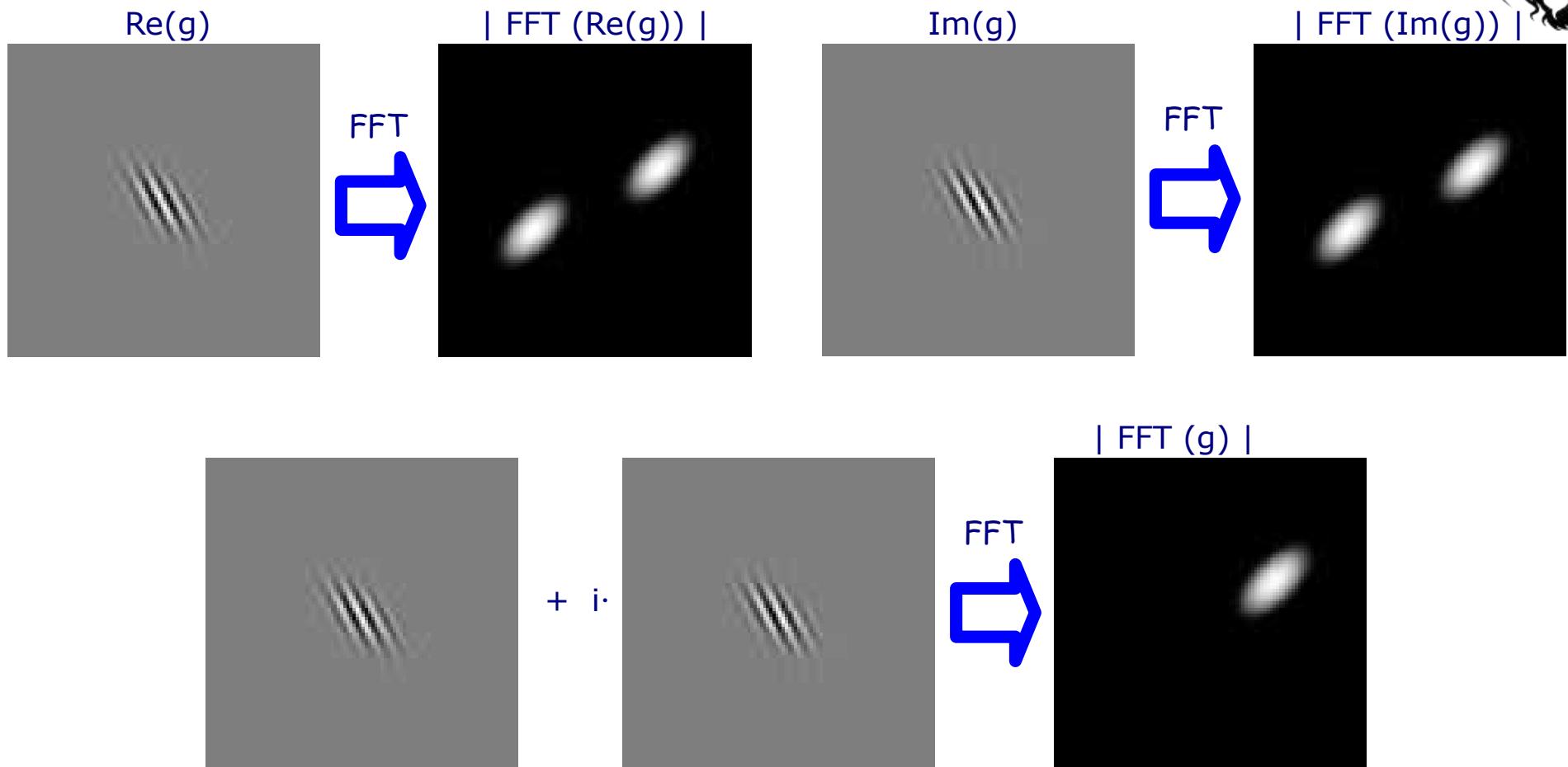


| Im (Image \* maska) |





## Transformata FFT składowych maski

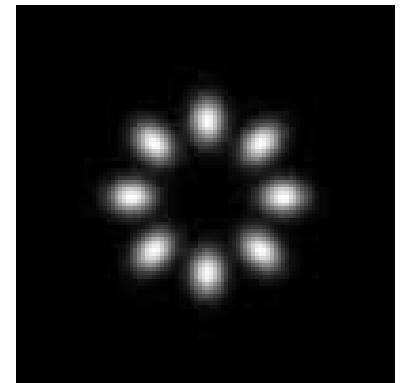


- wzięcie tylko części rzeczywistej lub urojonej maski powoduje dodanie drugiej filtrowanej częstotliwości
- pełna zespolona maska nie powoduje tego efektu



## Bank filtrów Gabora

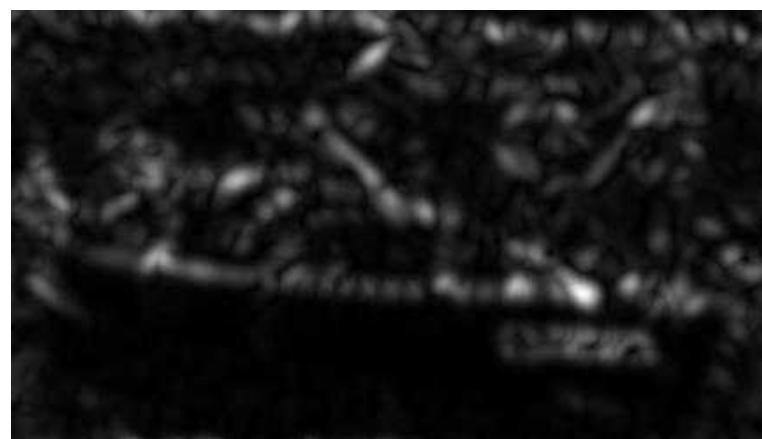
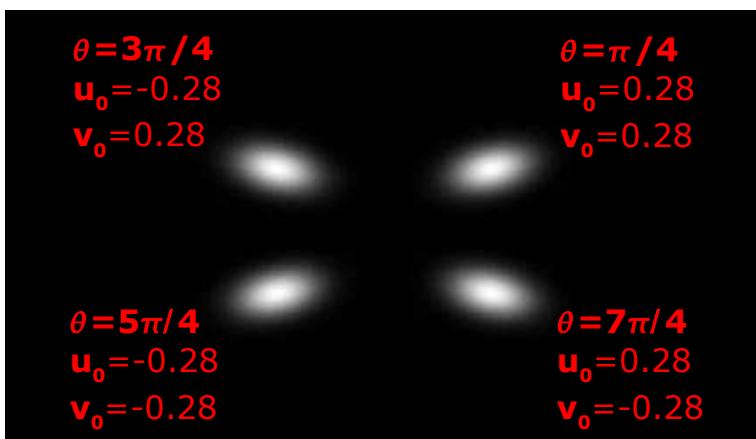
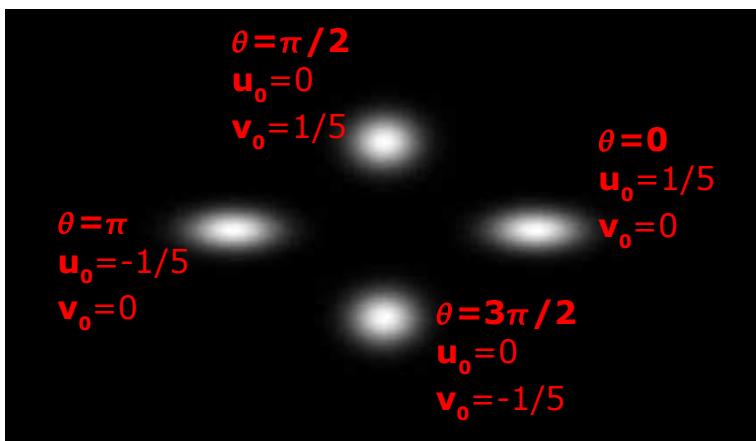
- pojedyncza zespolona maska wyodrębnia jedną częstotliwość określoną przez parametry  $(u_0, v_0)$  – z pewnym marginesem określonym przez  $a$ ,  $b$  i  $\theta$
- jeśli jesteśmy zainteresowani zbiorem częstotliwości należy zastosować kilka masek, z których każda wykrywa jedną częstotliwość
- często chodzi nam o jeden moduł częstotliwości i kilka orientacji
- tworzy się wtedy bank filtrów Gabora:
  - moduł częstotliwości jest stały,
  - zmieniają się kierunki: „płatki” banku leżą na okręgu
  - orientacja „płatka” określona przez  $\theta$  zmienia się aby orientować go do środka – częstotliwości 0
  - pojedynczy pątek to obraz FFT jednej maski
- każdy z filtrów daje jedną odpowiedź – tworzy się przez to tzw. **przestrzeń Gabora**
- przestrzeń tę można dalej przetwarzać, np. w celu klasyfikacji, np: bierzemy wartość maksymalną





## Orientacja płatków

- orientacja „steruje” wykrywanymi częstotliwościami
- parametry stałe: **moduł** = 0.2, **a**=0.1, **b**=0.075,  $(x_0, y_0) = (10, 5)$ , **K**=1, **P**=0
- parametry zmienne:  $\theta$ ,  $u_0$ ,  $v_0$  zależne od orientacji „płatków”
- w poniższych przykładach wykrywane są linie pionowe i poziome, potem ukośne

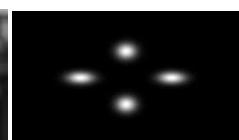




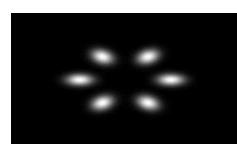
## Liczba płatków



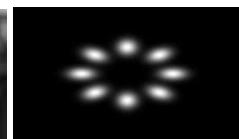
- czym więcej płatków, tym więcej „pokrytych” kierunków
- obraz się „zapełnia” – coraz więcej w nim szczegółów
- parametry stałe: **moduł** = 0.2, **a**=0.1, **b**=0.075, **(x<sub>0</sub>,y<sub>0</sub>)**=(10,5), **K**=1, **P**=0
- parametry zmienne:  $\theta$ , **u<sub>0</sub>**, **v<sub>0</sub>** zależne od  $n$  – liczby płatków



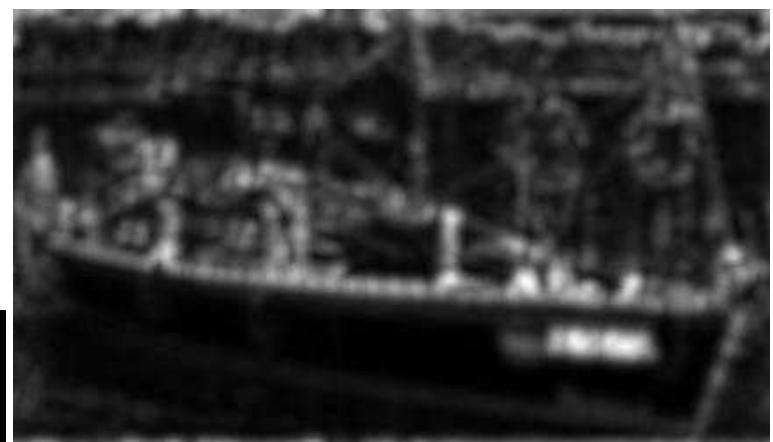
$n=4$



$n=6$



$n=8$

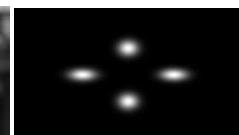


$n=16$



## Wielkość płatków

- podobny efekt „zapełniania” obrazu detalami można osiągnąć powiększając „płatki” – parametry  $a$  i  $b$
- powiększając parametr  $\mathbf{b}$  pokrywa się też sąsiednie kierunki



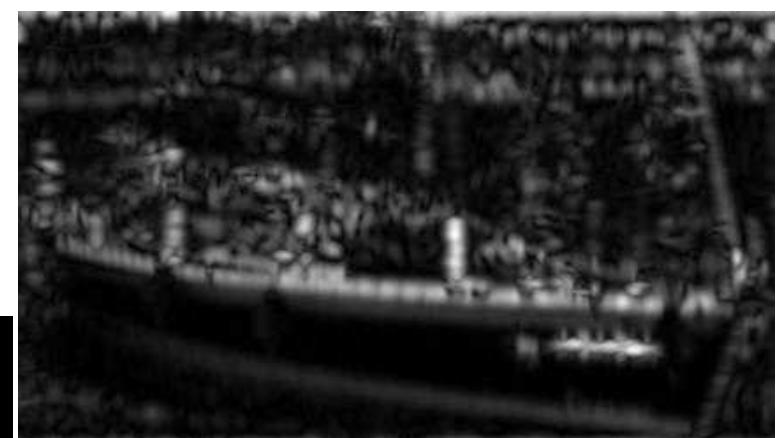
$b=0.075$



$b=0.10$



$b=0.15$



$b=0.25$



## Wielkość płatków

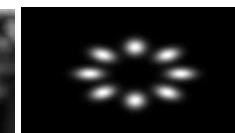
- zmniejszając parametr  $a$  zwiększa się też kopertę Gausa maski - filtr ma większy zasięg – z większej odległości „widzi” krawędź
- działa to też w drugą stronę – większy płatek  $\Rightarrow$  mniejsza koperta  $\Rightarrow$  precyzyjniejsza lokalizacja krawędzi



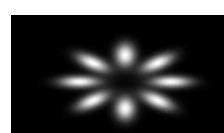
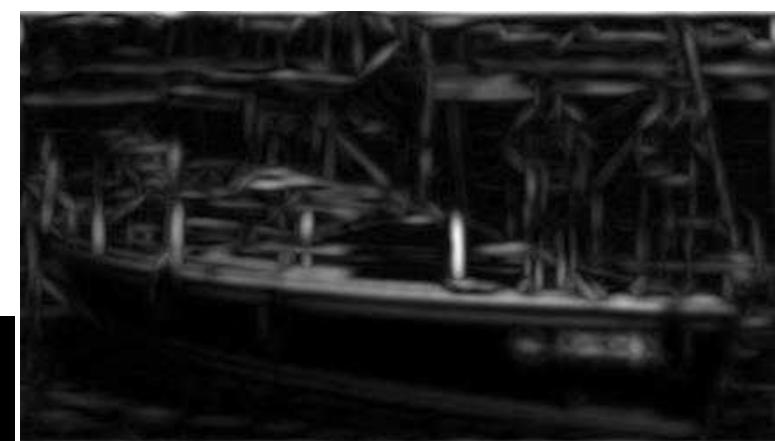
$a=0.03$



$a=0.066$



$a=0.1$

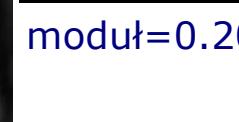
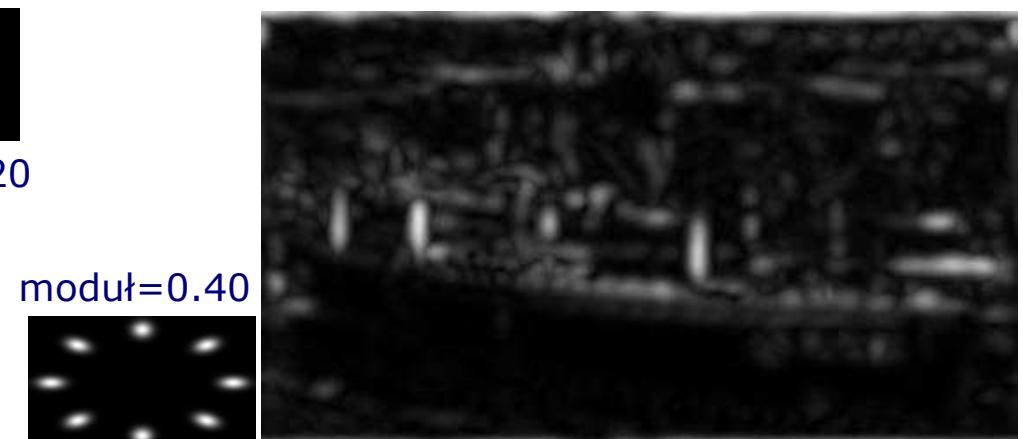
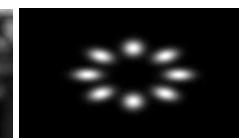


$a=0.15$



## Odległość płatków od środka

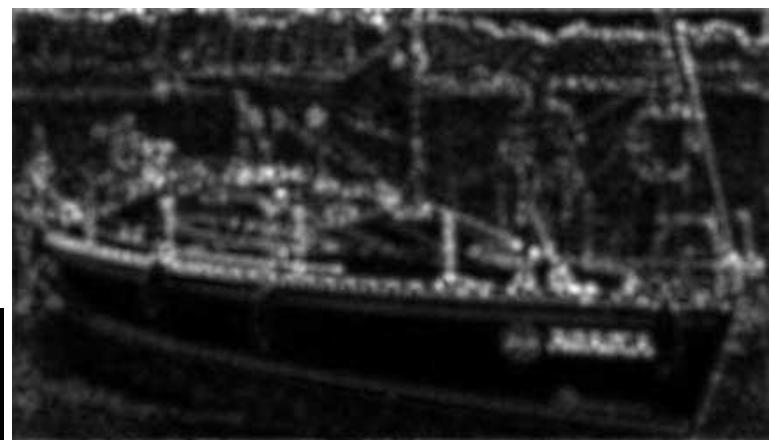
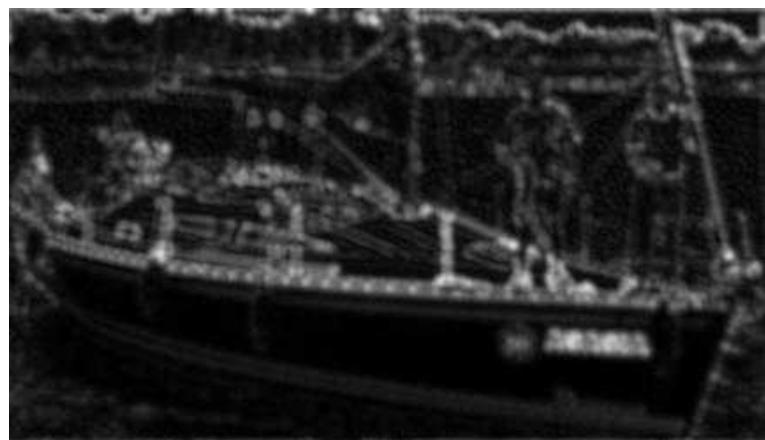
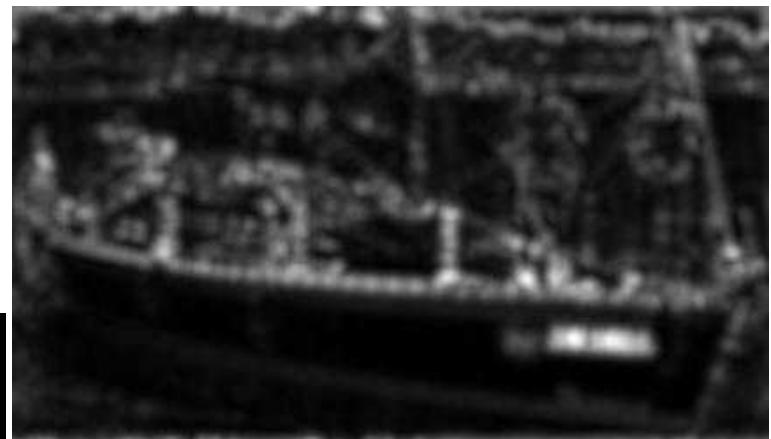
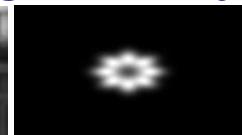
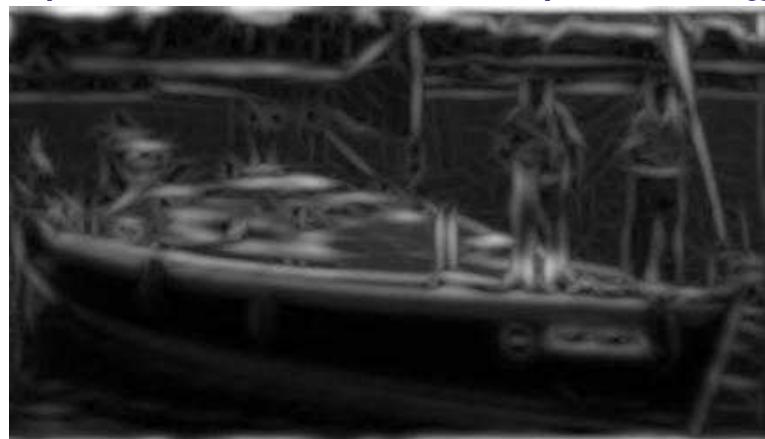
- rozsuwając „płatki” od środka. interesujemy się wyższymi częstotliwościami
- przesuwając je do środka, skupiamy się na mniejszych częstotliwościach
- jeśli płatki pokrywają częstotliwość 0, pojawia się składowa stała – obserwujemy pojawienie się oryginalnego obrazu (efekt uśredniania)





## Filtr kołowy

- filtr, który pokrywa wszystkie kierunki danej częstotliwości
- filtr typu *band-pass*
- powinien dawać w wyniku ciągłe krawędzie



# GNU Free Documentation License



Version 1.2, November 2002

Copyright (C) 2000,2001,2002 Free Software Foundation, Inc.  
51 Franklin St, Fifth Floor, Boston, MA 02110-1301 USA  
Everyone is permitted to copy and distribute verbatim copies  
of this license document, but changing it is not allowed.

## 0. PREAMBLE

The purpose of this License is to make a manual, textbook, or other functional and useful document "free" in the sense of freedom: to assure everyone the effective freedom to copy and redistribute it, with or without modifying it, either commercially or noncommercially. Secondly, this License preserves for the author and publisher a way to get credit for their work, while not being considered responsible for modifications made by others.

This License is a kind of "copyleft", which means that derivative works of the document must themselves be free in the same sense. It complements the GNU General Public License, which is a copyleft license designed for free software.

We have designed this License in order to use it for manuals for free software, because free software needs free documentation: a free program should come with manuals providing the same freedoms that the software does. But this License is not limited to software manuals; it can be used for any textual work, regardless of subject matter or whether it is published as a printed book. We recommend this License principally for works whose purpose is instruction or reference.

## 1. APPLICABILITY AND DEFINITIONS

This License applies to any manual or other work, in any medium, that contains a notice placed by the copyright holder saying it can be distributed under the terms of this License. Such a notice grants a world-wide, royalty-free license, unlimited in duration, to use that work under the conditions stated herein. The "Document", below, refers to any such manual or work. Any member of the public is a licensee, and is addressed as "you". You accept the license if you copy, modify or distribute the work in a way requiring permission under copyright law.

A "Modified Version" of the Document means any work containing the Document or a portion of it, either copied verbatim, or with modifications and/or translated into another language.

A "Secondary Section" is a named appendix or a front-matter section of the Document that deals exclusively with the relationship of the publishers or authors of the Document to the Document's overall subject (or to related matters) and contains nothing that could fall directly within that overall subject. (Thus, if the Document is in part a textbook of mathematics, a Secondary Section may not explain any mathematics.) The relationship could be a matter of historical connection with the subject or with related matters, or of legal, commercial, philosophical, ethical or political position regarding them.

The "Invariant Sections" are certain Secondary Sections whose titles are designated, as being those of Invariant Sections, in the notice that says that the Document is released under this License. If a section does not fit the above definition of Secondary then it is not allowed to be designated as Invariant. The Document may contain zero Invariant Sections. If the Document does not identify any Invariant Sections then there are none.

The "Cover Texts" are certain short passages of text that are listed, as Front-Cover Texts or Back-Cover Texts, in the notice that says that the Document is released under this License. A Front-Cover Text may be at most 5 words, and a Back-Cover Text may be at most 25 words.

A "Transparent" copy of the Document means a machine-readable copy, represented in a format whose specification is available to the general public, that is suitable for revising the document straightforwardly with generic text editors or (for images composed of pixels) generic paint programs or (for drawings) some widely available drawing editor, and that is suitable for input to text formatters or for automatic translation to a variety of formats suitable for input to text formatters. A copy made in an otherwise Transparent file format whose markup, or absence of markup, has been arranged to thwart or discourage subsequent modification by readers is not Transparent. An image format is not Transparent if used for any substantial amount of text. A copy that is not "Transparent" is called "Opaque".

Examples of suitable formats for Transparent copies include plain ASCII without markup, Texinfo input format, LaTeX input format, SGML or XML using a publicly available DTD, and standard-conforming simple HTML, PostScript or PDF designed for human modification. Examples of transparent image formats include PNG, XCF and JPG. Opaque formats include proprietary formats that can be read and edited only by proprietary word processors, SGML or XML for which the DTD and/or processing tools are not generally available, and the machine-generated HTML, PostScript or PDF produced by some word processors for output purposes only.

The "Title Page" means, for a printed book, the title page itself, plus such following pages as are needed to hold, legibly, the material this License requires to appear in the title page. For works in formats which do not have any title page as such, "Title Page" means the text near the most prominent appearance of the work's title, preceding the beginning of the body of the text.

A section "Entitled XYZ" means a named subunit of the Document whose title either is precisely XYZ or contains XYZ in parentheses following text that translates XYZ in another language. (Here XYZ stands for a specific section name mentioned below, such as "Acknowledgements", "Dedications", "Endorsements", or "History".) To "Preserve the Title" of such a section when you modify the Document means that it remains a section "Entitled XYZ" according to this definition.

The Document may include Warranty Disclaimers next to the notice which states that this License applies to the Document. These Warranty Disclaimers are considered to be included by reference in this License, but only as regards disclaiming warranties: any other implication that these Warranty Disclaimers may have is void and has no effect on the meaning of this License.

## 2. VERBATIM COPYING

You may copy and distribute the Document in any medium, either commercially or noncommercially, provided that this License, the copyright notices, and the license notice saying this License applies to the Document are reproduced in all copies, and that you add no other conditions whatsoever to those of this License. You may not use technical measures to obstruct or control the reading or further copying of the copies you make or distribute. However, you may accept compensation in exchange for copies. If you distribute a large enough number of copies you must also follow the conditions in section 3.

You may also lend copies, under the same conditions stated above, and you may publicly display copies.



### 3. COPYING IN QUANTITY

If you publish printed copies (or copies in media that commonly have printed covers) of the Document, numbering more than 100, and the Document's license notice requires Cover Texts, you must enclose the copies in covers that carry, clearly and legibly, all these Cover Texts: Front-Cover Texts on the front cover, and Back-Cover Texts on the back cover. Both covers must also clearly and legibly identify you as the publisher of these copies. The front cover must present the full title with all words of the title equally prominent and visible. You may add other material on the covers in addition. Copying with changes limited to the covers, as long as they preserve the title of the Document and satisfy these conditions, can be treated as verbatim copying in other respects.

If the required texts for either cover are too voluminous to fit legibly, you should put the first ones listed (as many as fit reasonably) on the actual cover, and continue the rest onto adjacent pages.

If you publish or distribute Opaque copies of the Document numbering more than 100, you must either include a machine-readable Transparent copy along with each Opaque copy, or state in or with each Opaque copy a computer-network location from which the general network-using public has access to download using public-standard network protocols a complete Transparent copy of the Document, free of added material. If you use the latter option, you must take reasonably prudent steps, when you begin distribution of Opaque copies in quantity, to ensure that this Transparent copy will remain thus accessible at the stated location until at least one year after the last time you distribute an Opaque copy (directly or through your agents or retailers) of that edition to the public.

It is requested, but not required, that you contact the authors of the Document well before redistributing any large number of copies, to give them a chance to provide you with an updated version of the Document.

### 4. MODIFICATIONS

You may copy and distribute a Modified Version of the Document under the conditions of sections 2 and 3 above, provided that you release the Modified Version under precisely this License, with the Modified Version filling the role of the Document, thus licensing distribution and modification of the Modified Version to whoever possesses a copy of it. In addition, you must do these things in the Modified Version:

- \* A. Use in the Title Page (and on the covers, if any) a title distinct from that of the Document, and from those of previous versions (which should, if there were any, be listed in the History section of the Document). You may use the same title as a previous version if the original publisher of that version gives permission.
- \* B. List on the Title Page, as authors, one or more persons or entities responsible for authorship of the modifications in the Modified Version, together with at least five of the principal authors of the Document (all of its principal authors, if it has fewer than five), unless they release you from this requirement.
- \* C. State on the Title page the name of the publisher of the Modified Version, as the publisher.
- \* D. Preserve all the copyright notices of the Document.
- \* E. Add an appropriate copyright notice for your modifications adjacent to the other copyright notices.
- \* F. Include, immediately after the copyright notices, a license notice giving the public permission to use the Modified Version under the terms of this License, in the form shown in the Addendum below.
- \* G. Preserve in that license notice the full lists of Invariant Sections and required Cover Texts given in the Document's license notice.
- \* H. Include an unaltered copy of this License.
- \* I. Preserve the section Entitled "History", Preserve its Title, and add to it an item stating at least the title, year, new authors, and publisher of the Modified Version as given on the Title Page. If there is no section Entitled "History" in the Document, create one stating the title, year, authors, and publisher of the Document as given on its Title Page, then add an item describing the Modified Version as stated in the previous sentence.
- \* J. Preserve the network location, if any, given in the Document for public access to a Transparent copy of the Document, and likewise the network locations given in the Document for previous versions it was based on. These may be placed in the "History" section. You may omit a network location for a work that was published at least four years before the Document itself, or if the original publisher of the version it refers to gives permission.
- \* K. For any section Entitled "Acknowledgements" or "Dedications", Preserve the Title of the section, and preserve in the section all the substance and tone of each of the contributor acknowledgements and/or dedications given therein.
- \* L. Preserve all the Invariant Sections of the Document, unaltered in their text and in their titles. Section numbers or the equivalent are not considered part of the section titles.
- \* M. Delete any section Entitled "Endorsements". Such a section may not be included in the Modified Version.
- \* N. Do not retitle any existing section to be Entitled "Endorsements" or to conflict in title with any Invariant Section.
- \* O. Preserve any Warranty Disclaimers.

If the Modified Version includes new front-matter sections or appendices that qualify as Secondary Sections and contain no material copied from the Document, you may at your option designate some or all of these sections as invariant. To do this, add their titles to the list of Invariant Sections in the Modified Version's license notice. These titles must be distinct from any other section titles.

You may add a section Entitled "Endorsements", provided it contains nothing but endorsements of your Modified Version by various parties--for example, statements of peer review or that the text has been approved by an organization as the authoritative definition of a standard.

You may add a passage of up to five words as a Front-Cover Text, and a passage of up to 25 words as a Back-Cover Text, to the end of the list of Cover Texts in the Modified Version. Only one passage of Front-Cover Text and one of Back-Cover Text may be added by (or through arrangements made by) any one entity. If the Document already includes a cover text for the same cover, previously added by you or by arrangement made by the same entity you are acting on behalf of, you may not add another; but you may replace the old one, on explicit permission from the previous publisher that added the old one.

The author(s) and publisher(s) of the Document do not by this License give permission to use their names for publicity for or to assert or imply endorsement of any Modified Version.



## 6. COLLECTIONS OF DOCUMENTS

You may make a collection consisting of the Document and other documents released under this License, and replace the individual copies of this License in the various documents with a single copy that is included in the collection, provided that you follow the rules of this License for verbatim copying of each of the documents in all other respects.

You may extract a single document from such a collection, and distribute it individually under this License, provided you insert a copy of this License into the extracted document, and follow this License in all other respects regarding verbatim copying of that document.

## 7. AGGREGATION WITH INDEPENDENT WORKS

A compilation of the Document or its derivatives with other separate and independent documents or works, in or on a volume of a storage or distribution medium, is called an "aggregate" if the copyright resulting from the compilation is not used to limit the legal rights of the compilation's users beyond what the individual works permit. When the Document is included in an aggregate, this License does not apply to the other works in the aggregate which are not themselves derivative works of the Document.

If the Cover Text requirement of section 3 is applicable to these copies of the Document, then if the Document is less than one half of the entire aggregate, the Document's Cover Texts may be placed on covers that bracket the Document within the aggregate, or the electronic equivalent of covers if the Document is in electronic form. Otherwise they must appear on printed covers that bracket the whole aggregate.

## 8. TRANSLATION

Translation is considered a kind of modification, so you may distribute translations of the Document under the terms of section 4. Replacing Invariant Sections with translations requires special permission from their copyright holders, but you may include translations of some or all Invariant Sections in addition to the original versions of these Invariant Sections. You may include a translation of this License, and all the license notices in the Document, and any Warranty Disclaimers, provided that you also include the original English version of this License and the original versions of those notices and disclaimers. In case of a disagreement between the translation and the original version of this License or a notice or disclaimer, the original version will prevail.

If a section in the Document is Entitled "Acknowledgements", "Dedications", or "History", the requirement (section 4) to Preserve its Title (section 1) will typically require changing the actual title.

## 9. TERMINATION

You may not copy, modify, sublicense, or distribute the Document except as expressly provided for under this License. Any other attempt to copy, modify, sublicense or distribute the Document is void, and will automatically terminate your rights under this License. However, parties who have received copies, or rights, from you under this License will not have their licenses terminated so long as such parties remain in full compliance.

## 10. FUTURE REVISIONS OF THIS LICENSE

The Free Software Foundation may publish new, revised versions of the GNU Free Documentation License from time to time. Such new versions will be similar in spirit to the present version, but may differ in detail to address new problems or concerns. See <http://www.gnu.org/copyleft/>.

Each version of the License is given a distinguishing version number. If the Document specifies that a particular numbered version of this License "or any later version" applies to it, you have the option of following the terms and conditions either of that specified version or of any later version that has been published (not as a draft) by the Free Software Foundation. If the Document does not specify a version number of this License, you may choose any version ever published (not as a draft) by the Free Software Foundation.

## 5. COMBINING DOCUMENTS

You may combine the Document with other documents released under this License, under the terms defined in section 4 above for modified versions, provided that you include in the combination all of the Invariant Sections of all of the original documents, unmodified, and list them all as Invariant Sections of your combined work in its license notice, and that you preserve all their Warranty Disclaimers.

The combined work need only contain one copy of this License, and multiple identical Invariant Sections may be replaced with a single copy. If there are multiple Invariant Sections with the same name but different contents, make the title of each such section unique by adding at the end of it, in parentheses, the name of the original author or publisher of that section if known, or else a unique number. Make the same adjustment to the section titles in the list of Invariant Sections in the license notice of the combined work.

In the combination, you must combine any sections Entitled "History" in the various original documents, forming one section Entitled "History"; likewise combine any sections Entitled "Acknowledgements", and any sections Entitled "Dedications". You must delete all sections Entitled "Endorsements."

(not as a draft) by the Free Software Foundation.

# Cyfrowe Przetwarzanie Obrazów

## Digital Image Processing



Aurelio A. Heckert

Copyright (c) 2008 Cezary Bołdak

Permission is granted to copy, distribute and/or modify this document under the terms of the GNU Free Documentation License, Version 1.2 or any later version published by the Free Software Foundation; with no Invariant Sections, no Front-Cover Texts, and no Back-Cover Texts. A copy of the license is included in the section entitled "GNU Free Documentation License"

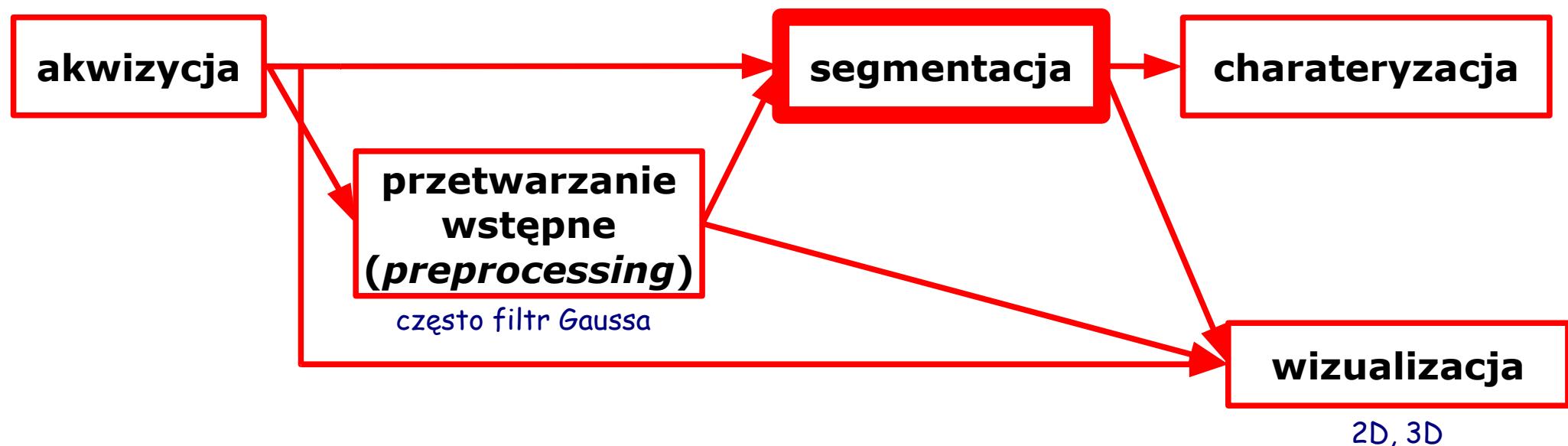


## Wykład 7

# Segmentacja



## Cykl przetwarzania obrazów cyfrowych

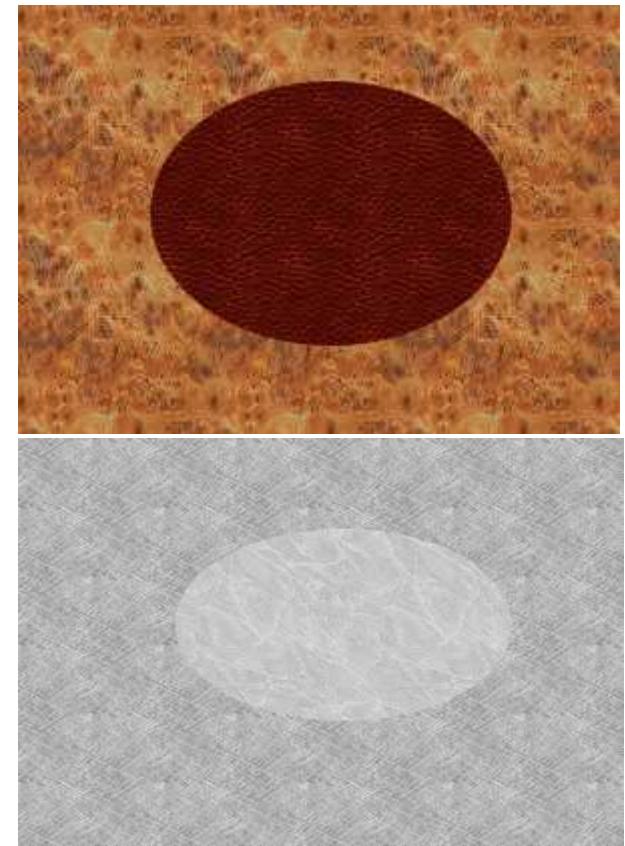
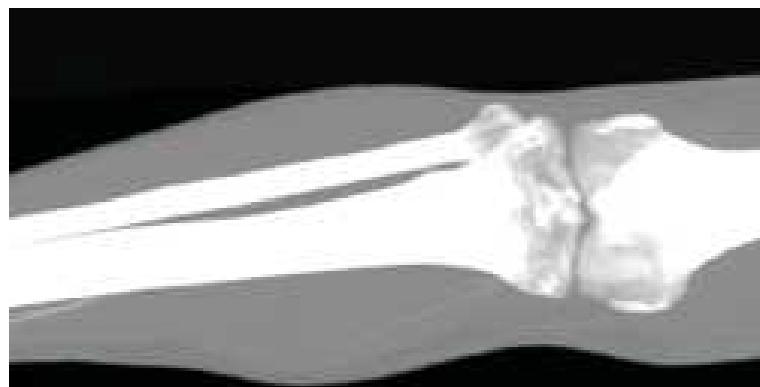




**Segmentacja** – podział obrazu na **spójne** fragmenty, różniące się w pewien sposób między sobą;  
czasami także: wyodrębnienie interesujących obiektów z tła

Co to znaczy spójne ? :

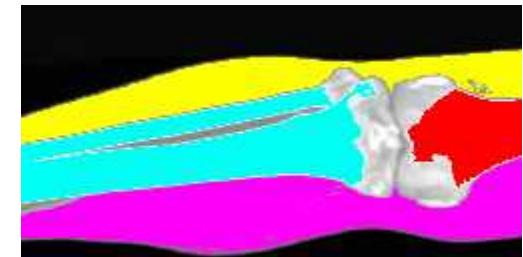
- o tym samym kolorze, jasności
- o podobnej teksturze
- bez wyraźnej granicy
- kryterium czasem trudne do określenia (subiektywne)





## Segmentacja

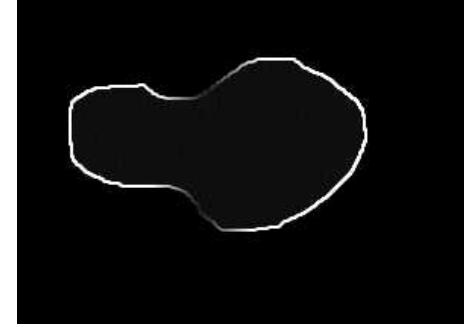
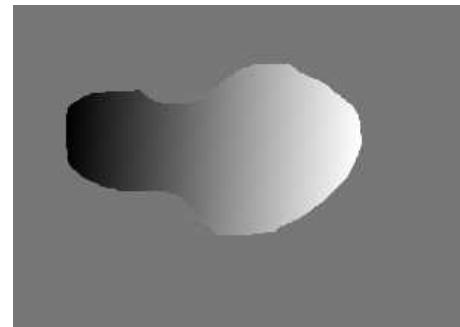
- często podział na 2 obszary – **tło i obiekt(y)**
- nie ma jednej metody segmentacji – określony jest tylko cel, sposobów jest wiele: konkurują ze sobą lub uzupełniają się
- metody uniwersalne i specjalizowane (często wykorzystujące wiedzę, oparte na modelu, np. struktury liniowe)
- dwuwymiarowe i trójwymiarowe
- automatyczne, półautomatyczne, (ręczne)  
nie wszędzie metody automatyczne są akceptowane (medycyna)
- metody często wieloetapowe, hybrydowe
- metody samouczące się
- metody globalne i lokalne
- **wynik:** obraz gdzie każdy piksel ma przyporządkowaną etykietę obszaru





## Segmentacja obszarowa i krawędziowa

- można wyróżnić 2 główne grupy metod segmentacji:
  - oparte na podobieństwach wewnętrz **obszarów**  
wynikiem jest zbiór pikseli które (lokalnie) nie różnią się między sobą
  - oparte na **granicach** między obszarami  
wynikiem jest zbiór krawędzi, w poprzek których piksele bardzo się różnią  
(detektory krawędzi)
- wyniki metod z obu grup zwykle są zbliżone, ale nie zawsze



- rodzina metod, która wydaje się umykać temu podziałowi – **progowanie**  
(metoda bezkontekstowa, nie liczy się tu otoczenie pikseli): czy na pewno ?



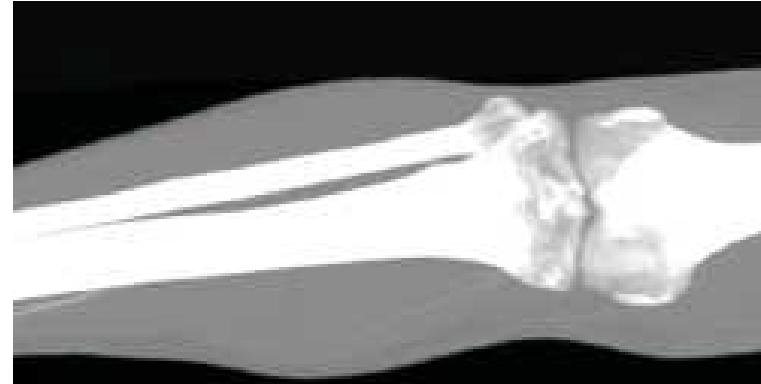
## Segmentacja przez progowanie

$$\begin{aligned} I(x,y) \geq T &\Rightarrow (x,y) \in \text{"obiekt"} \\ I(x,y) < T &\Rightarrow (x,y) \in \text{"tło"} \end{aligned}$$

- brana pod uwagę jest tylko intensywność piksela
- warunek konieczny dla zastosowania progowania:  
piksele obiektu muszą mieć unikalny zakres intensywności  
(np. obiekt jaśniejszy lub ciemniejszy niż tło)
- nawet gdy jest to zapewnione, pozostaje ważny problem:  
wybór progu **T**
  - wybór automatyczny
  - wybór ręczny (interaktywny)
- ten sam algorytm z różnymi progami wyodrębnia różne struktury



## Progowanie z pojedynczym progiem



$T = 67$   
cała nogą



$T = 176$   
struktura kości



$T = 230$   
kość bez chrząstki





## Progowanie z podwójnym progiem

$$T_1 \leq I(x,y) \leq T_2 \Rightarrow (x,y) \in \text{"obiekt"}$$

$$I(x,y) > T_1 \text{ lub } I(x,y) < T_2 \Rightarrow (x,y) \in \text{"tło"}$$

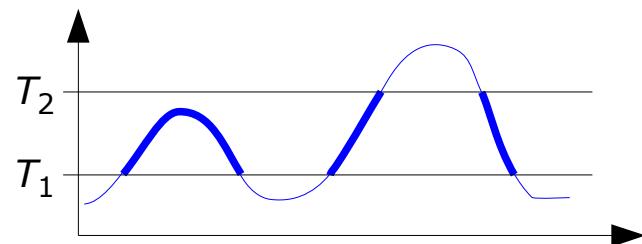
$$T_1 = 67, T_2 = 176$$

noga bez kości

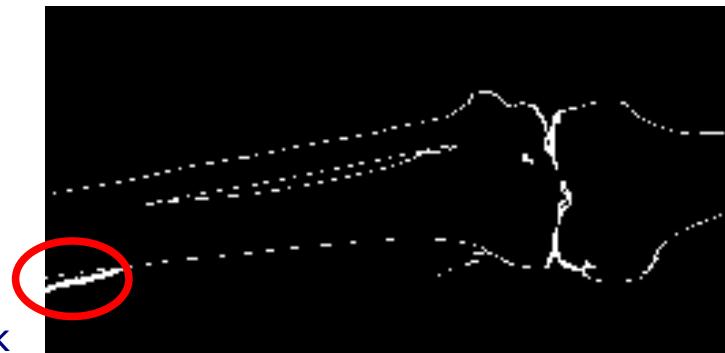


$$T_1 = 150, T_2 = 176$$

naczynie krwionośne



strefa przejściowa struktur jaśniejszych zakłóca wynik





## Progowanie obrazów barwnych

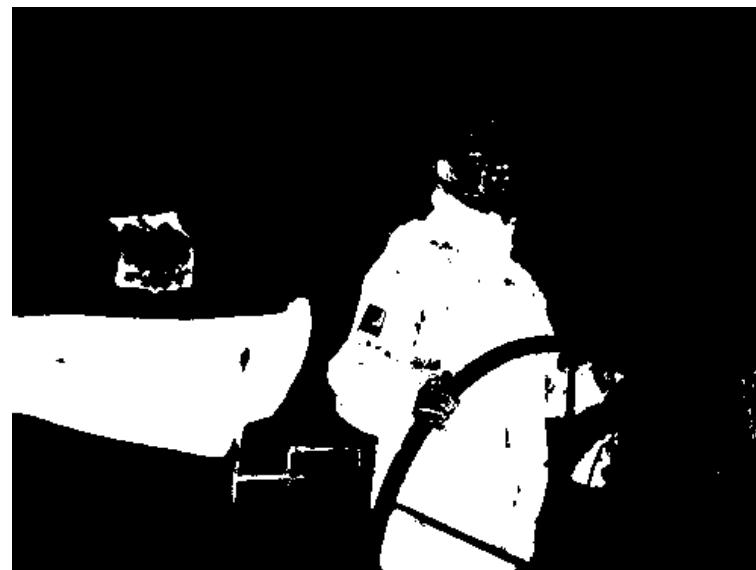
- aby wybrać jeden kolor, lepiej od modelu RGB sprawdza się HSV

progowanie kanału barwy (*Hue*)



progowanie kanału barwy (*Hue*)  
i nasycenia (Saturation)  
(iloczyn obrazów wynikowych)

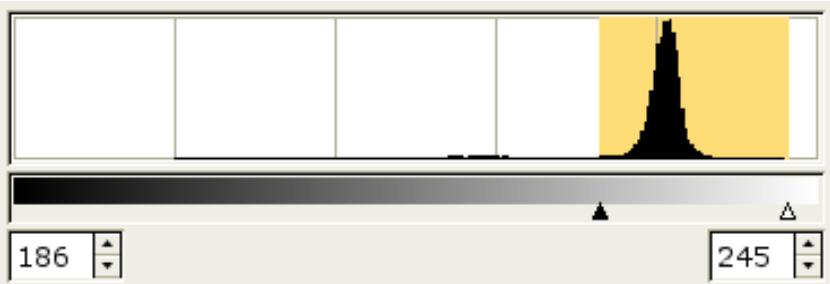
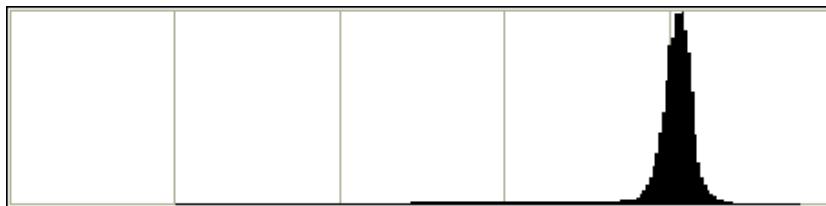
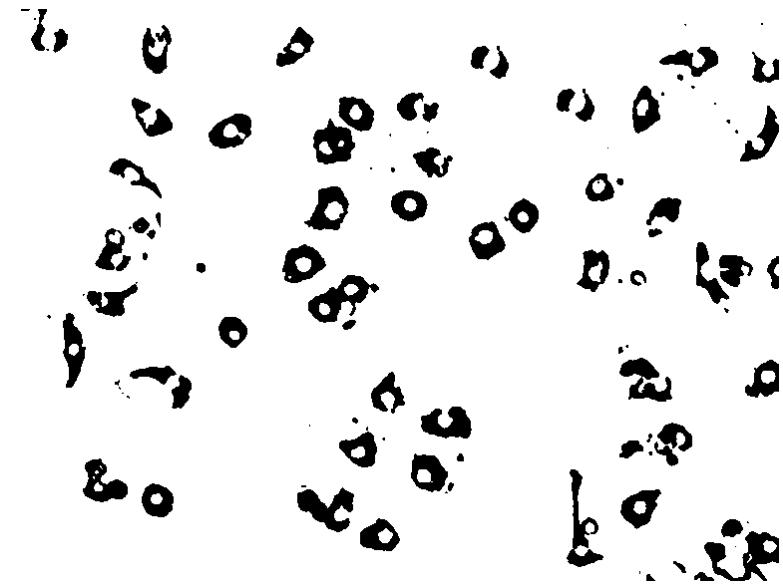
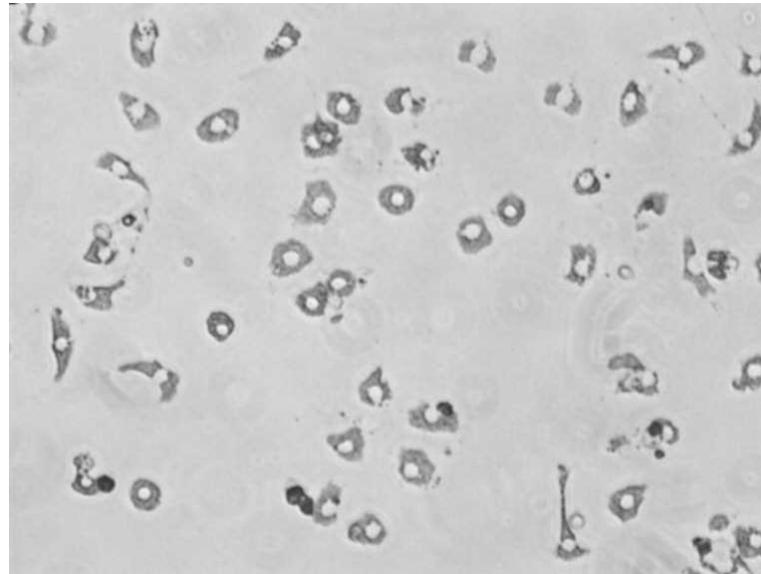
- aby lepiej wydobyć barwę można progować więcej kanałów jednocześnie





## Wyznaczanie progu na podstawie histogramu

- często histogram obrazu wykazuje maksima w częstotliwości pewnych intensywności
- granice tych maksymów mogą służyć za progi:

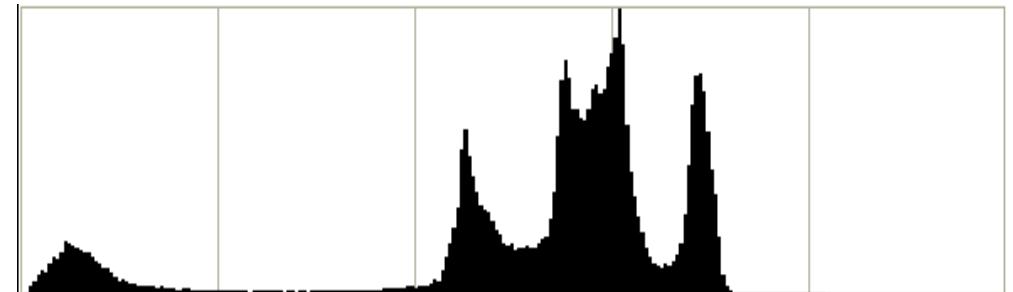
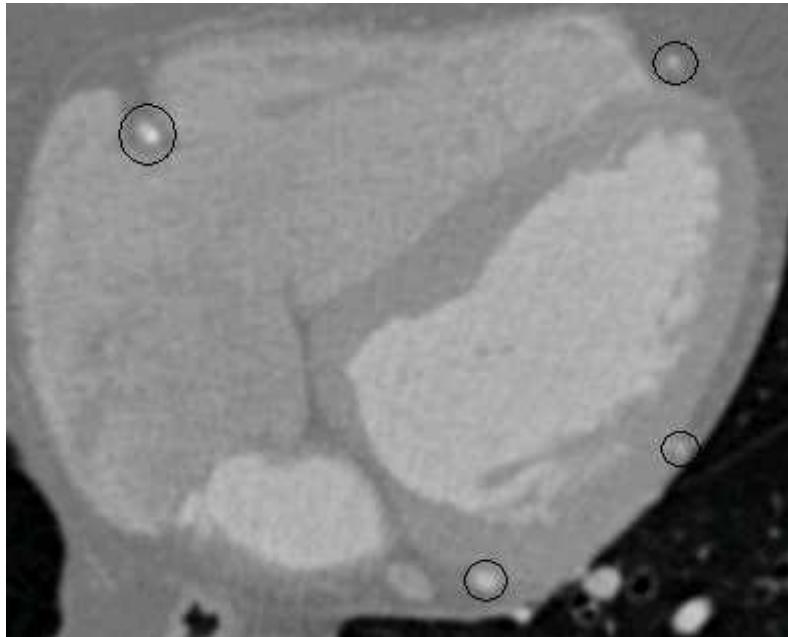


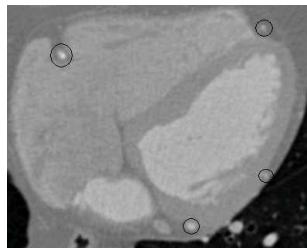
progowanie interaktywne (GIMP)



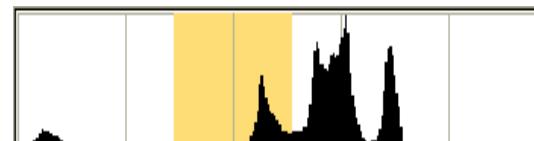
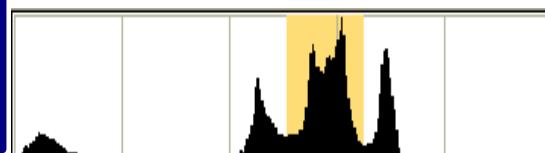
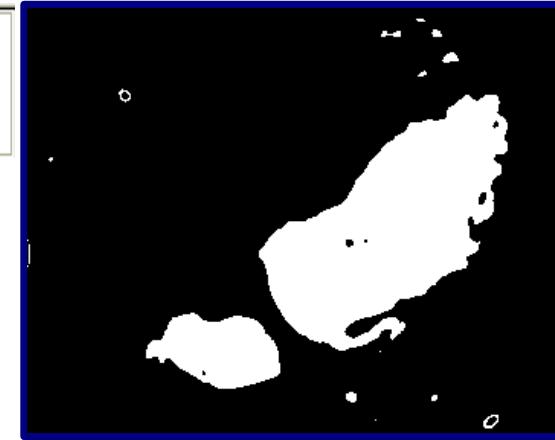
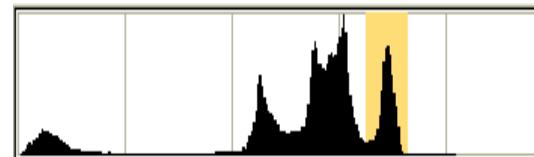
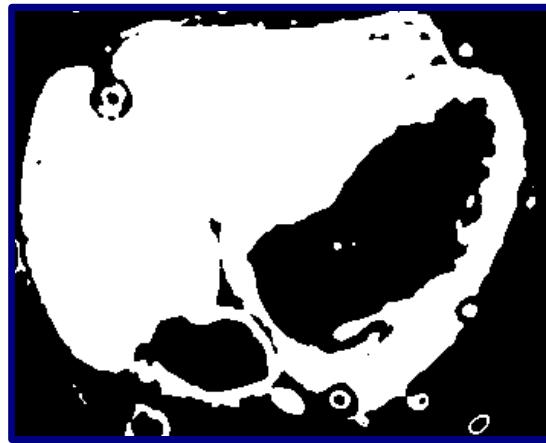
## Obraz z wieloma obiekty

- obiekty – różne tkanki na tomografii serca (przekrój poprzeczny)
- wyraźnie widać maksima częstości na histogramie
- możliwe jest interaktywne wyodrębnienie tych tkanek





## Wyznaczanie progów na podstawie histogramu



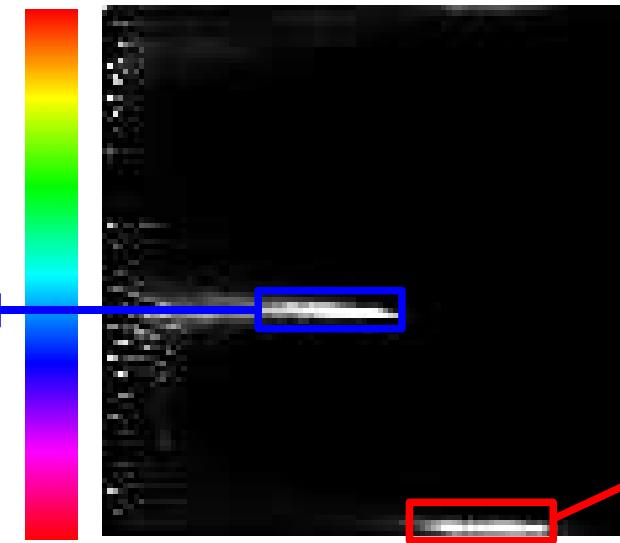


## Wykorzystanie histogramów 2D do progowania

- progowanie w przestrzeni HS (barwa, nasycenie)
- histogram dwuwymiarowy wykazuje maksima częstości
- ręczne ich wyznaczenie umożliwia wyróżnienie obiektów o tych samych barwach
- łatwe zautomatyzowanie
- można też wykorzystać histogramy 3D lub więcej (np. zdjęcia satelitarne w różnych częstotliwościach)



$0,54 < \text{Hue} < 0,6$   
 $0,25 < \text{Saturation} < 0,6$



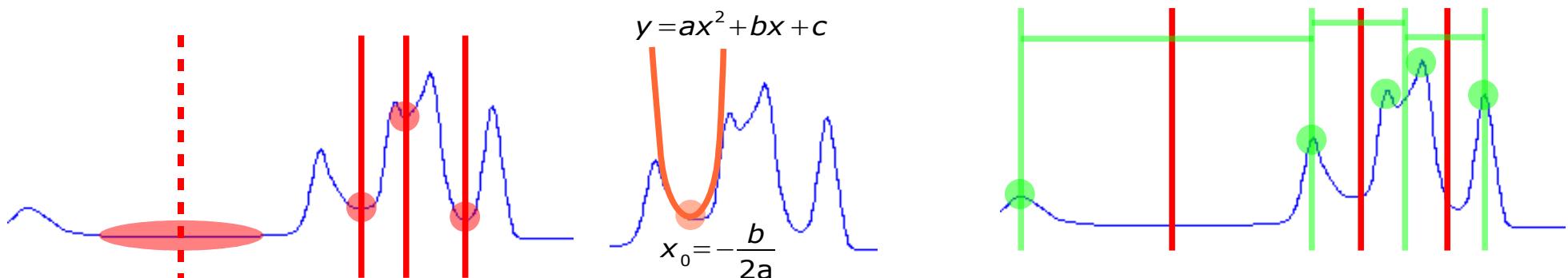
$0,95 < \text{Hue}$   
 $0,55 < \text{Saturation} < 0,85$



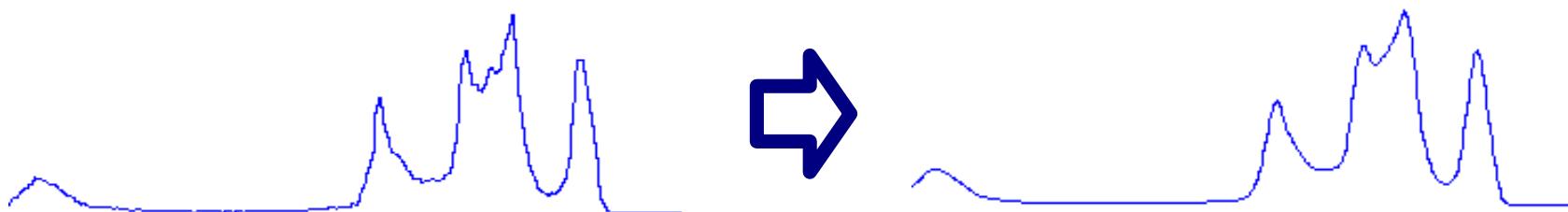
## Automatyczne wyznaczanie progów z histogramu

### Podejście geometryczne:

- wyznaczenie środków **dolin**: minimów lokalnych histogramu;
  - czasami płaskie doliny utrudniają zadanie
  - środki dolin można przybliżyć elipsami (minimalizacja błędu kwadratowego)
- wyznaczenie **szczytów**: maksimów lokalnych histogramu i przyjęcie jako progów geometrycznych środków między nimi
  - zbyt bliskie maksima można zastąpić jednym



- aby nie wpadać w ekstrema wynikające z szumu należy wygładzić histogram:





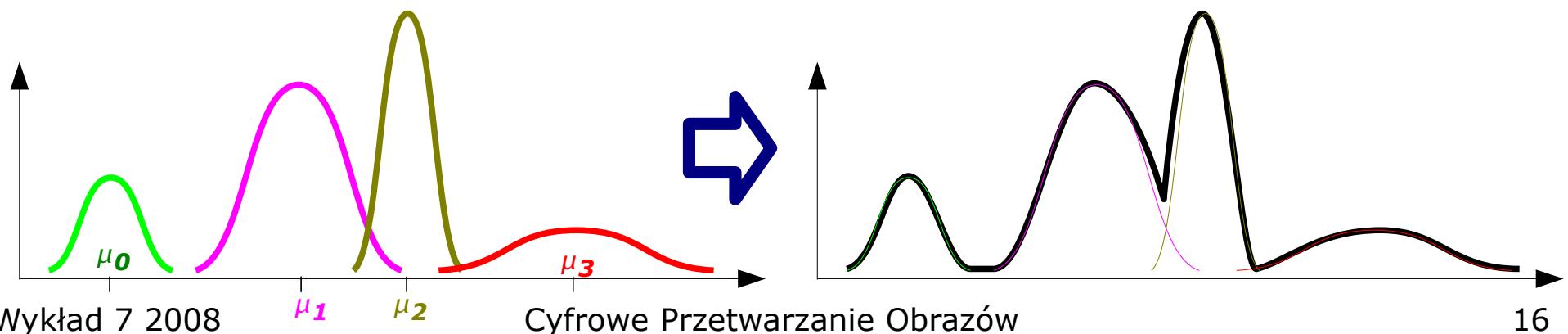
## Automatyczne wyznaczanie progów z histogramu

### Podejście zakładające określony model rozkładu intensywności pikseli obrazu

- zakłada się tu, że każdy z obiektów  $O_i$ , występujących w obrazie ma intensywność np. o rozkładzie normalnym  $N(\mu_i, \sigma_i)$  gdzie:  $\mu_i$  – reprezentuje średnią intensywność obiektu  
 $\sigma_i$  – odchylenie standardowe
- łączny rozkład prawdopodobieństwa intensywności pikseli obrazu jest sumą prawdopodobieństw warunkowych ważonych prawdopodobieństwem wystąpienia danego obiektu:

$$P(\text{Image}(x, y) = I) = \sum_i P(O_i) \cdot P(\text{Image}(x, y) = I | (x, y) \in O_i)$$

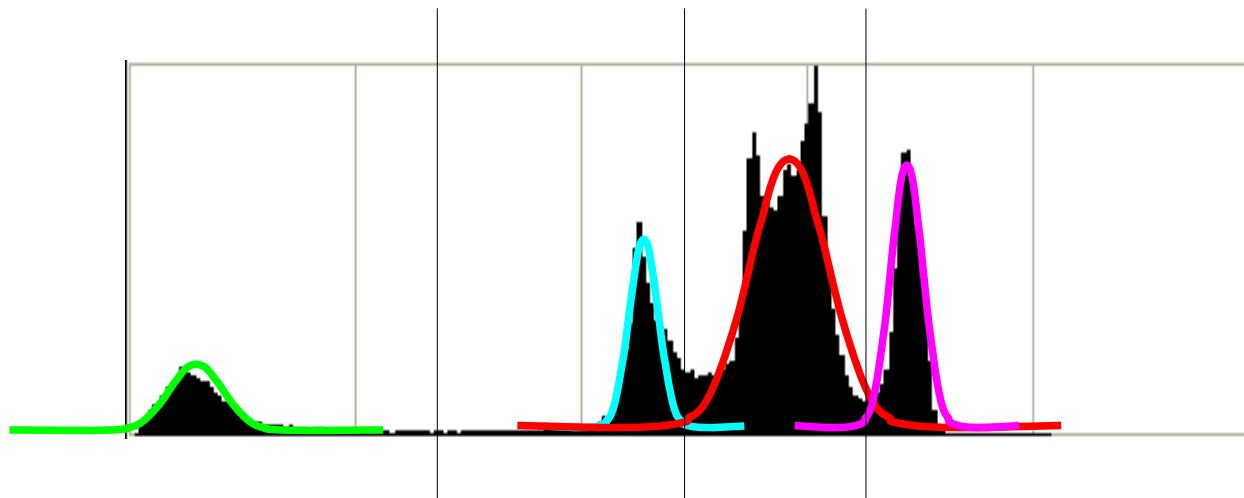
- zakładając, że każde z prawdopodobieństw warunkowych ma rozkład normalny a prawdopodobieństwo wystąpienia obiektu  $O_i$  pełni rolę czynnika skalującego mamy:





## Automatyczne wyznaczanie progów z histogramu

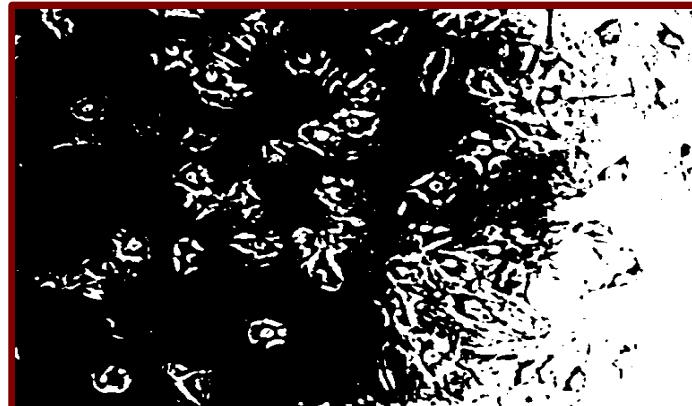
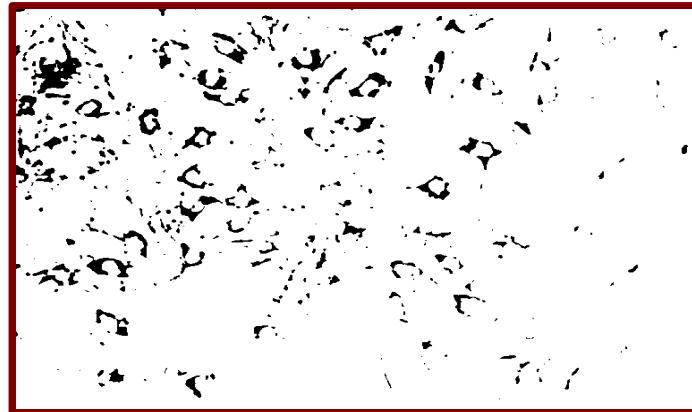
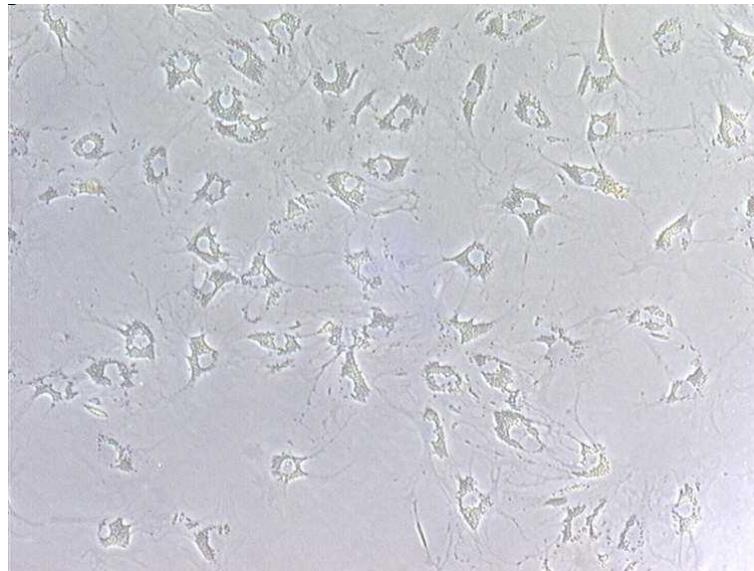
- mając histogram obrazu i informację o liczbie obiektów można aproksymować parametry rozkładów
- estymacja parametrów rozkładów można dokonać używając jednej z istniejących metod, np. algorytmu EM (expectation-maximization)
- mając parametry rozkładów można przyjąć wartości progów (np. wykorzystując regułę klasyfikacyjną Bayes'a)





## Progi lokalne

- czasami jasności obiektów i tła zmieniają w obrazie (np. wskutek nierównomiernego oświetlenia)
- nie można wtedy wyznaczyć jednego, globalnego zestawu progów
- progi muszą być lokalne, wyliczone dla fragmentów obrazu
- procedura interaktywna staje się uciążliwa – potrzeba automatyzacji



użycie  
progów  
globalnych

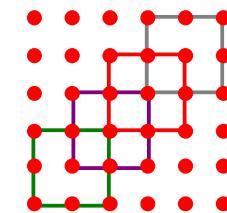


## Progowanie adaptacyjne

- obraz dzieli się na **obszary** (rozmiar dopasowany do rozmiaru obiektów, dynamiki zmian otoczenia)

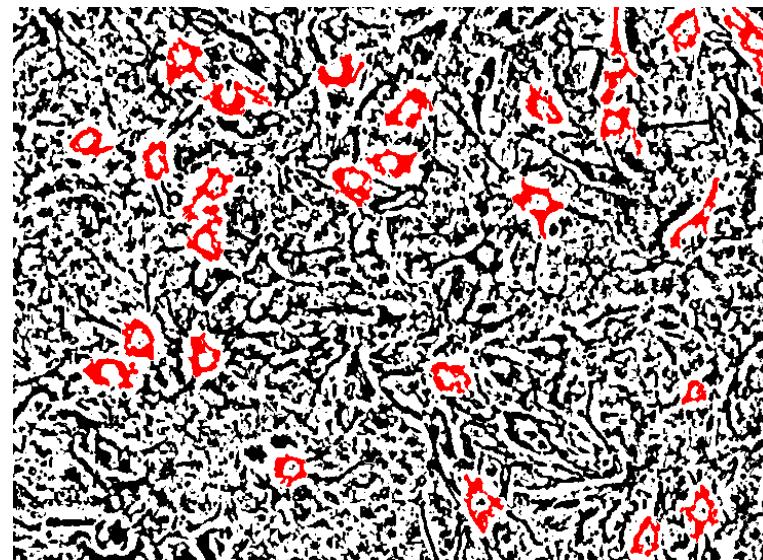
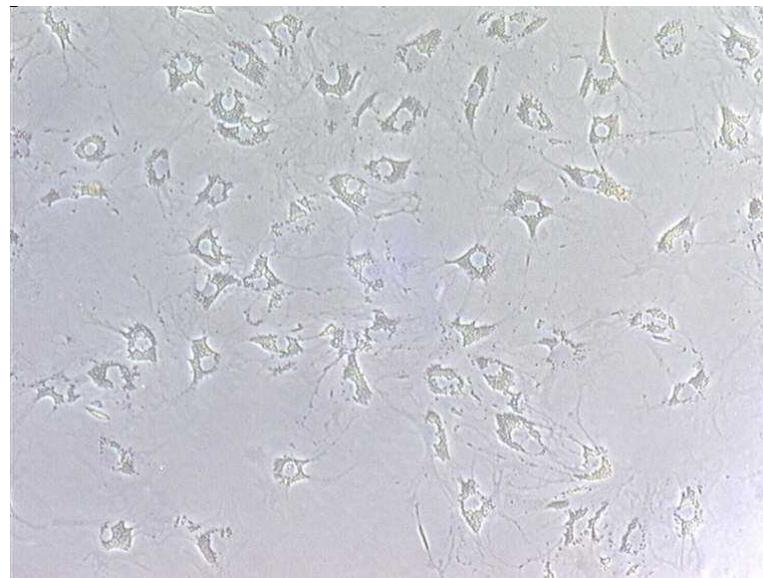
•	•	•	•	•
•	•	•	•	•
•	•	•	•	•

- w każdym fragmencie liczy się lokalne charakterystyki:
  - średnią wartość intensywności
  - minimalną i maksymalną wartość intensywności
- na tej podstawie oblicza się **próg lokalny** dla tego fragmentu
  - średnią wartość intensywności
  - średnią wartości minimalnej i maksymalnej
- aby uniknąć skokowej zmiany wartości lokalnego progu na granicach obszarów:
  - liczy się lokalne proggi dla **każdego piksela** obrazu wykorzystując jego lokalne otoczenie (długo czas obliczeń)
  - przyjmuje się obliczone wartości progów dla środków **rozłącznych obszarów** i dla pozostałych pikseli wartości progów interpoluje się



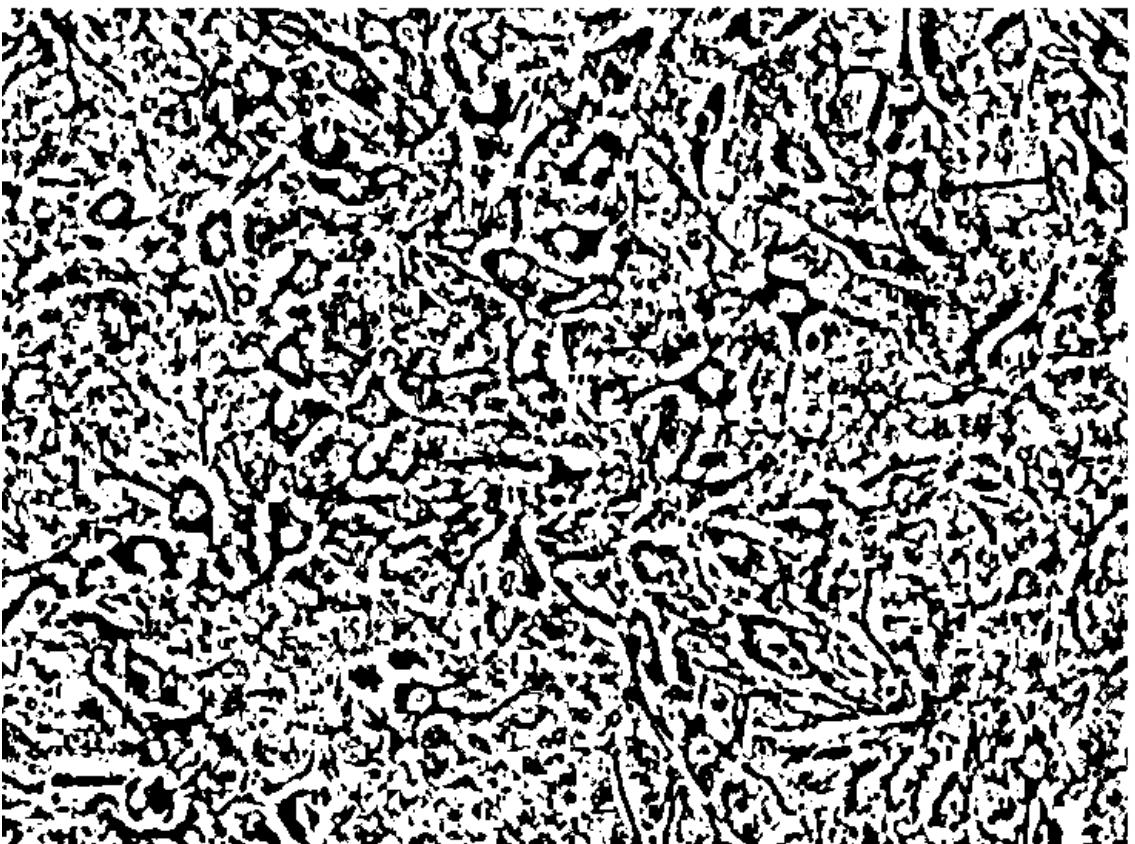


## Progowanie adaptacyjne



niektóre komórki ręcznie wyróżnione

- lokalne progi jako średnie
- progowanie poprawnie wykrywa krawędzie komórek
- zawodzi w obszarach bez komórek





## Progowanie adaptacyjne - dodatkowy próg

- aby odrzucić piksele tła, gdzie nie ma zróżnicowania intensywności, wprowadza się **dodatkowy globalny próg  $C$**  o niewielkiej wartości (podniesienie poprzeczki – trudniej jest włączyć piksel do obiektu):

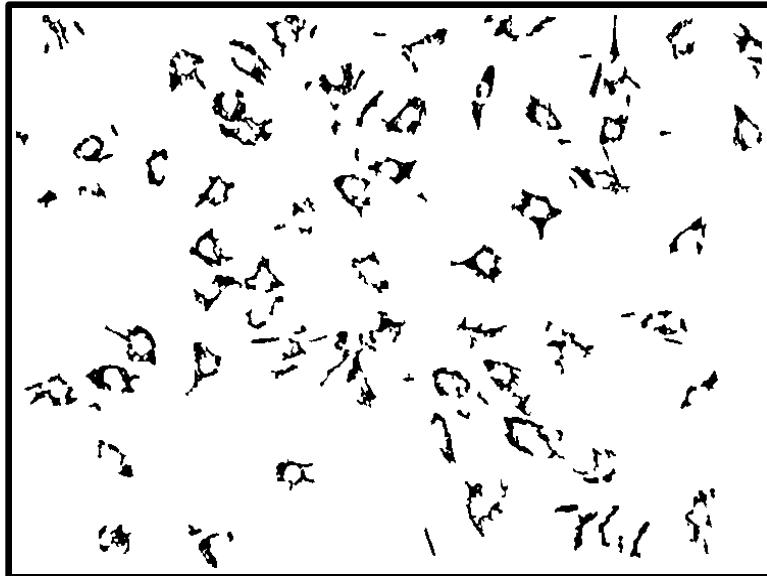
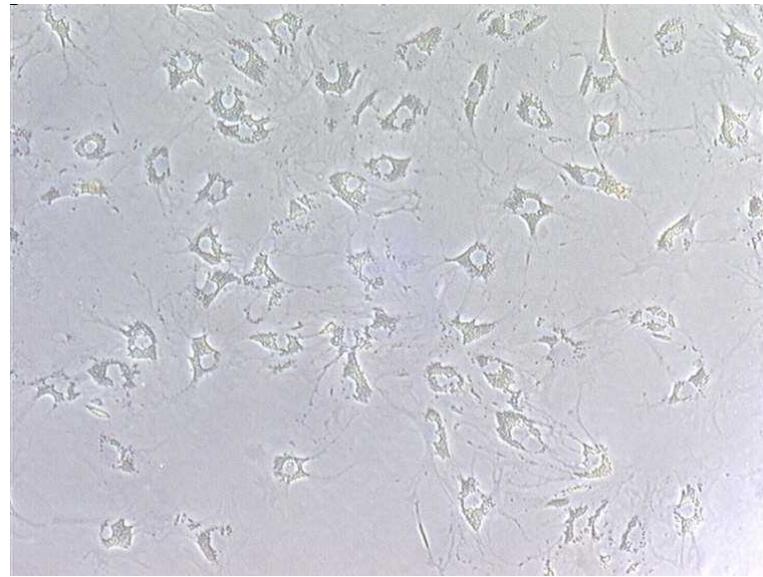
$$I(x,y) > \text{próg\_lokalny} + C \quad \Rightarrow \quad (x,y) \in \text{"obiekt"} \\ \text{w przeciwnym przypadku} \quad \Rightarrow \quad (x,y) \in \text{"tło"}$$

- w obszarach jednolitych wszystkie piksele mają intensywność zbliżoną do średniej – zwiększyły o  $C$  próg całkowity odrzuci je
- w przypadku obecności jaśniejszego obiektu średnia będzie mniejsza od jasności obiektu i większa od tła – stosunkowo mała wartość  $C$  nie wpłynie na segmentację obiektu
- zastosowanie: jeden rodzaj obiektu o unikalnym zakresie intensywności (metoda nie sprawdzi się dla wielu obiektów na obrazie)
- rozmiary obszaru liczenia progu lokalnego i obiektu muszą zapewniać, że nie będzie obszarów całkowicie zawartych we wnętrzu obiektu

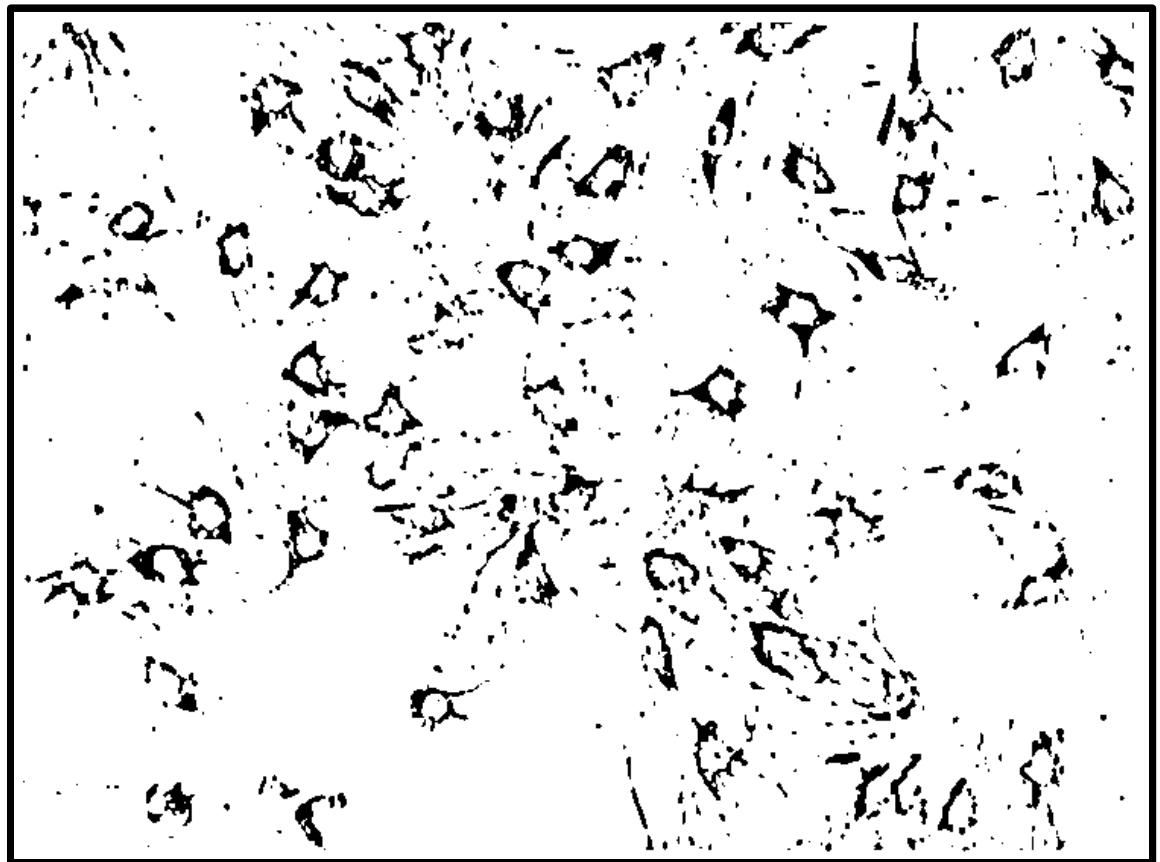


## Progowanie adaptacyjne

- lokalne progi jako średnie
- wprowadzony dodatkowy próg **C** powoduje odrzucenie tła
- wybór **C** jest interaktywny



po usunięciu najmniejszych grup





## Rozrost obszaru (*region growing*)

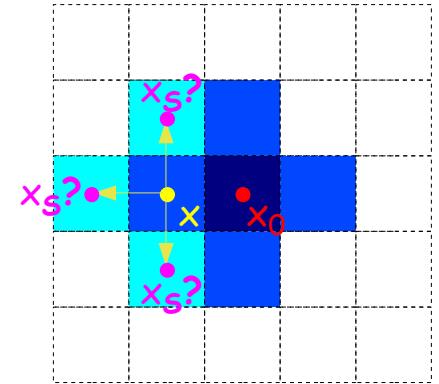
- progowanie nie wykorzystuje informacji o położeniu pikseli
- **rozrost obszaru** można traktować jako dodanie analizy położenia (**sąsiedztwa**) do techniki **progowania**
- oprócz kryterium odpowiedniej intensywności piksel zaliczany do obiektu musi sąsiadować z odpowiednimi pikselami – zaliczonymi już do obiektu
- metoda wymagająca podania dwóch parametrów wejściowych:
  - początkowego obiektu (który się sukcesywnie rozrasta), najczęściej w postaci punktu startowego (*seed point*): wybór ręczny, interaktywny lub z innego etapu analizy
  - kryterium dołączania nowych punktów
- metoda lokalna – nie wszystkie piksele są analizowane ! (wada i zaleta)
- podobieństwo do techniki grafiki komputerowej: wypełnianie *flood fill*



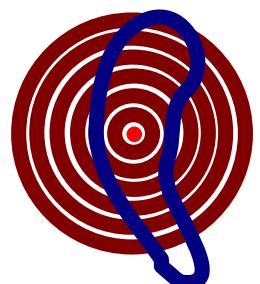
## Procedura rozrostu obszaru

- parametr wejściowy: punkt startowy  $x_0$
- pomocnicza struktura danych: stos  $S$  pikseli

```
S.push(x0);
zaznacz_jako_odwiedzony(x0);
obiekt.dodaj(x0);
while (!S.empty()) {
    x=S.top(); S.pop();
    foreach (xs: nieodwiedzony_sasiad(x) {
        zaznacz_jako_odwiedzony(xs);
        if (kryterium_włączenia(obiekt, xs)) {
            obiekt.dodaj(xs);
            S.push(xs);
        }
    }
}
```



- można użyć stosu programowego (rekurencja) ale jest to mało wydajne
- zamiast stosu (FILO, LIFO) można użyć kolejki FIFO – efekt końcowy zwykle ten sam (wyjątek: dynamicznie zmieniane kryterium włączania) ale inna kolejność włączania pikseli (od środka)
- różne definiowanie sąsiedztwa: 4, 8, siatka heksagonalna





## Kryteria włączania pikseli

- kryteria statyczne:

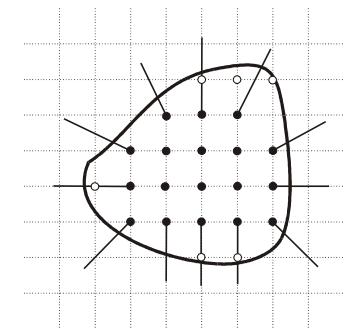
- progowanie ze stałym progiem (progami)
- włączanie pikseli z małą wartością amplitudy gradientu
- włączanie pikseli z małą wartością gradientu kierunkowego (np. kierunek prostopadły do bieżącej powierzchni obiektu)

- kryteria dynamiczne:

- porównywanie intensywności włączanego piksela ze średnią intensywnością bieżącego obszaru  
(średnia intensywność jest przeliczana po każdym kroku/  
po stałej liczbie kroków)
- sprawdzenie, czy włączenie nie zmieni homogeniczności obszaru  
(np. przekroczenie przez wariancję pewnego progu)
  - zależność od skali – inna waga piksela w małym i dużym obszarze
- dobór przedziału progowania na podstawie wariancji i średniej intensywności początkowego obszaru (nie punktu)

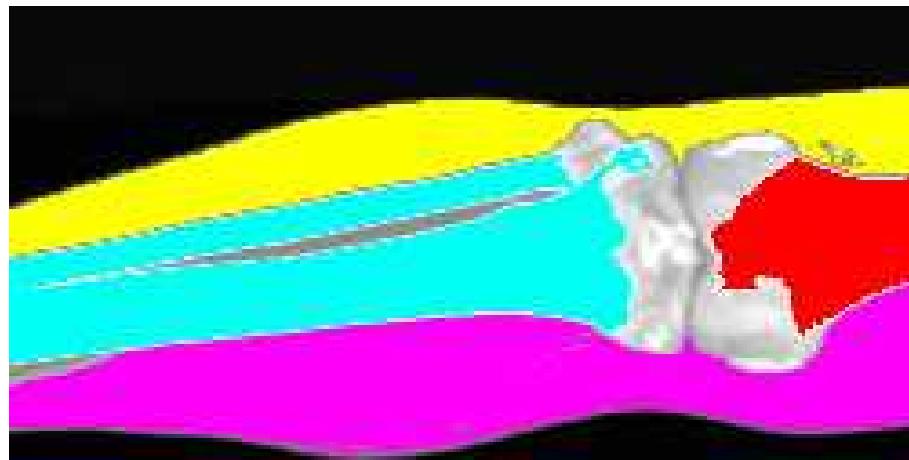
- wykorzystanie innych cech obrazu niż intensywność (np. cechy teksturalne)

- podawanie 2 punktów: startowego i nienależącego do obszaru





## Wyniki rozrostu obszaru

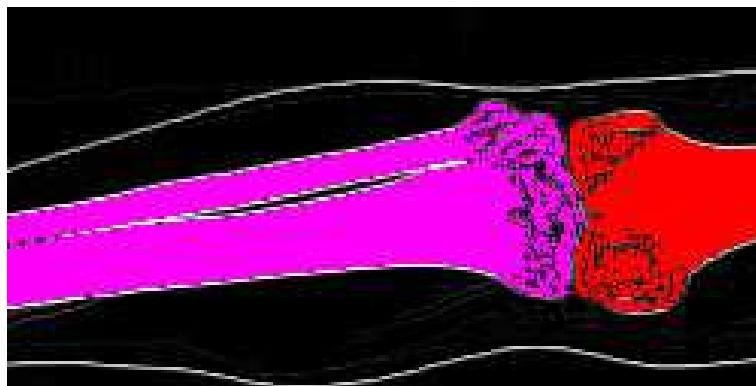
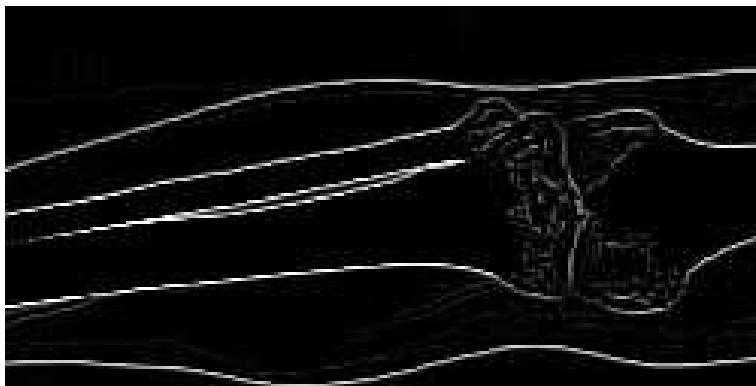


- każdy kolor oznacza inny punkt startowy
- kryterium:  
różnica intensywności piksela włączanego i startowego poniżej progu  
(wrażliwość na wybór punktu startowego)
- te same obiekty (tkanki) wyodrębniane jako oddzielne jeśli nie są połączone na obrazie
- obszary o zmiennej intensywności (chrząstka) niewyodrębnione



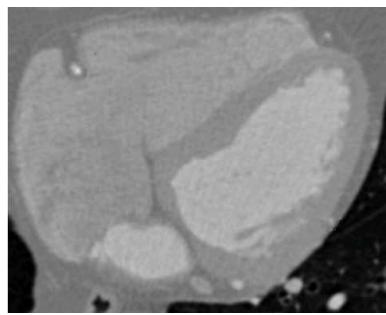
## Wyniki rozrostu obszaru

amplituda gradientu obrazu

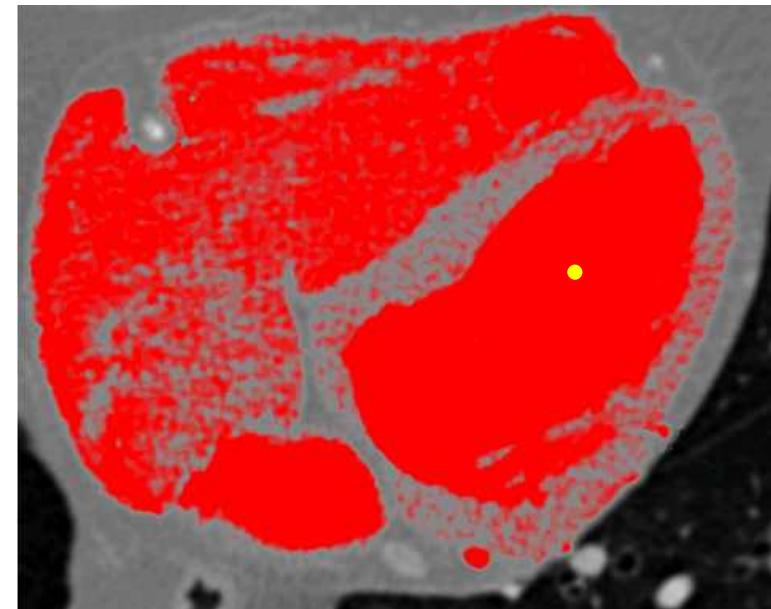
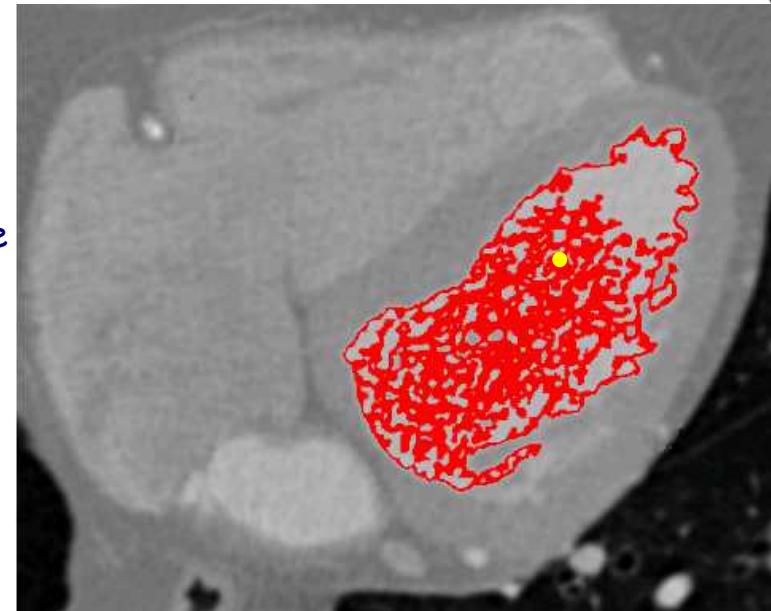
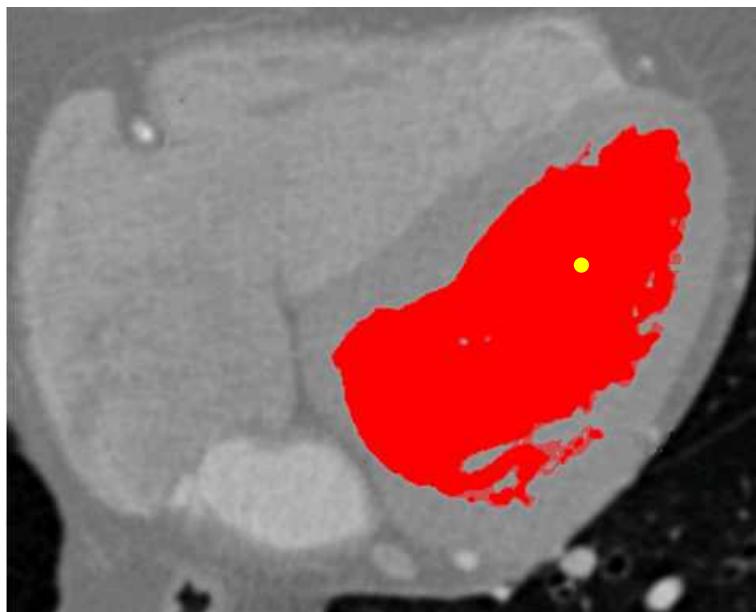


- kryterium:  
włączanie pikseli o małej wartości gradientu
- obszary o zmiennej intensywności (chrząstka) częściowo włączone
- potrzebna wyraźna granica aby algorytm „nie wypłynął”

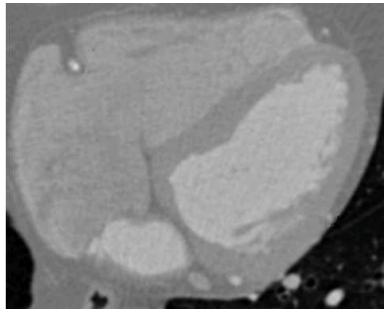
## Wyniki rozrostu obszaru



kryterium (progowanie) zbyt selektywne

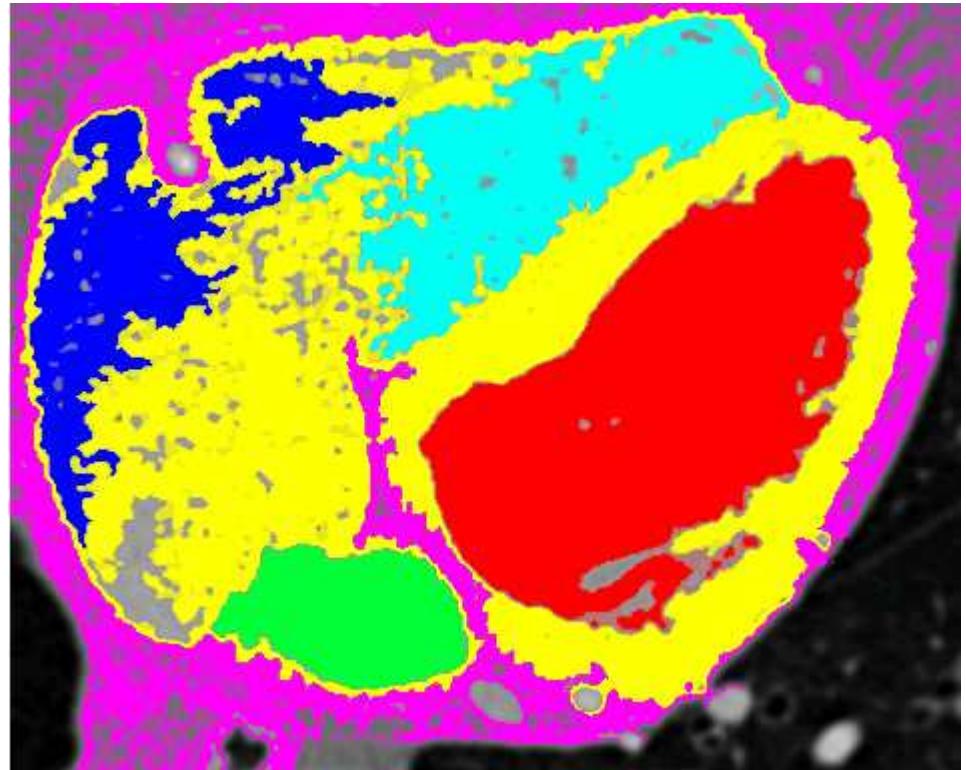


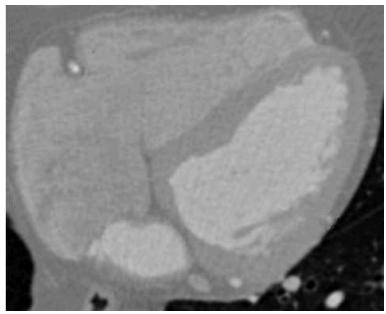
kryterium (progowanie) zbyt liberalne



## Wyniki rozrostu obszaru

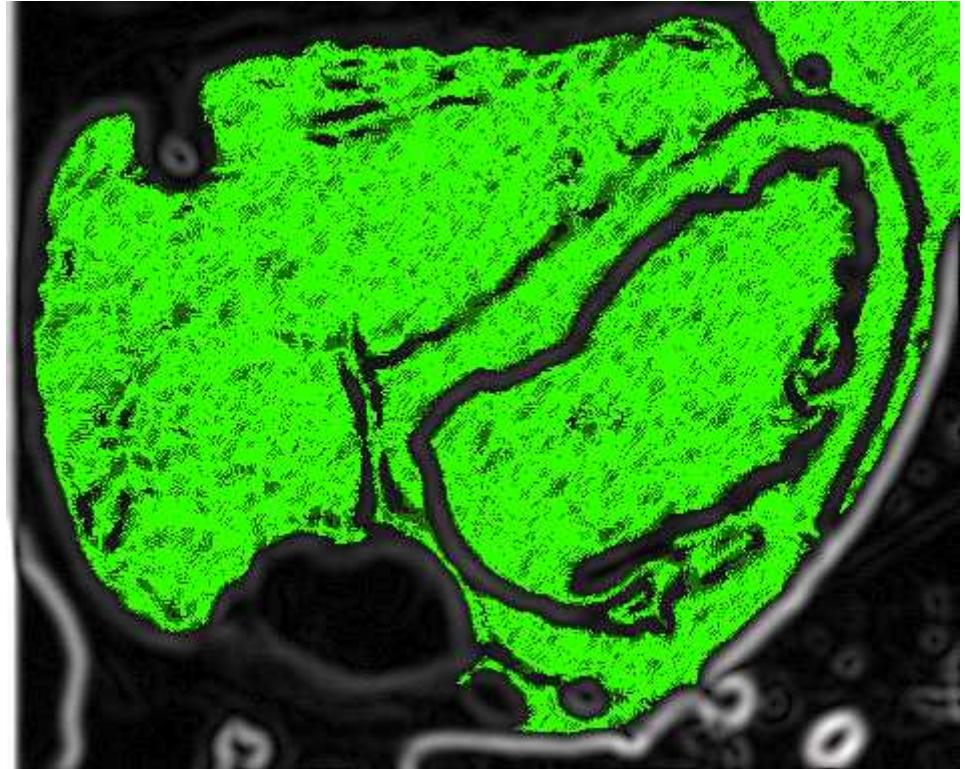
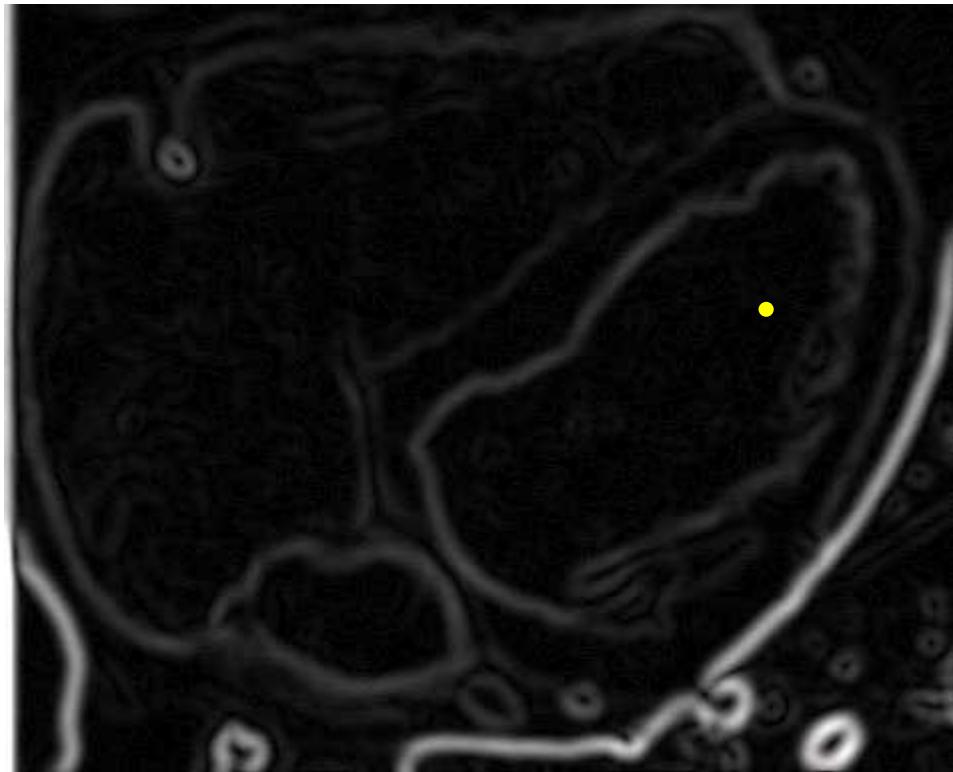
- każdy kolor oznacza inny punkt startowy
- kryterium:  
różnica intensywności piksela włączanego i startowego poniżej progu  
(wrażliwość na wybór punktu startowego)
- duży poziom szumu powoduje „poszarpane” granice





## Wyniki rozrostu obszaru

amplituda gradientu wygładzonego obrazu



- kryterium włączania: niska amplituda gradientu
- „rozlewanie się” algorytmu wskutek słabych krawędzi



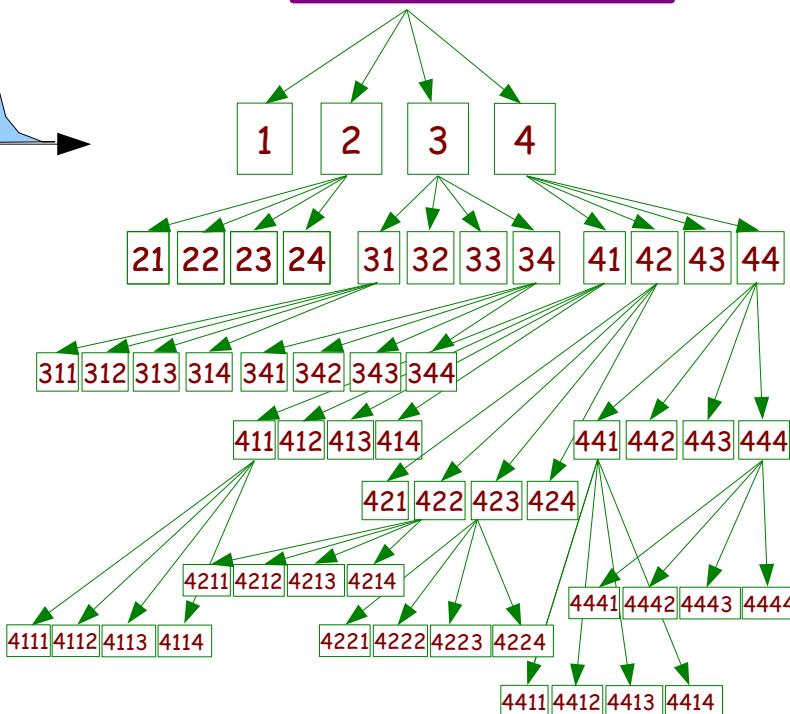
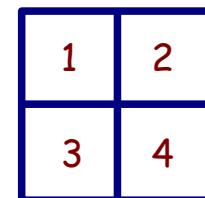
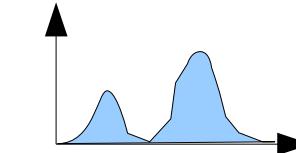
## Algorytm podziału i łączenia *(split & merge)*

- **rozrost obszaru** jest algorytmem interaktywnym – wymaga podawania punktów startowych – można to robić sukcesywnie aż do pełnego pokrycia obrazu obszarami wyodrębnionymi
- algorytm **podziału i łączenia** robi to automatycznie – po jego użyciu każdy piksel ma przyporządkowaną etykietę obszaru
- algorytm oparty na zasadzie podobieństw obszarów
- algorytm globalny – przetwarzany jest cały obraz
- 2 fazy algorytmu:
  - rekurencyjny podział na spójne (jednolite) obszary
  - łączenie podobnych obszarów



## Faza podziału

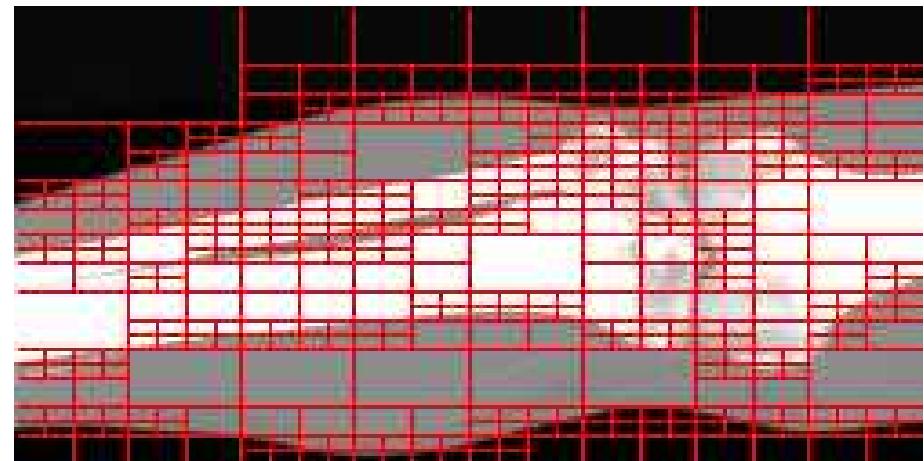
- obraz, początkowo traktowany jako jeden obszar, jest rekurencyjnie dzielony na coraz mniejsze obszary
- najczęściej podział na 4 równe prostokąty (kwadraty)
- kryteria podziału:
  - wariancja obszaru powyżej pewnego progu
  - histogram obszaru multimodalny
- można dopuścić podział do wielkości obszaru = 1 piksel lub założyć minimalny obszar nie podlegający już podziałowi
- kluczowa dla wydajności algorytmu reprezentacja podzielonych obszarów
  - numeracja obszarów cyframi na kolejnych pozycjach dziesiętnych
  - drzewa czwórkowe (*quad trees*)





## Przykłady podziału

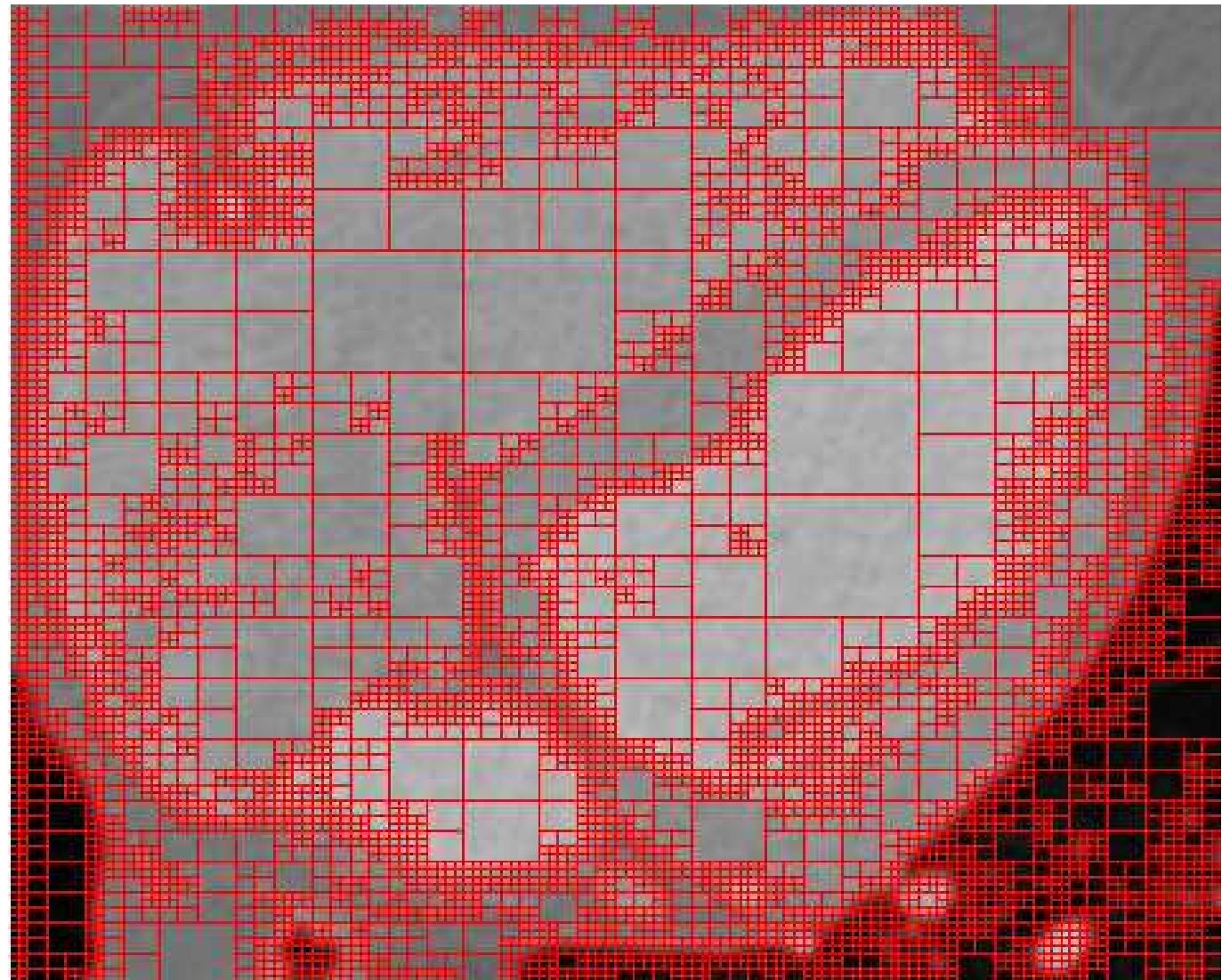
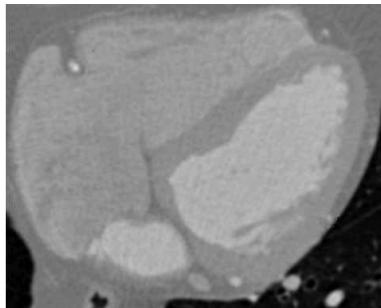
- obszar jest dzielony na 4 części jeśli jego wariancja przekracza próg
- minimalna wysokość obszaru – 5 pikseli
- jednolite obszary nie podlegają podziałowi
- najwięcej drobnych obszarów powstaje na granicach obiektów
- lepiej podzielić na zbyt wiele obszarów niż na zbyt mało



próg wariancji 0.005, intensywności pikseli z przedziału <0,1>



## Przykłady podziału

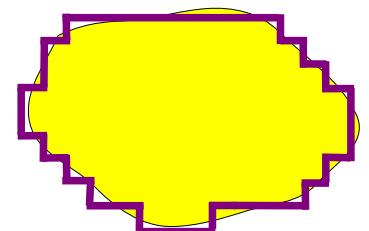
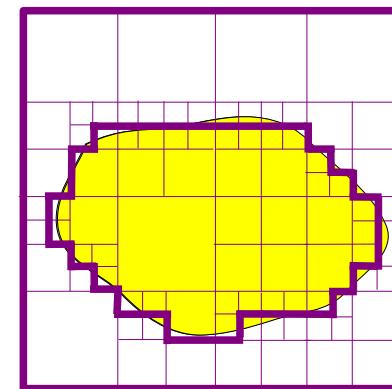
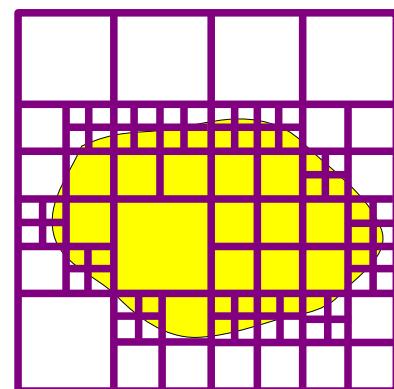
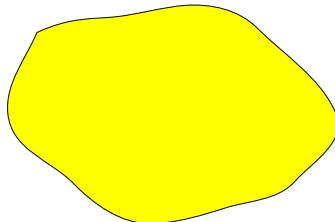


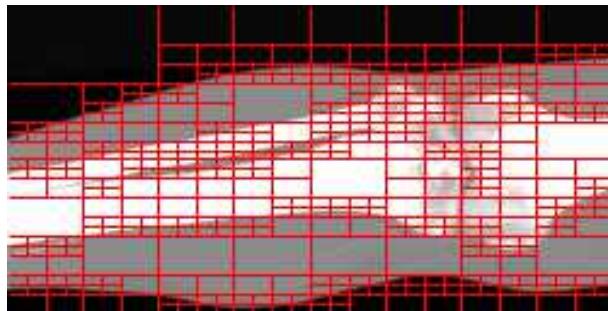
próg wariancji: 0.00025, minimalny rozmiar obszaru: 3 piksele



## Faza łączenia

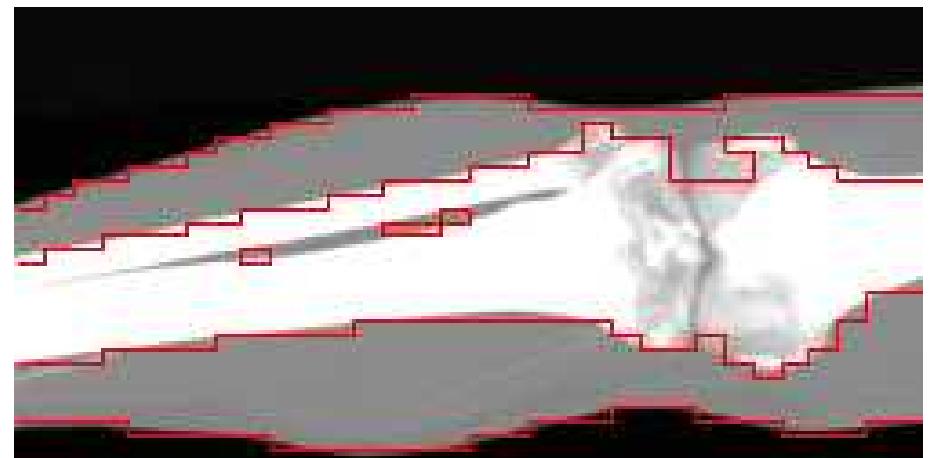
- obszary są już (względnie) jednolite
- niektóre sąsiadujące obszary są do siebie podobne (np. 2 podobszary jednego lub dwóch sąsiednich obszarów są jednolite )
- **sąsiadujące** podobszary łączy się, jeśli spełniają **kryterium łączenia**:
  - zblązona średnia intensywność
  - obszaru po złączeniu nie przekracza spełnia kryterium podziału:
    - histogram pozostaje unimodalny
    - wiariancja nie przekracza progu
  - krawędź między obszarami ni ma wysokiego gradientu (kryterium mało odporne na szum)
- procedura znajdywania sąsiednich obszarów może być złożona



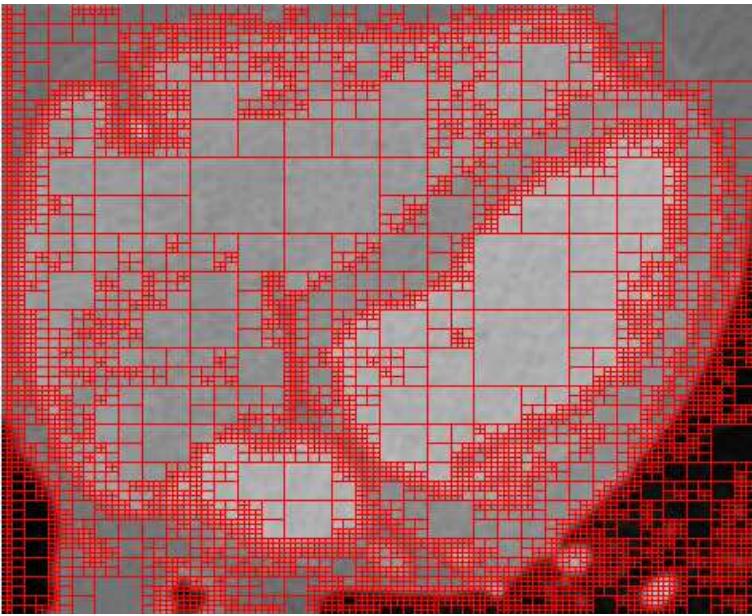


## Przykłady łączenia

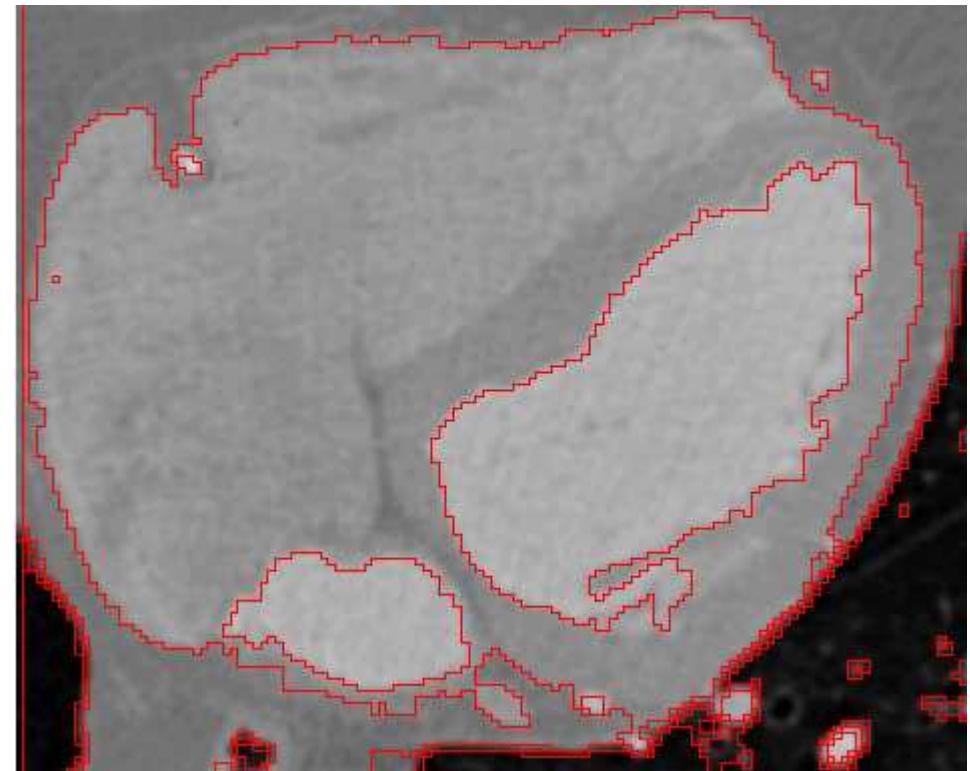
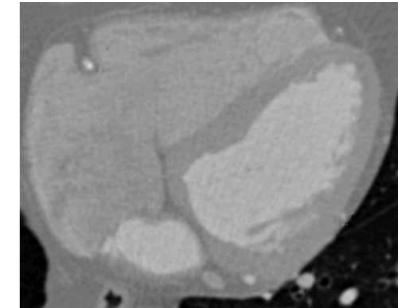
- obszary jednej tkanki połączone
- ze względu na ograniczoną rozdzielcość przestrzenną (minimalny rozmiar obszary):
  - krawędź jest łamana
  - małe podobszary nie są zidentyfikowane



kryterium łączenia: różnica średnich jasności poniżej 0.25



## Przykłady łączenia



- poprawnie zlokalizowane granice tkanek
- mniejszy obszar minimalny podziału skutkuje mniej „łamaną” granicą
- dużo dłuższa faza łączenia ze względu na wiele obszarów
- wrażliwość na parametr łączenia
- na granicy niepołączone obszary

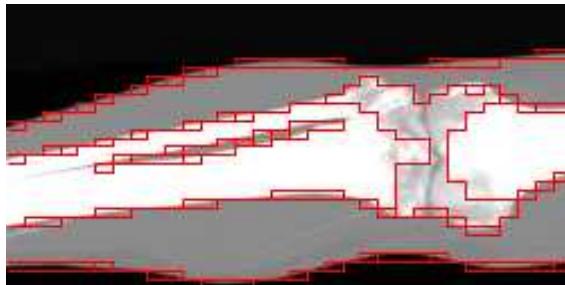
kryterium łączenia: różnica średnich jasności poniżej 0.07



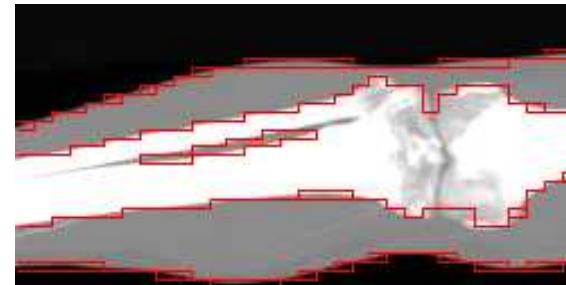
## Wrażliwość fazy łączenia na parametr

- niewłaściwie dobrane kryterium łączenia owocuje zbyt dużą lub zbyt małą liczbą obrazów wynikowych
- ale czasami brak jest jednoznacznej odpowiedzi: co jest już innym obszarem a co wciąż tym samym (obszar kolana)

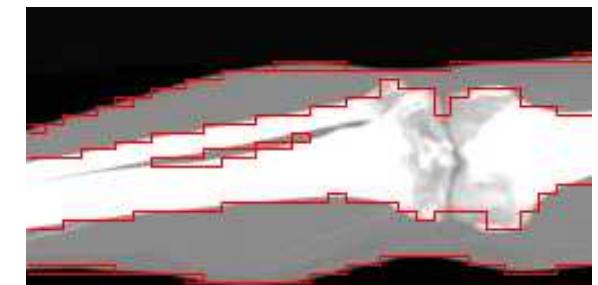
różnica średnich: 0.1



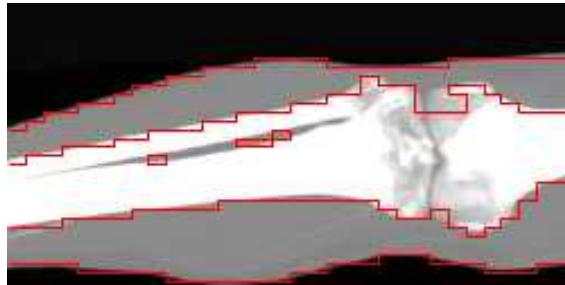
różnica średnich: 0.15



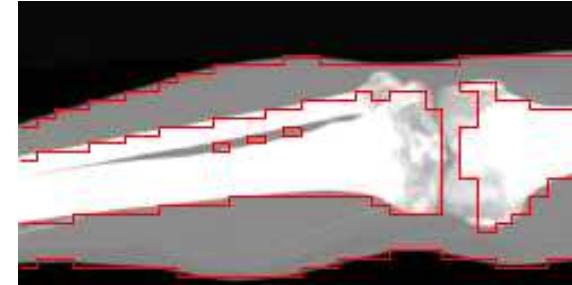
różnica średnich: 0.2



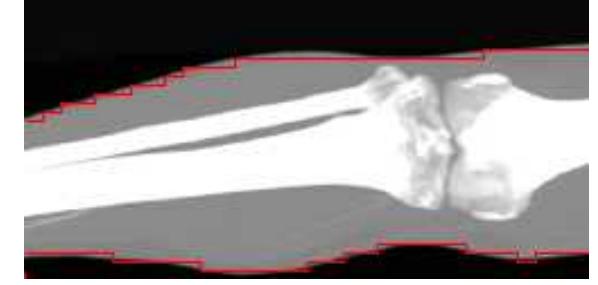
różnica średnich: 0.25



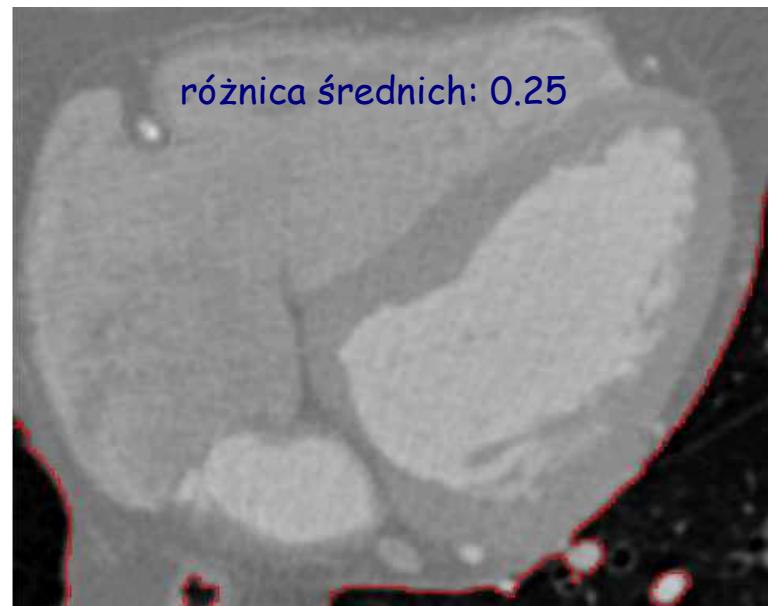
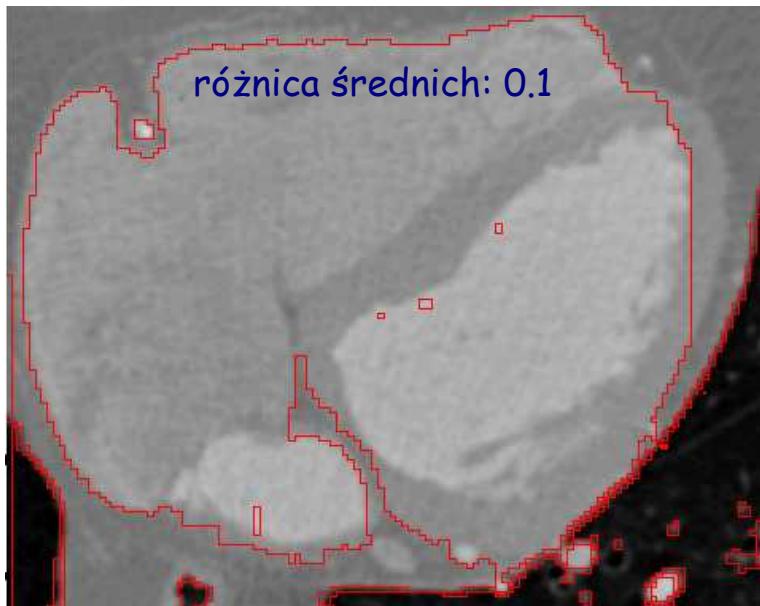
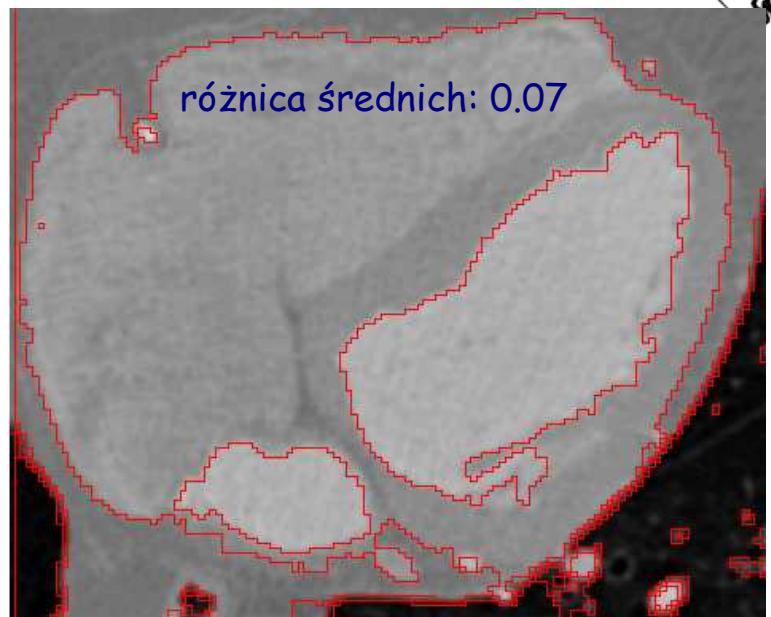
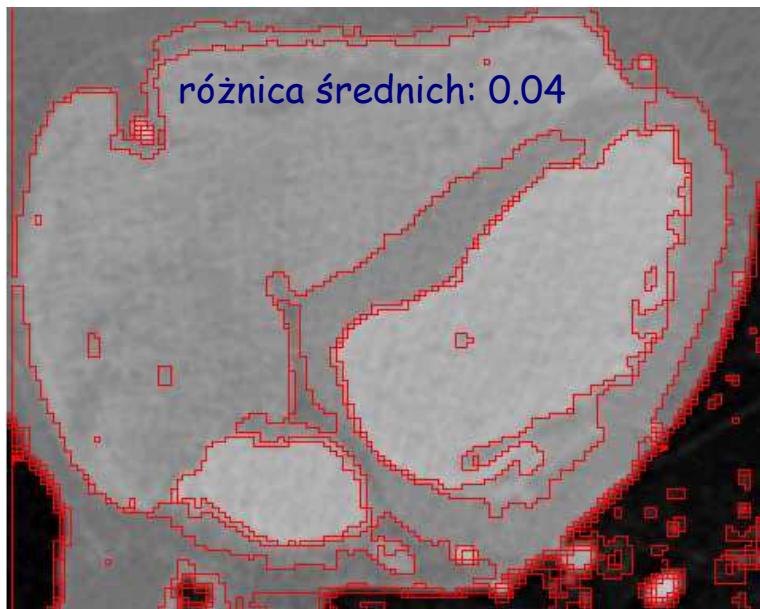
różnica średnich: 0.3



różnica średnich: 0.35



## Wrażliwość fazy łączenia na parametr





## Podsumowanie algorytmu podziału i łączenia

- można zdefiniować inną strategię podziału:
  - na obszary zbliżone do kwadratu: niezależność od prostokątności obrazu
  - na 2 podobszary powstałe w wyniku progowania, np. adaptacyjnego
    - dokładne odwzorowanie kształtów obiektu, bez bloków
    - trudna reprezentacja takich obszarów
- automatyczne określenie liczby obiektów
- wrażliwość na dobór parametrów
- duża złożoność obliczeniowa i czasowa (szczególnie faza łączenia)

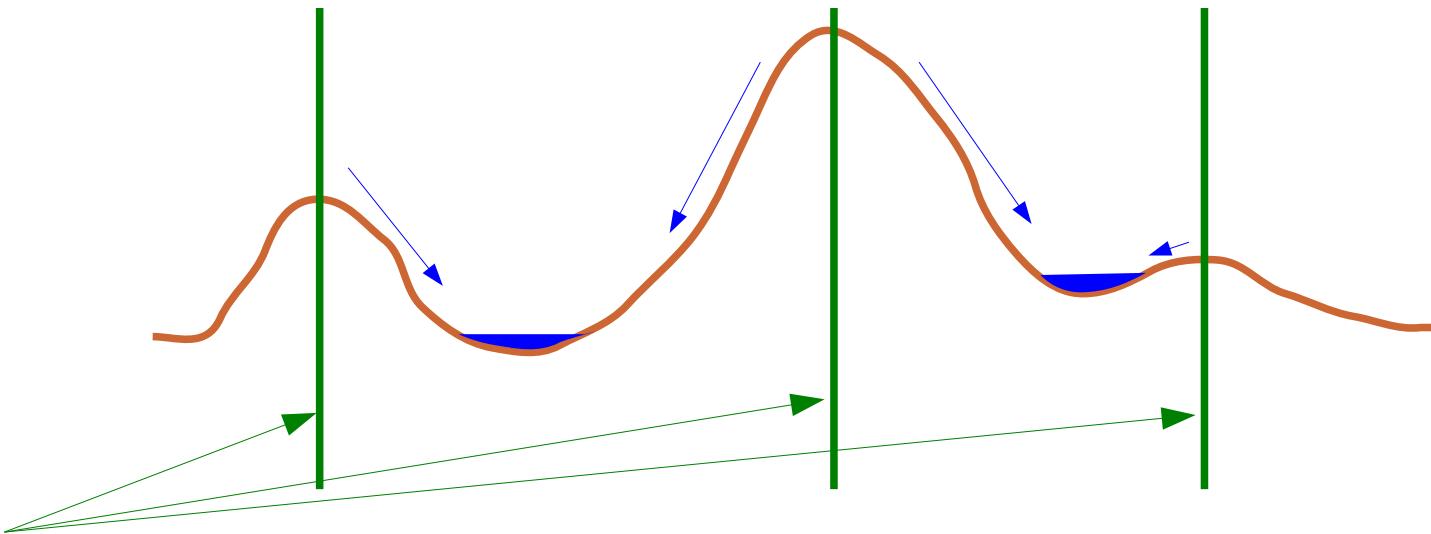


## Segmentacja wododziałowa *(watershed segmentation)*

- metoda segmentacji krawędziowej (jeśli opiera się na gradiencie)
- wywodzona z morfologii matematycznej
- czerpie z geografii – teorii zlewisk rzek
- może automatycznie dobierać liczbę wyodrębnionych obiektów lub pozwalać na podanie tej liczby przez operatora
- proces segmentacji jest tożsamy z zalewaniem wodą coraz wyżej położonych obszarów



## Pojęcia wododziału i zlewiska (watershed, catchment basin)

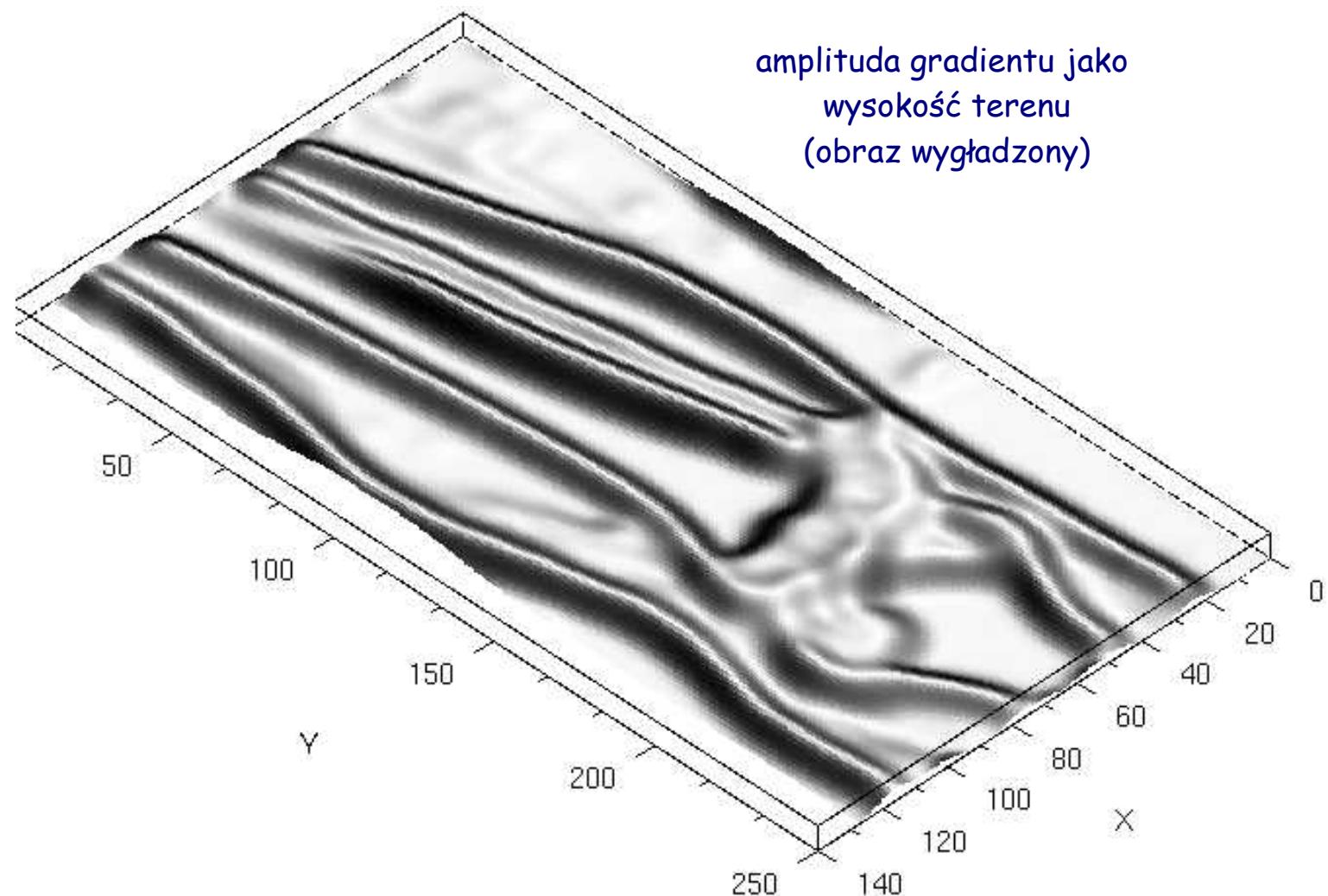


- **wododział** jest to linia rozgraniczająca tereny **zlewisk** – obszarów z których wody opadowe spływają do wspólnej rzeki lub zbiornika wodnego
- wododziały leżą na **grzbietach** (*ridges*) funkcji wysokości terenu
- grzbiet to punkt który nie jest pełnym maksimum ale maksimum w co najmniej jednym kierunku (w N-1 wymiarach)
  - są punktami maksymalnymi w przypadku funkcji 1D (maksima lokalne)
  - są liniami w przypadku funkcji 2D (rzeczywista rzeźba terenu)
  - są powierzchniami w przypadku funkcji 3D
- obszary (punktowe) do których spływają wody ze zlewisk leżą na **minimach** funkcji wysokości terenu



## Pojęcie wododziału w segmentacji

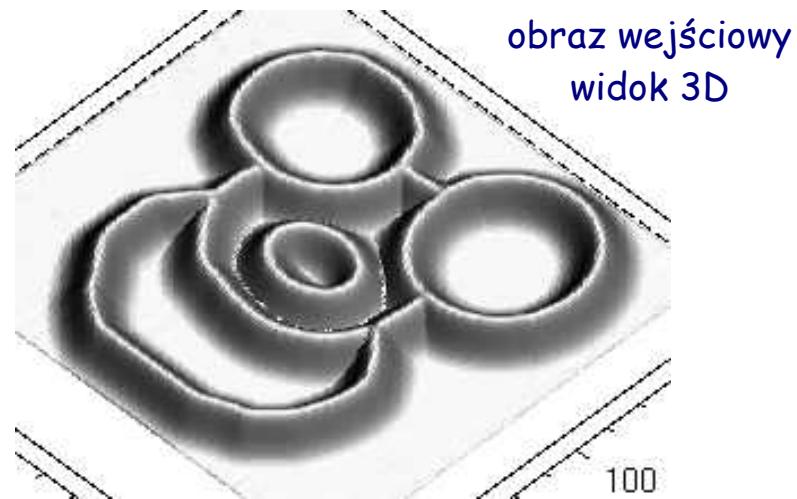
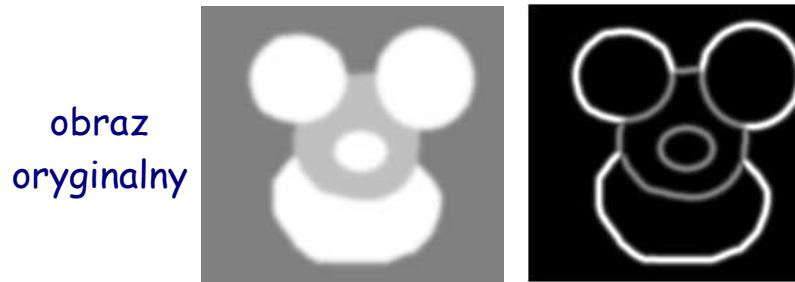
- jako wysokość terenu w segmentacji obrazów przyjmuje się najczęściej **amplitudę gradientu** obrazu lub (rzadziej) samą **intensywność** obrazu
- jako wododział zaś **grzbiety** tych funkcji





## Proces segmentacji wododziałowej podejście oparte o opad deszczu (rainfall)

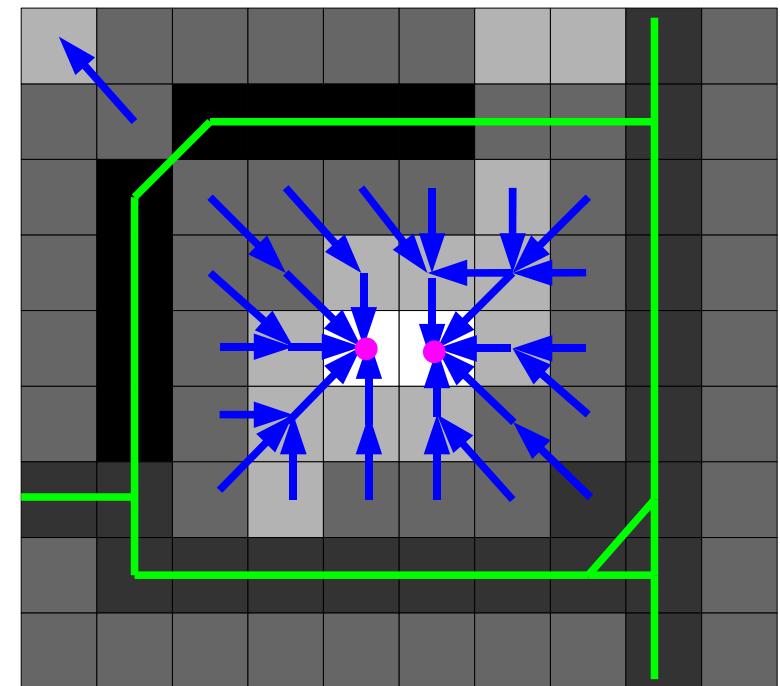
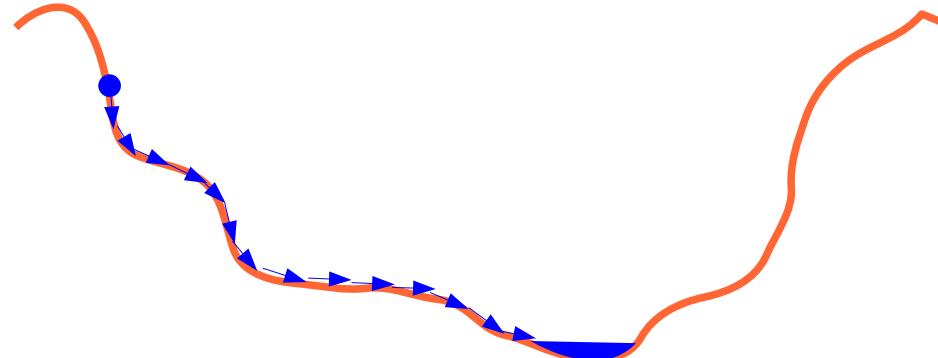
- na obrazie wejściowym wyliczany jest **gradient**
- ponieważ obrazem wejściowym jest najczęściej już przetworzony, poprzez wyliczenie amplitudy gradientu, obraz oryginalny, chodzi tu o gradient gradientu





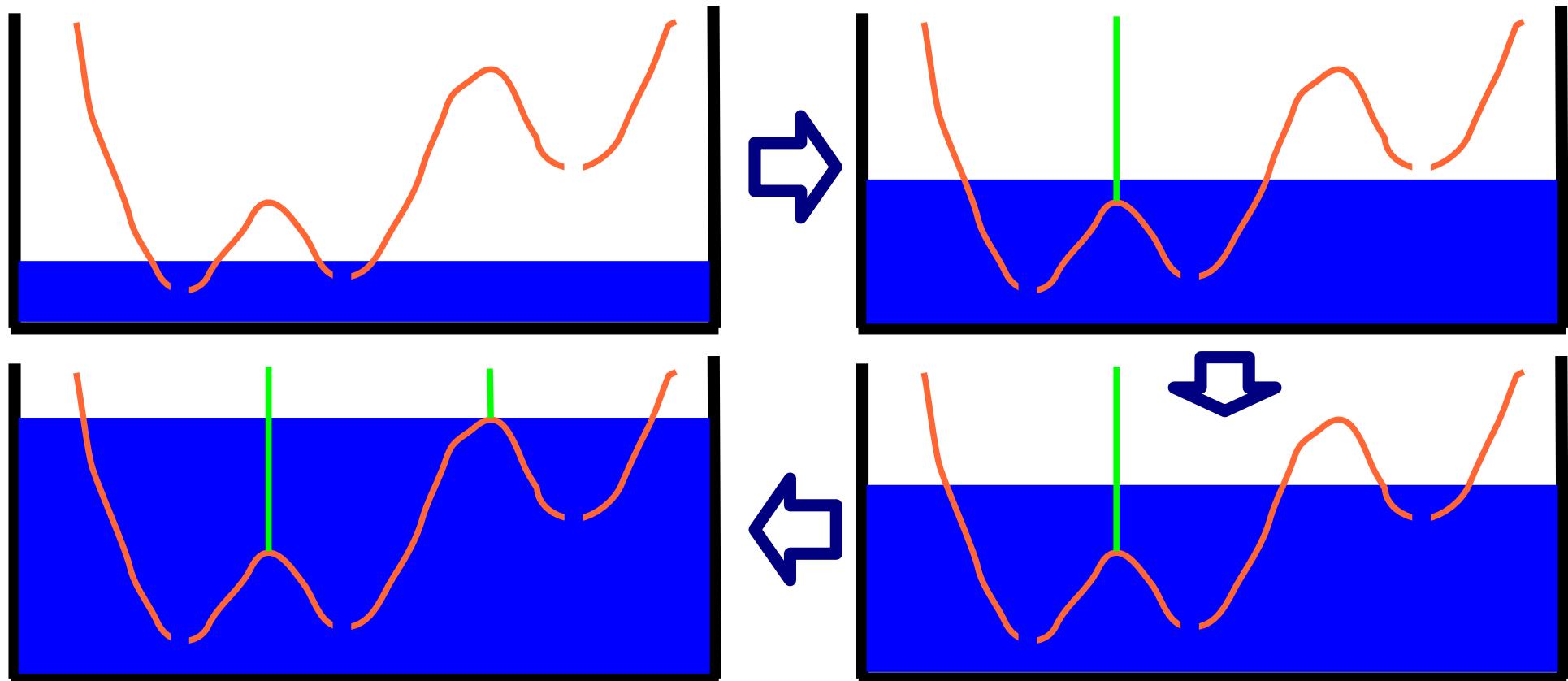
## Proces segmentacji wododziałowej podejście oparte o opad deszczu (rainfall)

- z każdego piksela obrazu wyliczana jest **droga do minimum lokalnego**
- droga ta wiedzie przeciwnie do kierunków lokalnego gradientu
- wszystkie piksele z których droga wiedzie do tego samego **minimum lokalnego** zaliczane są do jednego zlewiska
- granice między zlewiskami wyznaczają **wododział** – wynik segmentacji (kontury segmentowanego obiektu)



- obszary minimów mogą być płaskie, może to utrudniać ich identyfikację
- w przypadku dyskretnym 2D (obraz cyfrowy) kierunek gradientu określa do którego z 8 (4) sąsiadujących pikseli „przepływa woda opadowa”

## Proces segmentacji wododziałowej podejście oparte o zatapianie (sinking)



- w obszarze znajdują się **lokalne minima**
- „wierci się” w ich miejscach **dziury** w powierzchni „terenu”
- cały „teren” zanurza stopniowo w **wodzie**
- w miejscach, gdzie woda z dwóch otworów zetknie się ze sobą, stawia się „**tamy**” wyznaczające **wododział**



## Algorytm segmentacji wododziałowej (zatapianie)

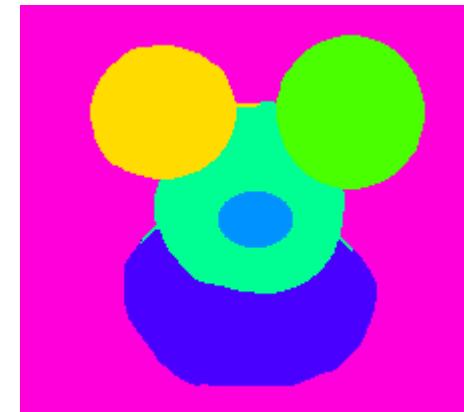
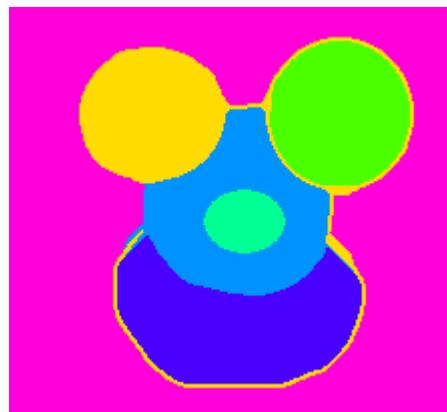
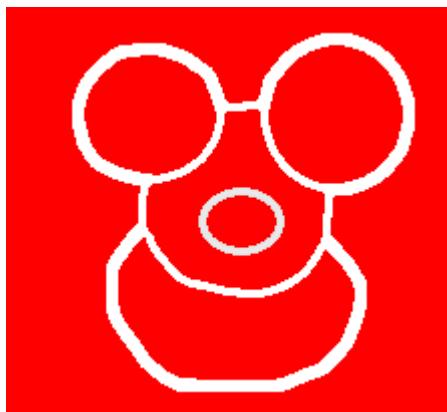
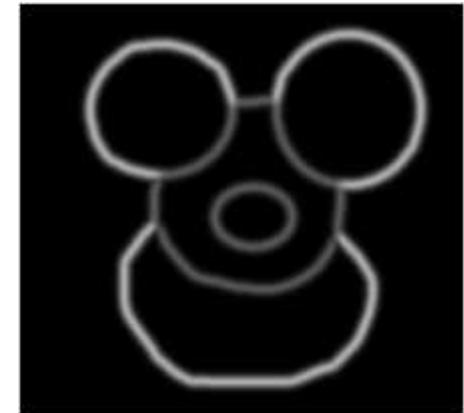
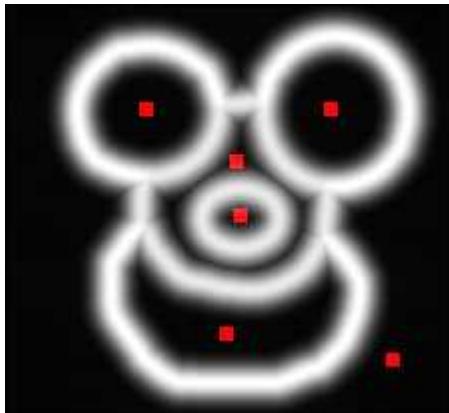
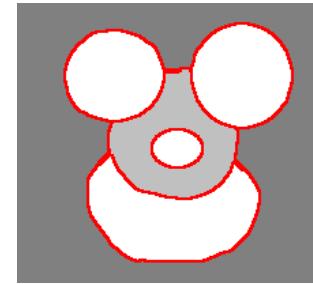
- oparta na algorytmie rozrostu obszaru
- duża złożoność (ale możliwość optymalizacji)

```
1. wyznacz lokalne minima jako początkowe zlewiska Z
2. uporządkuj je wg rosnącego poziomu intensywności:  $Z_0, Z_1, \dots, Z_N$ 
3. for (i=0;i<N;++i) //***wyrównywanie poziomów w zlewiskach
   while (nie osiągniesz poziomu wody  $Z_{i+1}$ )
      //***wszyskie zlewiska  $Z_0, \dots, Z_i$  mają już ten sam poziom wody
      for (j=0;j<=i;++j)
         - podnieś poziom wody  $Z_j$  o 1
            tzn. wykonaj algorytm rozrostu obszaru startując od obszaru  $Z_j$ ;
            kryterium włączania: intensywność pikseli == max(I( $Z_j$ ))+1;
            włączaj tylko piksele nie włączone jeszcze od innych zlewisk
4. while (nie wszystkie piksele włączone do zlewisk)
   //***równomierne podnoszenie poziomu wody we wszystkich zlewiskach
   for (i=0;i<N;++i)
      - podnieś poziom wody  $Z_i$  o 1
         tzn. wykonaj algorytm rozrostu obszaru .....
```

- każde ze zlewisk otrzymuje inny kolor
- granice między kolorami wyznaczają wododział (wynik segmentacji)



## Segmentacja przez zatapianie



kolejne fazy segmentacji wododziałowej,  
obraz krawędzi (amplitudy gradientu) został spłaszczony - efektem są  
otoczki obszarów przez inny obszar;  
ostatni obraz pokazuje wyodrębnione obszary

efekt segmentacji  
przy użyciu oryginalnej  
amplitudy gradientu



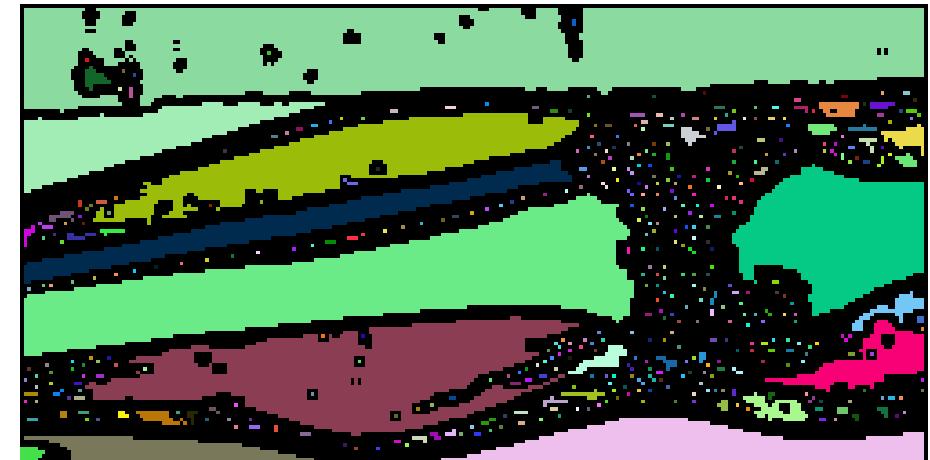
## Segmentacja wododziałowa obrazów rzeczywistych



obraz oryginalny



amplituda gradientu



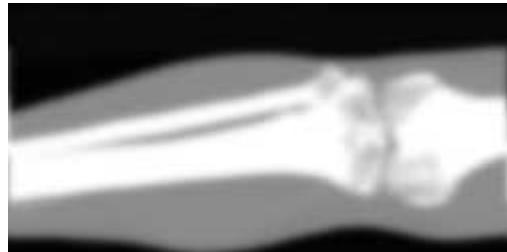
wyliczone minima lokalne (478 minimów)



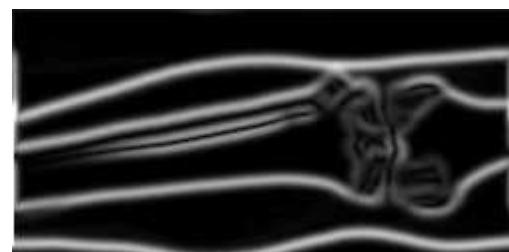
segmentacja wododziałowa: 478 obszarów



## Ograniczanie liczby minimów przez wygładzenie



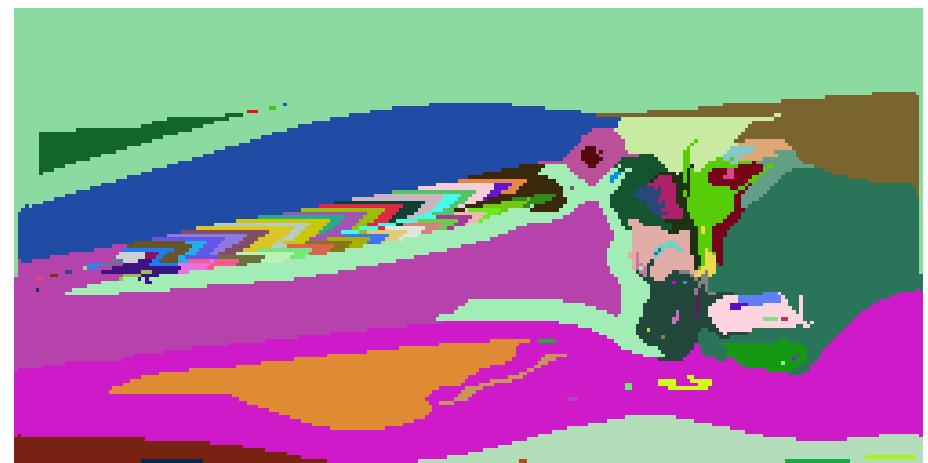
obraz oryginalny po wygładzeniu



amplituda gradientu



wyliczone minima lokalne (115 minów)

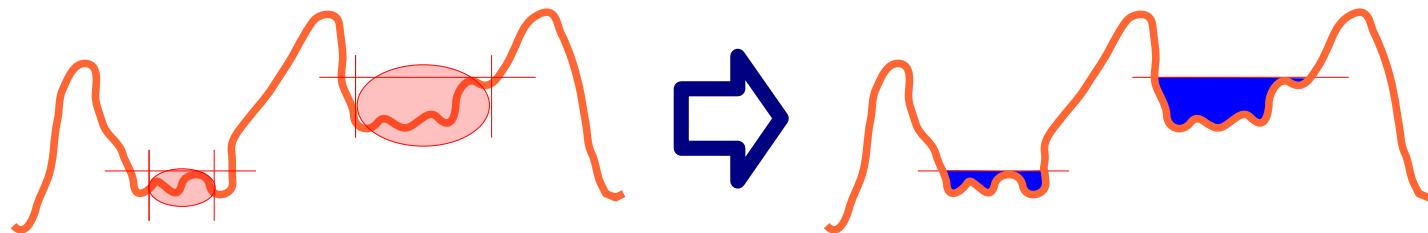


segmentacja wododziałowa: 115 obszarów



## Technika markerów

- jeśli znana jest liczba i orientacyjne położenie segmentowanych obiektów można poprawić efektywność segmentacji wododziałowej
- na obrazie zaznacza się tzw. **markery**: obszary, które leżą wewnątrz segmentowanych obiektów
- zastępują one (bardzo) liczne niekiedy minima lokalne
- zaznaczenie jest zwykle interaktywne lub z procesu wyższego rzędu



- markery są od razы „wypełniane wodą” do maksymalnej wysokości w ich obszarze
- dalej procedura zatapiania odbywa się standardowo
- nie pojawiają się już nowe zlewiska



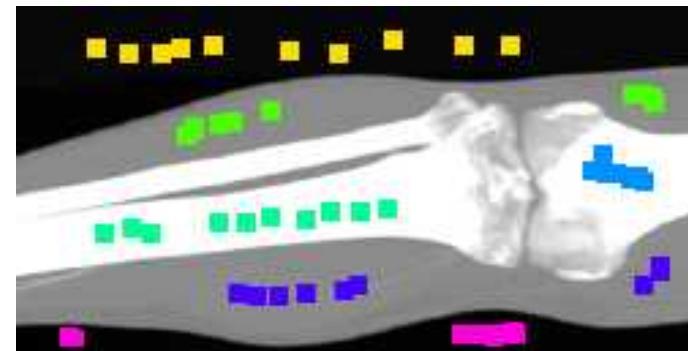
## Segmentacja z wykorzystaniem markerów



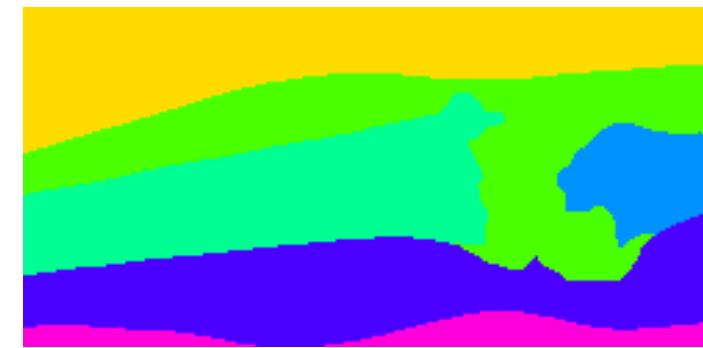
obraz oryginalny



amplituda gradientu



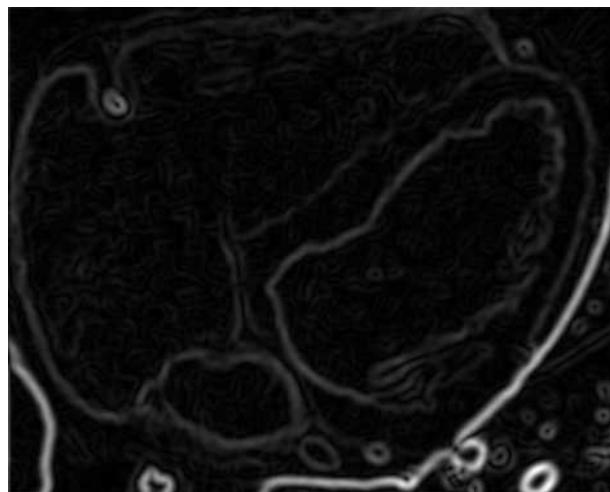
interaktywnie zaznaczone markery



wynik segmentacji wododziałowej

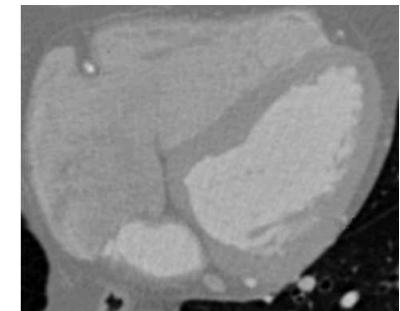
- poprawna detekcja obszarów z wyraźnymi krawędziami
- problemy w obszarach, gdzie krawędź jest niewyraźna
- niezaznaczone obszary pozostają niewykryte

## Segmentacja z wykorzystaniem markerów

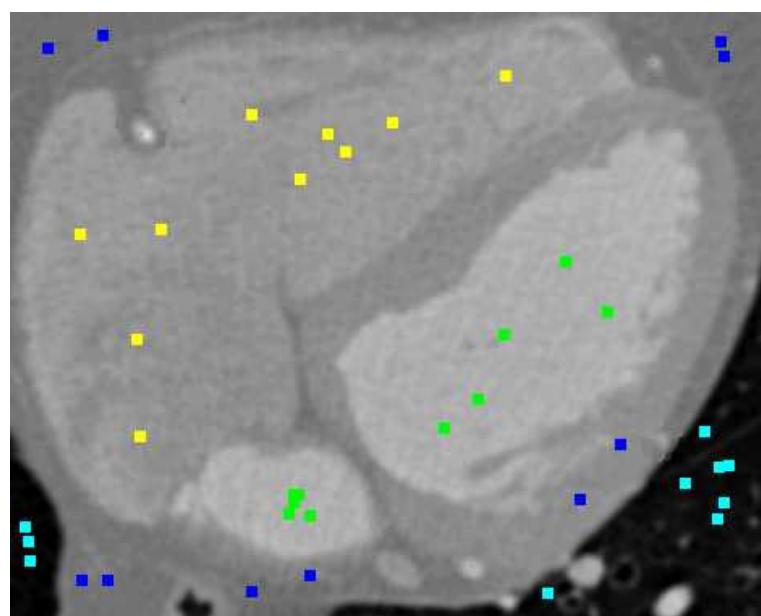


amplituda gradientu

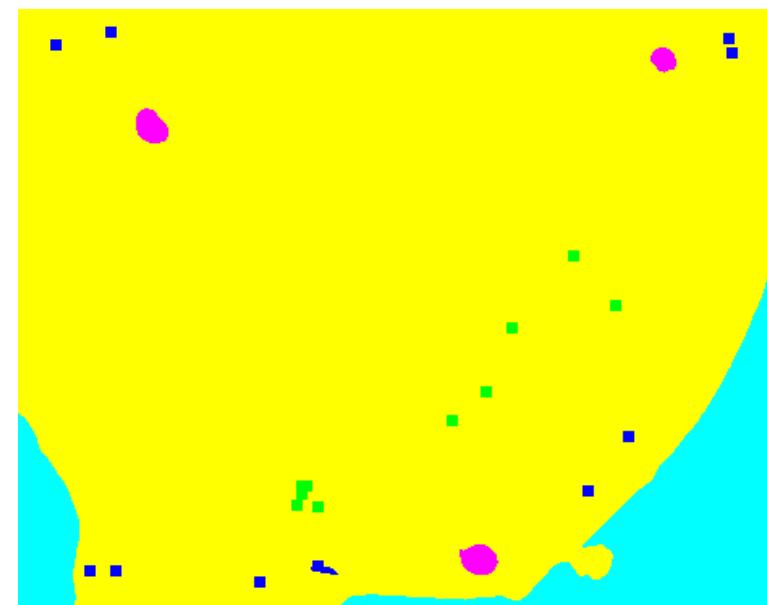
- słabe krawędzie powodują „zalanie” obszarów o różnej intensywności
- długie czas segmentacji



obraz oryginalny

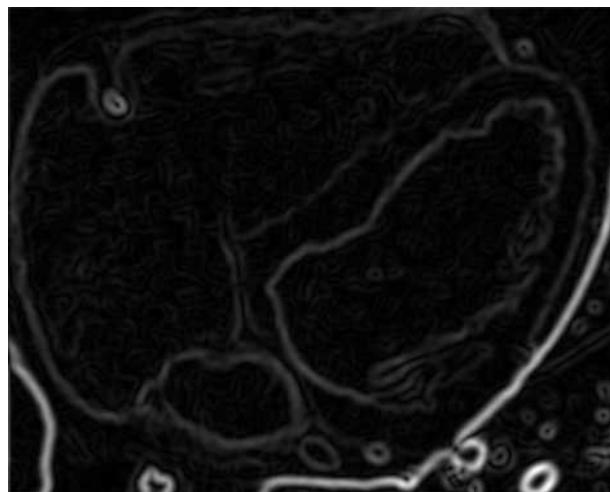


interaktywnie zaznaczone markery



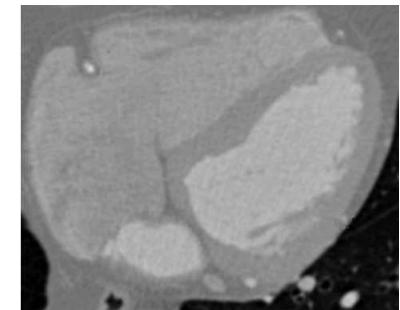
wynik segmentacji wododziałowej

## Segmentacja z wykorzystaniem markerów

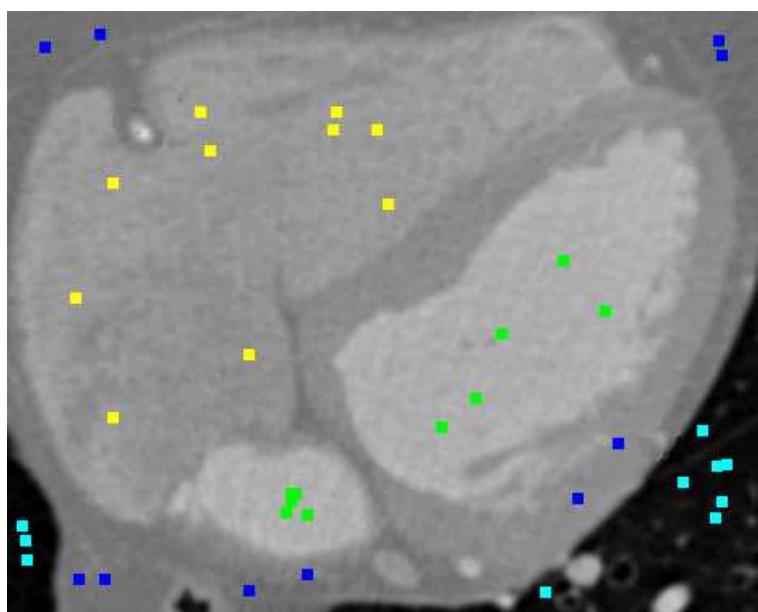


amplituda gradientu

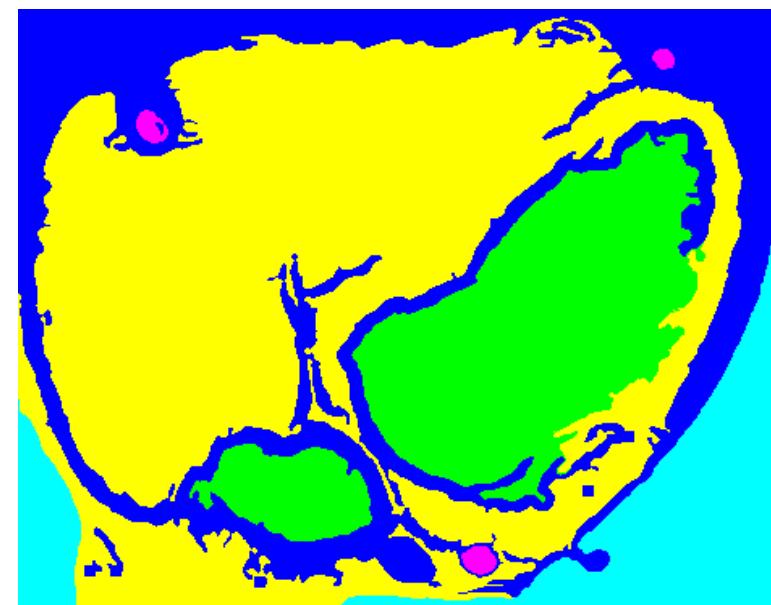
- wybór markerów które nie leżą na krawędziach ogranicza „rozłanie” się tych markerów
- **metoda bardzo wrażliwa na ciągłość granic !!!**



obraz oryginalny



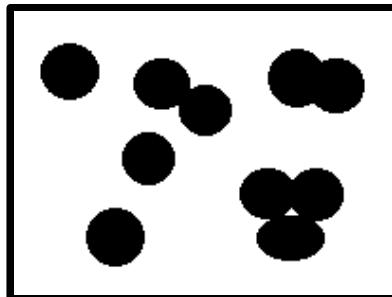
inne pozycje markerów żółtych



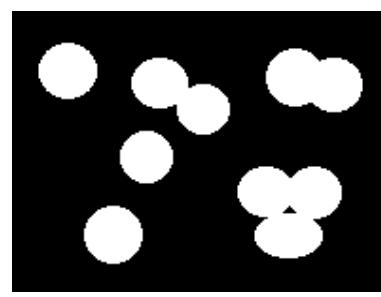
wynik segmentacji wododziałowej



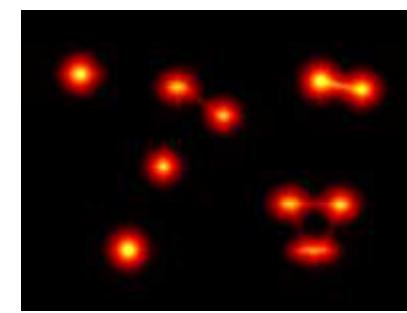
## Wykorzystanie segmentacji wododziałowej do separacji nakładających się obiektów (przykład ze Scilaba)



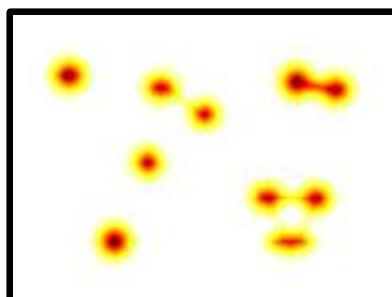
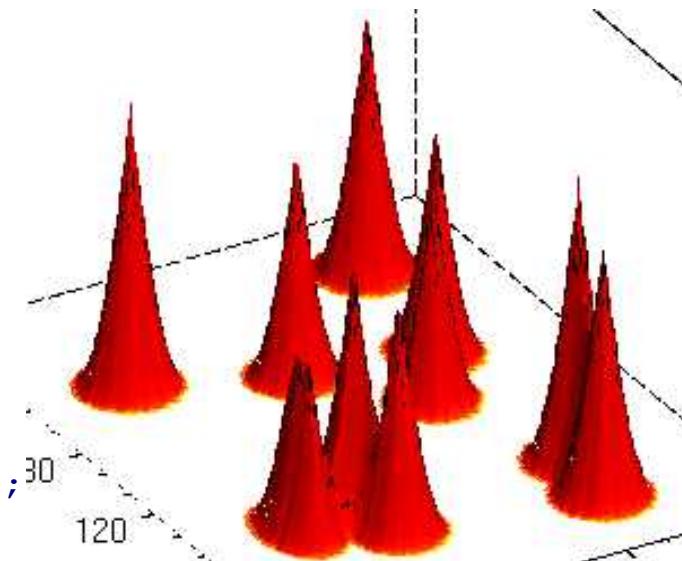
obraz oryginalny  
a



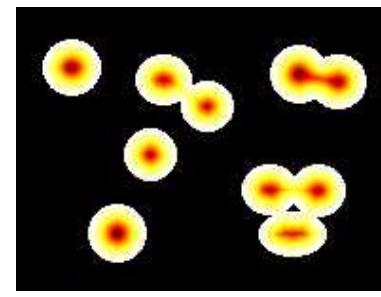
inwersja kolorów  
 $a=1-a;$



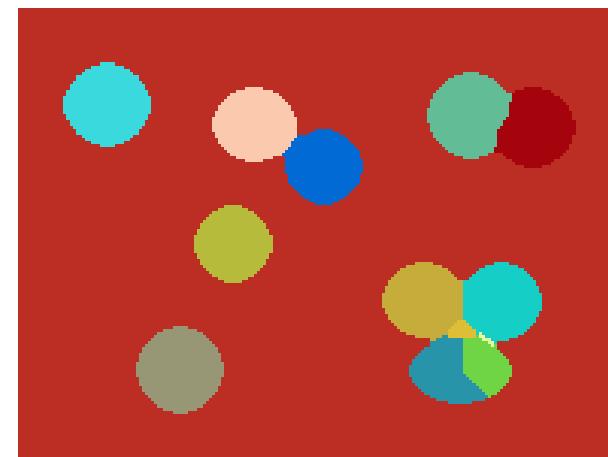
transformata odległościowa  
 $d=\text{normal}(\text{bwdist}(a), 255);$



zamiana maximów na minima  
 $d=255-d;$



wyzerowanie tła  
 $d=d.*a;$



tsegmentacja wododziałowa  
 $w=\text{watershed}(d/255);$

- problem z detekcją lokalnych minimumw
- są lepsze metody

# GNU Free Documentation License



Version 1.2, November 2002

Copyright (C) 2000,2001,2002 Free Software Foundation, Inc.  
51 Franklin St, Fifth Floor, Boston, MA 02110-1301 USA  
Everyone is permitted to copy and distribute verbatim copies  
of this license document, but changing it is not allowed.

## 0. PREAMBLE

The purpose of this License is to make a manual, textbook, or other functional and useful document "free" in the sense of freedom: to assure everyone the effective freedom to copy and redistribute it, with or without modifying it, either commercially or noncommercially. Secondly, this License preserves for the author and publisher a way to get credit for their work, while not being considered responsible for modifications made by others.

This License is a kind of "copyleft", which means that derivative works of the document must themselves be free in the same sense. It complements the GNU General Public License, which is a copyleft license designed for free software.

We have designed this License in order to use it for manuals for free software, because free software needs free documentation: a free program should come with manuals providing the same freedoms that the software does. But this License is not limited to software manuals; it can be used for any textual work, regardless of subject matter or whether it is published as a printed book. We recommend this License principally for works whose purpose is instruction or reference.

## 1. APPLICABILITY AND DEFINITIONS

This License applies to any manual or other work, in any medium, that contains a notice placed by the copyright holder saying it can be distributed under the terms of this License. Such a notice grants a world-wide, royalty-free license, unlimited in duration, to use that work under the conditions stated herein. The "Document", below, refers to any such manual or work. Any member of the public is a licensee, and is addressed as "you". You accept the license if you copy, modify or distribute the work in a way requiring permission under copyright law.

A "Modified Version" of the Document means any work containing the Document or a portion of it, either copied verbatim, or with modifications and/or translated into another language.

A "Secondary Section" is a named appendix or a front-matter section of the Document that deals exclusively with the relationship of the publishers or authors of the Document to the Document's overall subject (or to related matters) and contains nothing that could fall directly within that overall subject. (Thus, if the Document is in part a textbook of mathematics, a Secondary Section may not explain any mathematics.) The relationship could be a matter of historical connection with the subject or with related matters, or of legal, commercial, philosophical, ethical or political position regarding them.

The "Invariant Sections" are certain Secondary Sections whose titles are designated, as being those of Invariant Sections, in the notice that says that the Document is released under this License. If a section does not fit the above definition of Secondary then it is not allowed to be designated as Invariant. The Document may contain zero Invariant Sections. If the Document does not identify any Invariant Sections then there are none.

The "Cover Texts" are certain short passages of text that are listed, as Front-Cover Texts or Back-Cover Texts, in the notice that says that the Document is released under this License. A Front-Cover Text may be at most 5 words, and a Back-Cover Text may be at most 25 words.

A "Transparent" copy of the Document means a machine-readable copy, represented in a format whose specification is available to the general public, that is suitable for revising the document straightforwardly with generic text editors or (for images composed of pixels) generic paint programs or (for drawings) some widely available drawing editor, and that is suitable for input to text formatters or for automatic translation to a variety of formats suitable for input to text formatters. A copy made in an otherwise Transparent file format whose markup, or absence of markup, has been arranged to thwart or discourage subsequent modification by readers is not Transparent. An image format is not Transparent if used for any substantial amount of text. A copy that is not "Transparent" is called "Opaque".

Examples of suitable formats for Transparent copies include plain ASCII without markup, Texinfo input format, LaTeX input format, SGML or XML using a publicly available DTD, and standard-conforming simple HTML, PostScript or PDF designed for human modification. Examples of transparent image formats include PNG, XCF and JPG. Opaque formats include proprietary formats that can be read and edited only by proprietary word processors, SGML or XML for which the DTD and/or processing tools are not generally available, and the machine-generated HTML, PostScript or PDF produced by some word processors for output purposes only.

The "Title Page" means, for a printed book, the title page itself, plus such following pages as are needed to hold, legibly, the material this License requires to appear in the title page. For works in formats which do not have any title page as such, "Title Page" means the text near the most prominent appearance of the work's title, preceding the beginning of the body of the text.

A section "Entitled XYZ" means a named subunit of the Document whose title either is precisely XYZ or contains XYZ in parentheses following text that translates XYZ in another language. (Here XYZ stands for a specific section name mentioned below, such as "Acknowledgements", "Dedications", "Endorsements", or "History".) To "Preserve the Title" of such a section when you modify the Document means that it remains a section "Entitled XYZ" according to this definition.

The Document may include Warranty Disclaimers next to the notice which states that this License applies to the Document. These Warranty Disclaimers are considered to be included by reference in this License, but only as regards disclaiming warranties: any other implication that these Warranty Disclaimers may have is void and has no effect on the meaning of this License.

## 2. VERBATIM COPYING

You may copy and distribute the Document in any medium, either commercially or noncommercially, provided that this License, the copyright notices, and the license notice saying this License applies to the Document are reproduced in all copies, and that you add no other conditions whatsoever to those of this License. You may not use technical measures to obstruct or control the reading or further copying of the copies you make or distribute. However, you may accept compensation in exchange for copies. If you distribute a large enough number of copies you must also follow the conditions in section 3.

You may also lend copies, under the same conditions stated above, and you may publicly display copies.



### 3. COPYING IN QUANTITY

If you publish printed copies (or copies in media that commonly have printed covers) of the Document, numbering more than 100, and the Document's license notice requires Cover Texts, you must enclose the copies in covers that carry, clearly and legibly, all these Cover Texts: Front-Cover Texts on the front cover, and Back-Cover Texts on the back cover. Both covers must also clearly and legibly identify you as the publisher of these copies. The front cover must present the full title with all words of the title equally prominent and visible. You may add other material on the covers in addition. Copying with changes limited to the covers, as long as they preserve the title of the Document and satisfy these conditions, can be treated as verbatim copying in other respects.

If the required texts for either cover are too voluminous to fit legibly, you should put the first ones listed (as many as fit reasonably) on the actual cover, and continue the rest onto adjacent pages.

If you publish or distribute Opaque copies of the Document numbering more than 100, you must either include a machine-readable Transparent copy along with each Opaque copy, or state in or with each Opaque copy a computer-network location from which the general network-using public has access to download using public-standard network protocols a complete Transparent copy of the Document, free of added material. If you use the latter option, you must take reasonably prudent steps, when you begin distribution of Opaque copies in quantity, to ensure that this Transparent copy will remain thus accessible at the stated location until at least one year after the last time you distribute an Opaque copy (directly or through your agents or retailers) of that edition to the public.

It is requested, but not required, that you contact the authors of the Document well before redistributing any large number of copies, to give them a chance to provide you with an updated version of the Document.

### 4. MODIFICATIONS

You may copy and distribute a Modified Version of the Document under the conditions of sections 2 and 3 above, provided that you release the Modified Version under precisely this License, with the Modified Version filling the role of the Document, thus licensing distribution and modification of the Modified Version to whoever possesses a copy of it. In addition, you must do these things in the Modified Version:

- \* A. Use in the Title Page (and on the covers, if any) a title distinct from that of the Document, and from those of previous versions (which should, if there were any, be listed in the History section of the Document). You may use the same title as a previous version if the original publisher of that version gives permission.
- \* B. List on the Title Page, as authors, one or more persons or entities responsible for authorship of the modifications in the Modified Version, together with at least five of the principal authors of the Document (all of its principal authors, if it has fewer than five), unless they release you from this requirement.
- \* C. State on the Title page the name of the publisher of the Modified Version, as the publisher.
- \* D. Preserve all the copyright notices of the Document.
- \* E. Add an appropriate copyright notice for your modifications adjacent to the other copyright notices.
- \* F. Include, immediately after the copyright notices, a license notice giving the public permission to use the Modified Version under the terms of this License, in the form shown in the Addendum below.
- \* G. Preserve in that license notice the full lists of Invariant Sections and required Cover Texts given in the Document's license notice.
- \* H. Include an unaltered copy of this License.
- \* I. Preserve the section Entitled "History", Preserve its Title, and add to it an item stating at least the title, year, new authors, and publisher of the Modified Version as given on the Title Page. If there is no section Entitled "History" in the Document, create one stating the title, year, authors, and publisher of the Document as given on its Title Page, then add an item describing the Modified Version as stated in the previous sentence.
- \* J. Preserve the network location, if any, given in the Document for public access to a Transparent copy of the Document, and likewise the network locations given in the Document for previous versions it was based on. These may be placed in the "History" section. You may omit a network location for a work that was published at least four years before the Document itself, or if the original publisher of the version it refers to gives permission.
- \* K. For any section Entitled "Acknowledgements" or "Dedications", Preserve the Title of the section, and preserve in the section all the substance and tone of each of the contributor acknowledgements and/or dedications given therein.
- \* L. Preserve all the Invariant Sections of the Document, unaltered in their text and in their titles. Section numbers or the equivalent are not considered part of the section titles.
- \* M. Delete any section Entitled "Endorsements". Such a section may not be included in the Modified Version.
- \* N. Do not retitle any existing section to be Entitled "Endorsements" or to conflict in title with any Invariant Section.
- \* O. Preserve any Warranty Disclaimers.

If the Modified Version includes new front-matter sections or appendices that qualify as Secondary Sections and contain no material copied from the Document, you may at your option designate some or all of these sections as invariant. To do this, add their titles to the list of Invariant Sections in the Modified Version's license notice. These titles must be distinct from any other section titles.

You may add a section Entitled "Endorsements", provided it contains nothing but endorsements of your Modified Version by various parties--for example, statements of peer review or that the text has been approved by an organization as the authoritative definition of a standard.

You may add a passage of up to five words as a Front-Cover Text, and a passage of up to 25 words as a Back-Cover Text, to the end of the list of Cover Texts in the Modified Version. Only one passage of Front-Cover Text and one of Back-Cover Text may be added by (or through arrangements made by) any one entity. If the Document already includes a cover text for the same cover, previously added by you or by arrangement made by the same entity you are acting on behalf of, you may not add another; but you may replace the old one, on explicit permission from the previous publisher that added the old one.

The author(s) and publisher(s) of the Document do not by this License give permission to use their names for publicity for or to assert or imply endorsement of any Modified Version.



## 6. COLLECTIONS OF DOCUMENTS

You may make a collection consisting of the Document and other documents released under this License, and replace the individual copies of this License in the various documents with a single copy that is included in the collection, provided that you follow the rules of this License for verbatim copying of each of the documents in all other respects.

You may extract a single document from such a collection, and distribute it individually under this License, provided you insert a copy of this License into the extracted document, and follow this License in all other respects regarding verbatim copying of that document.

## 7. AGGREGATION WITH INDEPENDENT WORKS

A compilation of the Document or its derivatives with other separate and independent documents or works, in or on a volume of a storage or distribution medium, is called an "aggregate" if the copyright resulting from the compilation is not used to limit the legal rights of the compilation's users beyond what the individual works permit. When the Document is included in an aggregate, this License does not apply to the other works in the aggregate which are not themselves derivative works of the Document.

If the Cover Text requirement of section 3 is applicable to these copies of the Document, then if the Document is less than one half of the entire aggregate, the Document's Cover Texts may be placed on covers that bracket the Document within the aggregate, or the electronic equivalent of covers if the Document is in electronic form. Otherwise they must appear on printed covers that bracket the whole aggregate.

## 8. TRANSLATION

Translation is considered a kind of modification, so you may distribute translations of the Document under the terms of section 4. Replacing Invariant Sections with translations requires special permission from their copyright holders, but you may include translations of some or all Invariant Sections in addition to the original versions of these Invariant Sections. You may include a translation of this License, and all the license notices in the Document, and any Warranty Disclaimers, provided that you also include the original English version of this License and the original versions of those notices and disclaimers. In case of a disagreement between the translation and the original version of this License or a notice or disclaimer, the original version will prevail.

If a section in the Document is Entitled "Acknowledgements", "Dedications", or "History", the requirement (section 4) to Preserve its Title (section 1) will typically require changing the actual title.

## 9. TERMINATION

You may not copy, modify, sublicense, or distribute the Document except as expressly provided for under this License. Any other attempt to copy, modify, sublicense or distribute the Document is void, and will automatically terminate your rights under this License. However, parties who have received copies, or rights, from you under this License will not have their licenses terminated so long as such parties remain in full compliance.

## 10. FUTURE REVISIONS OF THIS LICENSE

The Free Software Foundation may publish new, revised versions of the GNU Free Documentation License from time to time. Such new versions will be similar in spirit to the present version, but may differ in detail to address new problems or concerns. See <http://www.gnu.org/copyleft/>.

Each version of the License is given a distinguishing version number. If the Document specifies that a particular numbered version of this License "or any later version" applies to it, you have the option of following the terms and conditions either of that specified version or of any later version that has been published (not as a draft) by the Free Software Foundation. If the Document does not specify a version number of this License, you may choose any version ever published (not as a draft) by the Free Software Foundation.

## 5. COMBINING DOCUMENTS

You may combine the Document with other documents released under this License, under the terms defined in section 4 above for modified versions, provided that you include in the combination all of the Invariant Sections of all of the original documents, unmodified, and list them all as Invariant Sections of your combined work in its license notice, and that you preserve all their Warranty Disclaimers.

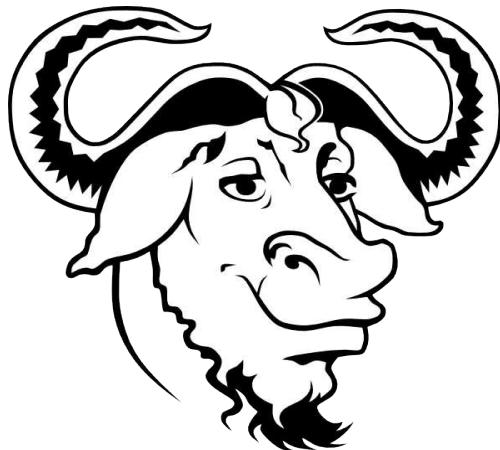
The combined work need only contain one copy of this License, and multiple identical Invariant Sections may be replaced with a single copy. If there are multiple Invariant Sections with the same name but different contents, make the title of each such section unique by adding at the end of it, in parentheses, the name of the original author or publisher of that section if known, or else a unique number. Make the same adjustment to the section titles in the list of Invariant Sections in the license notice of the combined work.

In the combination, you must combine any sections Entitled "History" in the various original documents, forming one section Entitled "History"; likewise combine any sections Entitled "Acknowledgements", and any sections Entitled "Dedications". You must delete all sections Entitled "Endorsements."

(not as a draft) by the Free Software Foundation.

# Cyfrowe Przetwarzanie Obrazów

## Digital Image Processing



Aurelio A. Heckert

Copyright (c) 2008 Cezary Bołdak

Permission is granted to copy, distribute and/or modify this document under the terms of the GNU Free Documentation License, Version 1.2 or any later version published by the Free Software Foundation; with no Invariant Sections, no Front-Cover Texts, and no Back-Cover Texts. A copy of the license is included in the section entitled "GNU Free Documentation License"



## Wykład 8

# Modele odkształcalne (*Deformable models*)



## Modele odkształcalne

- inna nazwa: **aktywny kontur** (*active contour*)
- metoda segmentacji krawędziowej
- zakłada ciągłość krawędzi, nawet jeśli w rzeczywistości jest on nieciągła
- 2 rodziny metod:
  - **parametryczny** aktywny kontur – **wąż** (*snake*)
    - Kass, M., Witkin, A. and Terzopolous, D.: *Snakes: Active Contour Models*, Int. J. Computer Vision, 1(4):321-331, 1987.
  - **geometryczny, geodezyjny** aktywny kontur (*geodesic*)
    - Caselles, V., Catte, F., Coll, T. and Dibos, F.: *A geometric model for active contours*. Numerische Mathematik, 66:1-31, 1993.
    - Malladi, R., Sethian, J.A. and Vemuri, B.C.: *Shape modeling with front propagation: A level set approach*. IEEE Trans. on Pattern Anal. Machine Intell., 17(2):158-175, 1995.
- ewolucja kompletnych konturów pod wpływem przyłożonych do nich **sił wewnętrznych i zewnętrznych**  
(inspiracja modelem fizycznym)
- ewolucja aktywnego kontury sprowadza się do procesu **minimalizacji** jego **energii**
- zastosowania:
  - segmentacja
  - śledzenie ruchu (*motion tracking*)
  - modelowanie

<http://iacl.ece.jhu.edu/projects/gvf/>

Xu, c. and Prince, J.L.: *Active Contours, Deformable Models, and Gradient Vector Flow*



## Parametryczny aktywny kontur (*snake*)

- **kontur reprezentowany jest jawnie:**

- w postaci analitycznej (równania krzywych, płaszczyzn)
- w postaci punktów-węzłów konturu

- do każdego punktu kontury przyłożone są **siły**:

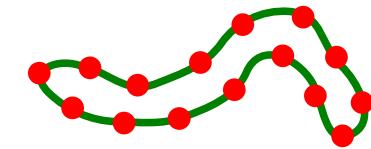
- **wewnętrzne** – ich zadaniem jest wygładzenie kształtu konturu
- **zewnętrzne** – pochodzące z otoczenia (np. gradientu obrazu)  
ich zadaniem jest dopasowanie węza do środowiska
- **ograniczeń** – rzadko używane, mają za zadanie przyciągnąć lub odrepchnąć węza do/od pewnych obszarów

- każda z sił wnosi **energię**, którą wąż stara się zminimalizować

- oryginalny wąż nie zmienia swojej topologii

- kontur może być **otwarty i zamknięty, 2D i 3D** (4D, ...)

- wąż jest **metodą lokalną** – przetwarza otoczenie (obraz) tylko w swoim lokalnym sąsiedztwie





## Oryginalny wąż Kassa

- wąż jako krzywa 2D:  $v(s) = (x(s), y(s)), s \in <0, 1>$

- postać analityczna (np. funkcje sklejane)
- postać parametryczna (punktowa)

- dla każdego  $s$  określona jest energia węża:

$$E_{\text{snake}} = \int_0^1 E_{\text{snake}}(v(s)) ds = \int_0^1 E_{\text{int}}(v(s)) ds + \int_0^1 E_{\text{image}}(v(s)) ds + \int_0^1 E_{\text{con}}(v(s)) ds$$

- $E_{\text{int}}$  – energia wewnętrzna, kontroluje jego gładkość i sztywność
- $E_{\text{image}}$  – energia obrazu, kontroluje, gdzie w otoczeniu (obrazie) wąż jest położony
- $E_{\text{con}}$  – energia związana z zewnętrznymi ograniczeniami narzuconymi przez użytkownika



## Energia wewnętrzna

$$E_{\text{int}} = \frac{\alpha(s) |v_s(s)|^2 + \beta(s) |v_{ss}(s)|^2}{2}$$

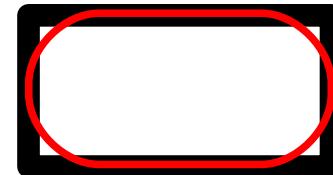
gdzie  $v_s(s) = \frac{d v(s)}{d s}$ ,  $v_{ss}(s) = \frac{d^2 v(s)}{d s^2}$

- kontroluje 2 aspekty kształtu krzywej:
  - pierwszy człon (pierwsza pochodna) sprawia, że krzywa zachowuje się jak **sznurek/guma** i nie rozciąga się (a nawet kurzy się – efekt uboczny)
  - drugi człon (druga pochodna) sprawia, że krzywa zachowuje się jak **pręt** – sprawia opory przy wyginaniu
- parametry ważace  $\alpha$  i  $\beta$  określają wpływ każdego z członów na całkowitą energię – często są stałe na całej długości węza:  $\alpha(s)=\alpha$ ,  $\beta(s)=\beta$
- pochodne liczne są w sposób analityczny lub przybliżane przez różnice skończone
- wyzerowanie (lokalne) parametru  $\alpha$  powoduje pozwolenie na formowanie nieciągłości, zmiana znaku na ujemny – powoduje premiowanie nieciągłości
- wyzerowanie (lokalne) parametru  $\beta$  pozwala na formowanie rogów, zmiana znaku na ujemny powoduje premiowanie takich rogów:

$\beta \approx 0$



$\beta \gg 0$





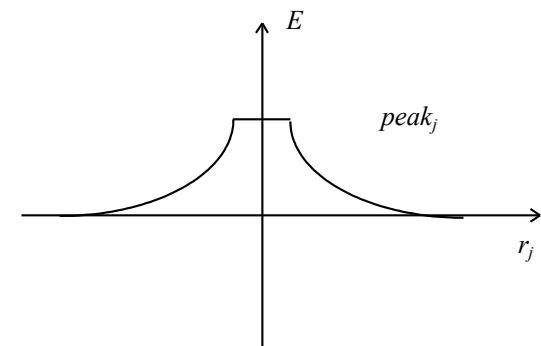
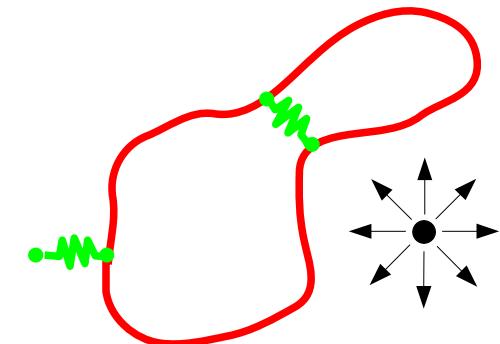
## Zewnętrzne ograniczenia

- energia ta powoduje wprowadzenie sił pochodzących z procesu wyższego rzędu (nadzorcy) który dysponuje pewną wiedzą
- Kass zaproponował 2 jej formy:
  - **energia sprężyny:** działa selektywnie na 1 punkt konturu:

$$E_{\text{spring}}^i(s=s_i) = -k^i |x_1^i, x_2^i|$$

- jeden punkt  $x_1$  jest usytuowany w stałym punkcie węża
- drugi  $x_2$  może być zamocowany do stałego punktu obrazu lub innego punktu węża
- **energia wulkanu:** odpycha węża od pewnych obszarów otoczenia:

$$E_{\text{volcano}}^i(s) = \min(\text{peak}_i, \frac{1}{|\nu(s), \text{volcano}^i|})$$





## Energia zewnętrzna obrazu

$$E_{\text{image}} = w_{\text{line}} E_{\text{line}} + w_{\text{edge}} E_{\text{edge}} + w_{\text{term}} E_{\text{term}}$$

- energia ta sprawia, że wąż przemieszcza się w kierunku tych obszarów, które mają pewne interesujące cechy/właściwości:
  - w kierunku **linii**:

$$E_{\text{line}}(s) = I(x(s), y(s))$$

wąż dąży do ciemnych ( $w_{\text{line}} > 0$ ) lub jasnych ( $w_{\text{line}} < 0$ ) rejony obrazu

- w kierunku **krawędzi**:

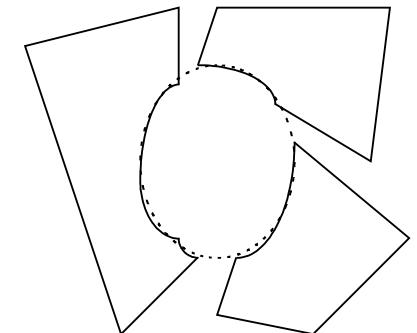
$$E_{\text{edge}}(s) = -|\nabla I(x(s), y(s))|^2$$

wąż dąży do obszarów o wysokim gradiencie

- w kierunku **zakończeń linii**:

$$C_x = \frac{\partial I}{\partial x}, \quad C_y = \frac{\partial I}{\partial y}$$
$$C_{xx} = \frac{\partial^2 I}{\partial x^2}, \quad C_{yy} = \frac{\partial^2 I}{\partial y^2}, \quad C_{xy} = \frac{\partial^2 I}{\partial x \partial y}$$
$$E_{\text{term}} = \frac{C_{yy} C_x^2 - 2 C C_{xy} C_x C_y + C_{xx} C_y^2}{(C_x^2 + C_y^2)^{\frac{3}{2}}}$$

- wpływ każdej z tych energii (zachowanie węża) można kontrolować przez dobór współczynników  $w$
- zwykle energia wewnętrzna obrazu wyliczana jest z obrazu wygładzonego filtrem Gaussa



# Numeryczne rozwiązanie problemu minimalizacji energii



- jeśli oznaczyć  $E_{\text{ext}} = E_{\text{image}} + E_{\text{con}}$  to rozwiązanie  $\mathbf{v}^*$  minimalizujące energię kontury spełnia równanie **Eulera-Lagrange'a** (równowaga sił):

$$-(\alpha \mathbf{v}')' + (\beta \mathbf{v}''')''' + \nabla E_{\text{ext}}(\mathbf{v}) = 0$$

- równanie to można łatwo rozwiązać numerycznie:

- przechodząc do dyskretnej formulacji węza (wąż złożony z punktów)
- zastępując pochodne różnicami skończonymi z krokiem  $\Delta x = \Delta y = 1$
- traktując minimalizację jako ewolucję krzywej  $\mathbf{v}$ ,  $E_{\text{snake}} = \sum_{i=0}^n (E_{\text{int}}(i) + E_{\text{ext}}(i))$  której pochodna po czasie dąży do zera
- rozdzielając równania po zmiennej  $x$  i  $y$   $f_x(i) = \frac{\partial E_{\text{ext}}}{\partial x}(i), f_y(i) = \frac{\partial E_{\text{ext}}}{\partial y}(i)$
- $v(0)=v(N)$  – wąż zamknięty

$$\begin{aligned} \alpha_i \cdot (v_i - v_{i-1}) - \alpha_{i+1} \cdot (v_{i+1} - v_i) + \beta_{i-1} \cdot (v_{i-2} - 2v_{i-1} + v_i) - 2\beta_i \cdot (v_{i-1} - 2v_i + v_{i+1}) + \\ \beta_{i+1} \cdot (v_i - 2v_{i+1} + v_{i+2}) + (f_x(i), f_y(i)) = 0 \end{aligned}$$

lub w formie macierzowej:

$$A_{(N \times N)} V + F = 0$$

$$A_{(N \times N)} = \begin{bmatrix} c_1 & d_1 & e_1 & 0 & \dots & 0 & a_1 & b_1 \\ b_2 & c_2 & d_2 & e_2 & 0 & \dots & 0 & a_2 \\ \dots & \dots \\ d_N & e_N & 0 & \dots & 0 & a_N & b_N & c_n \end{bmatrix} \quad \begin{aligned} a_i &= \beta_{i-1}, & b_i &= -2\beta_i - 2\beta_{i-1} - \alpha_i, & c_i &= \beta_{i+1} + 4\beta_i + \beta_{i-1} + \alpha_{i+1} + \alpha_i, \\ d_i &= -2\beta_{i+1} - 2\beta_i - \alpha_{i+1}, & e_i &= \beta_{i+1} \end{aligned}$$

# Numeryczne rozwiązanie problemu minimalizacji energii



- aby rozwiązać równanie za jego prawą stronę podstawiamy pochodną po czasie aproksymowaną różnicą skońzoną z krokiem  $\gamma$ :

$$A_{(N \times N)} x_t + f_x(x_{t-1}, y_{t-1}) = -\gamma \cdot (x_t - x_{t-1})$$

$$A_{(N \times N)} y_t + f_y(x_{t-1}, y_{t-1}) = -\gamma \cdot (y_t - y_{t-1})$$

po przekształceniu:

$$x_t = (A_{(N \times N)} + \gamma I_{(N \times N)})^{-1} (\gamma \cdot x_{t-1} - f_x(x_{t-1}, y_{t-1}))$$

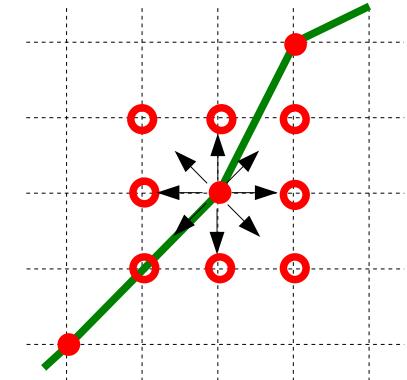
$$y_t = (A_{(N \times N)} + \gamma I_{(N \times N)})^{-1} (\gamma \cdot y_{t-1} - f_y(x_{t-1}, y_{t-1}))$$

- macierz  $(A_{(N \times N)} + \gamma I_{(N \times N)})$  jest macierzą pentagonalną, niezmienną w czasie całej ewolucji – można ją odwrócić przez dekompozycję **LU** w czasie  $O(N)$  na początku procesu
- proces zatrzymujemy, kiedy prawdziwy jest warunek stopu:
  - $x$  i  $y$  przestają się zmieniać (uwaga na oscylacje)
  - energia  $E$  przestaje (szybko) maleć
  - wykona się zadana z góry liczba kroków iteracji
  - proces wyższego rzędu (np. użytkownik obserwujący wyniki pośrednie) zakończy minimalizację

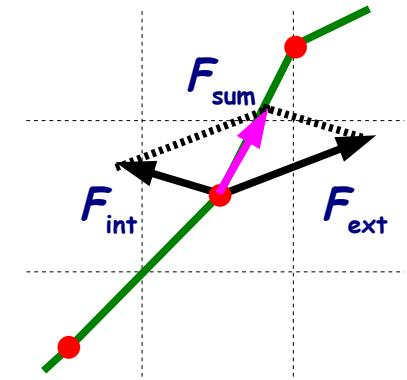
# Konkurencyjne (lokalne) sposoby minimalizacji energii



- podejście oparte na szukaniu położen o mniejszej energii
  - dla każdego **punktu węza położonego na pikselu obrazu** rozpatruje się jego sąsiedztwo
  - dla każdego z sąsiadów liczy się energię, którą miałby ten punkt gdyby się tam znalazł
  - jeśli któryś z sąsiadów wykazuje mniejszą energię, punkt konturu przemieszcza się tam



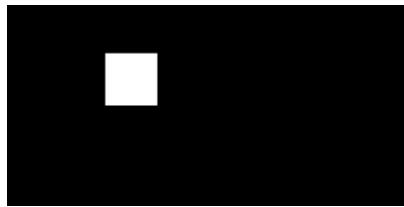
- podejście oparte na wyliczaniu sił i przemieszczaniu punktów w kierunku siły wypadkowej
  - dla każdego punktu liczone są wszystkie siły na niego działające
  - sumowane są one wektorowo aby wyliczyć siłę wypadkową
  - punkt przesuwany jest z pewnym krokiem w kierunku tej siły
  - położenie punktów może nie pokrywać się z pikselami, wartości określone tylko w punktach siatki interpoluje się



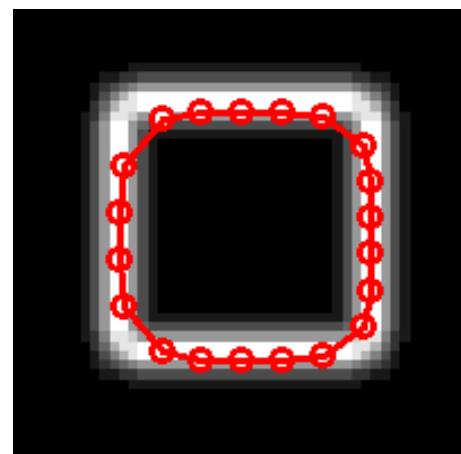


## Wpływ parametru $\alpha$

obraz oryginalny

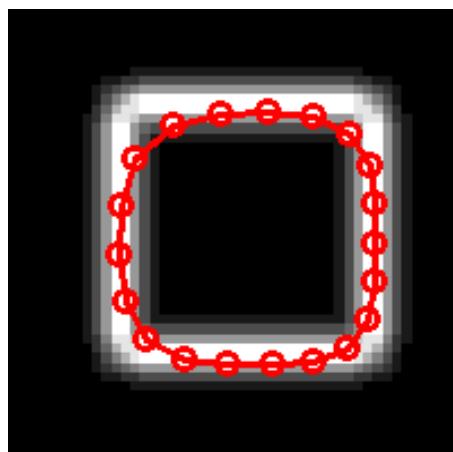


kontur początkowy  
(20 punktów)  
na tle obrazu  
amplitudy gradientu

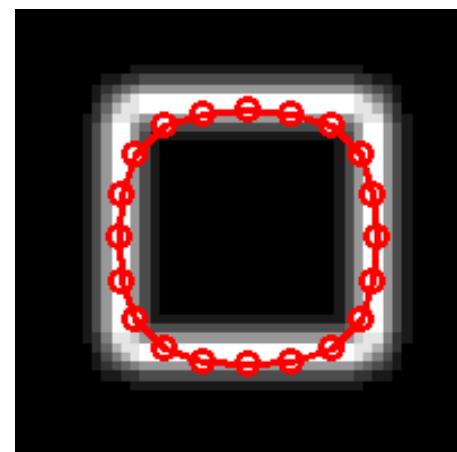


- parametr ten jest odpowiedzialny za regularny kształt konturu
- wprowadza siłę, która przyciąga do siebie punkty konturu
- efektem ubocznym jest tendencja do kurczenia się węza
- jeśli waga ta jest zbyt duża w stosunku do siły obrazu, wąż „odrywa” się od krawędzi
- alternatywne formy (odstępstwo odległości między punktami od średniej)

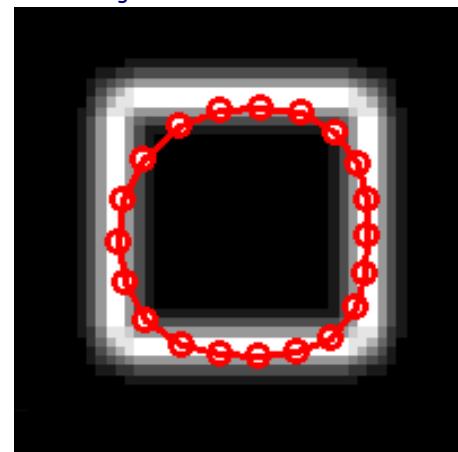
kontur po 30 iteracjach,  $\beta=0.3$ ,  $w_{edge}=0.5$ ,  $\gamma=1$



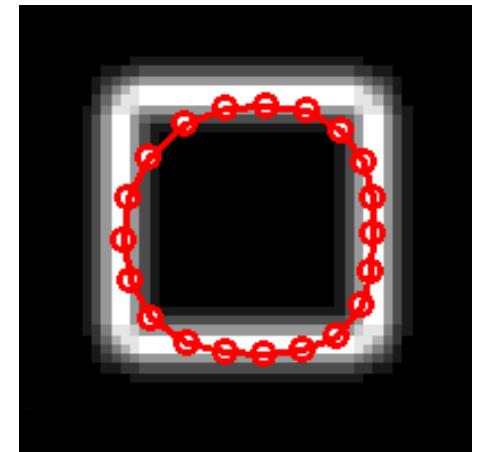
$\alpha=0$



$\alpha=0.2$



$\alpha=0.5$



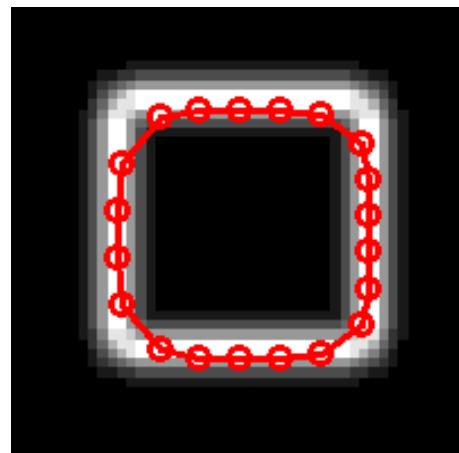
$\alpha=1.0$



## Wpływ parametru $\beta$

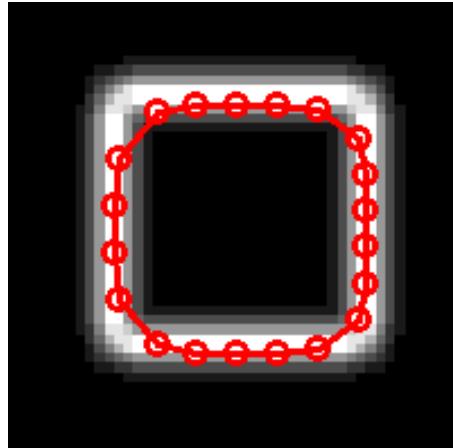
kontur początkowy (20 punktów)  
na tle obrazu amplitudy gradientu

obraz oryginalny

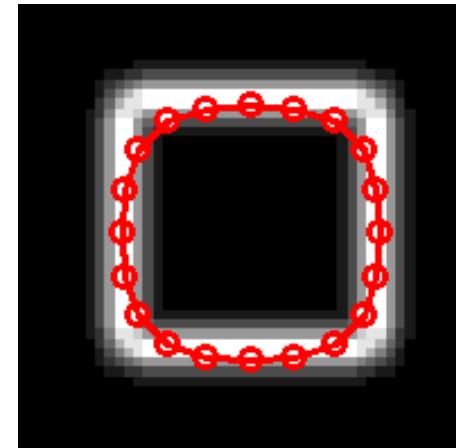


- parametr ten jest odpowiedzialny  
gładki kształt konturu
- utrudnia tworzenie się rogów
- w przypadku konturu zamkniętego  
dąży do kształtu koła

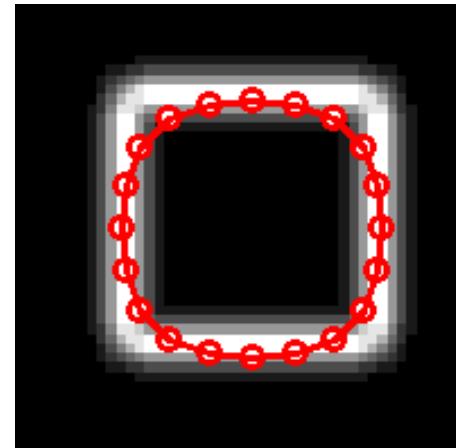
kontur wynikowy,  $\alpha=0.2$ ,  $w_{edge}=0.5$ ,  $\gamma=1$



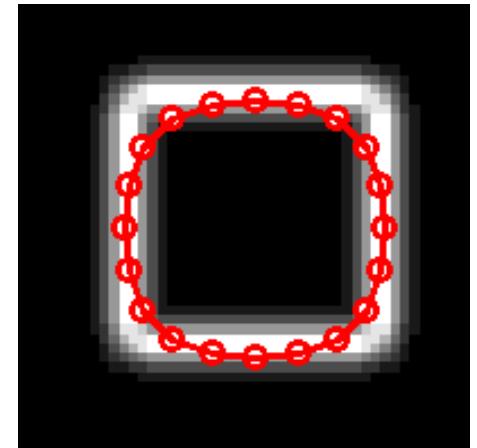
$\beta=0$



$\beta=0.3$



$\beta=0.7$

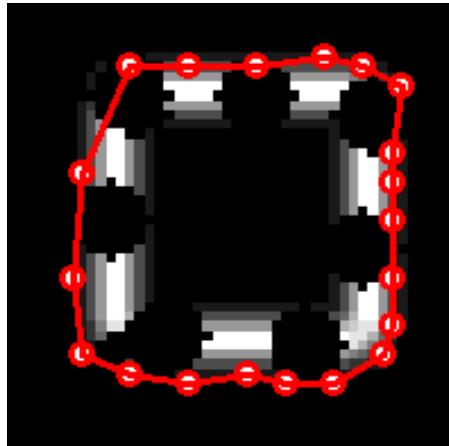


$\beta=2.0$



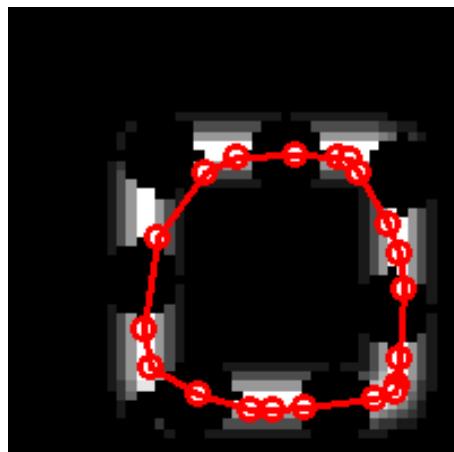
## Rekonstrukcja kształtów

kontur początkowy (20 punktów)  
na tle obrazu wejściowego

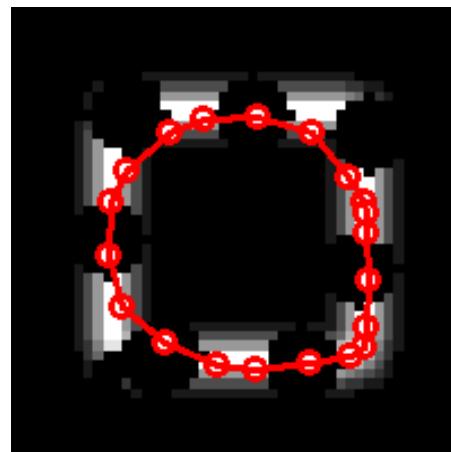


- jeśli obiekt jest nieciągły, aktywny kontur wciąż może go wykryć
- użyta jest energia wewnętrzna odpowiedzialna za wykrywanie linii ( $E_{line}$ )

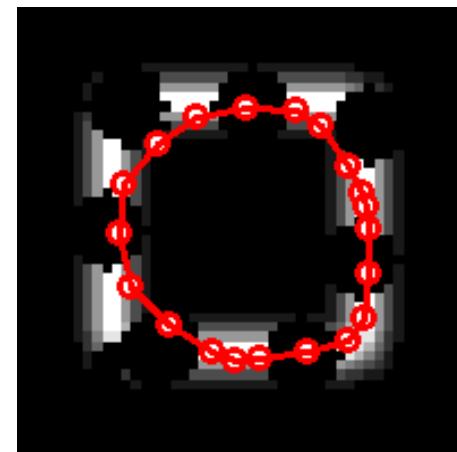
kontur wynikowy,  $\beta=0.3$ ,  $w_{line}=2$ ,  $\gamma=1$



$\alpha=0.2$

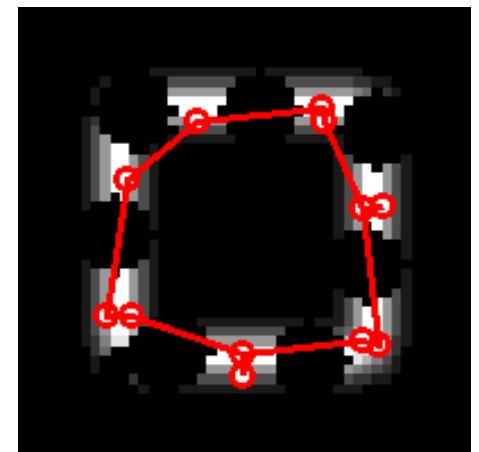


$\alpha=0.3$



$\alpha=0.4$

$\alpha=0$ ,  $\beta=0$ ,  $w_{line}=2$ ,  $\gamma=1$





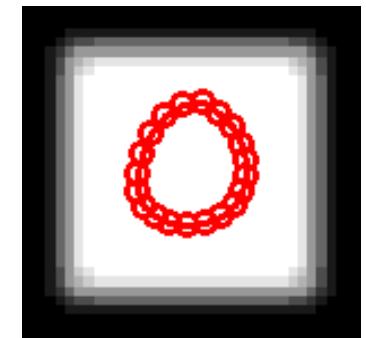
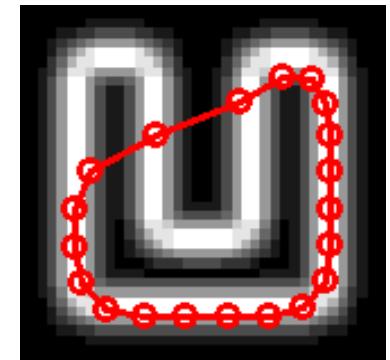
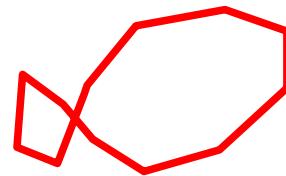
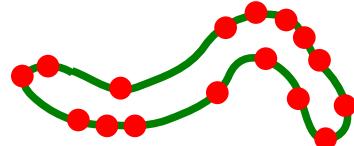
## Zalety bazowego aktywnego konturu

- jako **wynik** zwraca **ciągłe obiekty**, pomimo potencjalnych przerw w krawędziach
- może współpracować z procesem wyższego rzędu w celu:
  - wniesienia wiedzy *a priori* (energia ograniczeń)
  - interaktywnego nadzoru nad procesem minimalizacji:
    - ręczna korekcja położenia konturu
    - dynamiczna zmiana parametrów metody
- **metoda lokalna**, co przyspiesza obliczenia
- wyodrębnia **obiekty o dowolnym kształcie**
- dowolność w określaniu formy konturu 2D i 3D (punkty, trójkąty, funkcje sklejane, B-spline, elementy skończone, ...)
- baza do licznych modyfikacji
- parametry mają fizyczną interpretację



## Wady i wyzwania

- **wrażliwość na dobór parametrów**, różne obrazy wymagają różnych ich wartości, wartości parametrów oddziałują na siebie
- **wrażliwość na początkowe położenie konturu** – lokalizuje najbliższe minimum lokalne
- **wąż „widzi” tylko najbliższe obiekty**, rozmycie obrazu wejściowego pozwala rozszerzyć zakres działania (widoczności) węża, ale wciąż jest to rząd pikseli ⇒ **konieczność bliskiej inicjalizacji**
- **zależność energii od skali**
- **tendencja do zmniejszania rozmiarów**
- **trudności z wyodrębnianiem kształtów wklęsłych**
- **trudne zmiany w topologii**
- **problemy z formą konturu** – konieczność reformulacji
  - nierównomierne rozmieszczenie punktów
  - formowanie pętli





## Aktywny kontur w śledzeniu ruchu

- aktywny kontur idealnie nadaje się do śledzenia obiektu w ruchu
- warunkiem są niewielkie zmiany położenia i kształtu obiektu na kolejnych klatkach filmu
- kontur wynikowy na bieżącej klatce staje się konturem wynikowym na klatce kolejnej
- jedynie pierwsza klatka wymaga oddzielnej inicjalizacji
- zmiany z klatki na klatkę są niewielkie ⇒ kolejne kontury nie różnią, więc czas segmentacji jest zwykle niewielki
- (film)



## Wąż - balon

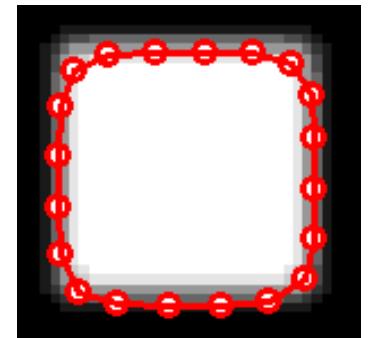
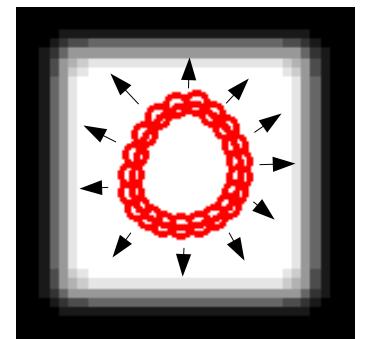
- aby zniwelować efekt kurczenia się węża można wprowadzić dodatkową siłę skierowaną na zewnątrz węża
- zachowuje się on wtedy jak pompujący się balon

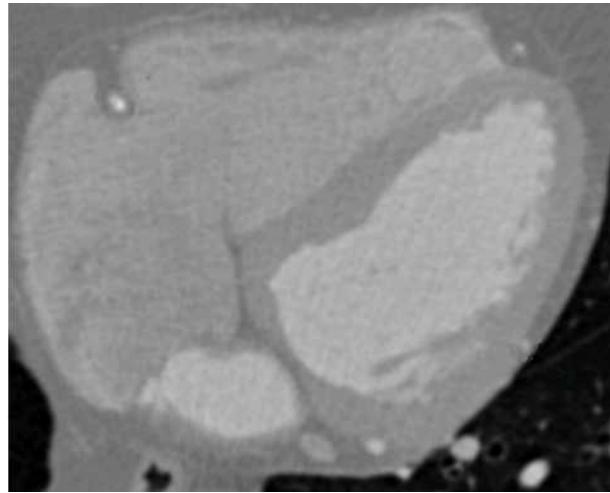
$$F_{\text{balloon}}(s) = k \cdot n(v(s))$$

$k$  – współczynnik wagi,

$n(v(s))$  – wektor normalny do konturu

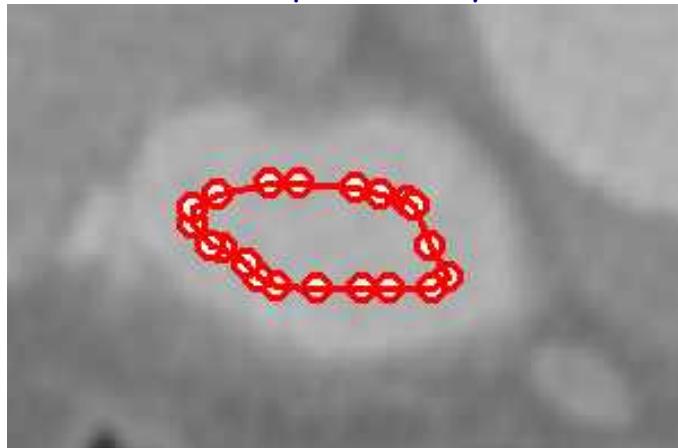
- dodatkowy parametr, ale znacznie ulepsza działania węża w przypadku inicjacji wewnętrz obiektu
- waga  $k$  musi być dobrze dobrana
  - za mała nie pozwoli na wzrost węża
  - za duża nie pozwoli na zatrzymanie się na krawędzi
- Cohen,L.D. and Cohen, I.: *Finite-Element Method for Active Contour Models and Balloons for 2D and 3D Images*, IEEE Transactions on Pattern Analysis and Machine Intelligence, 15(11): 1131-1147, 1993



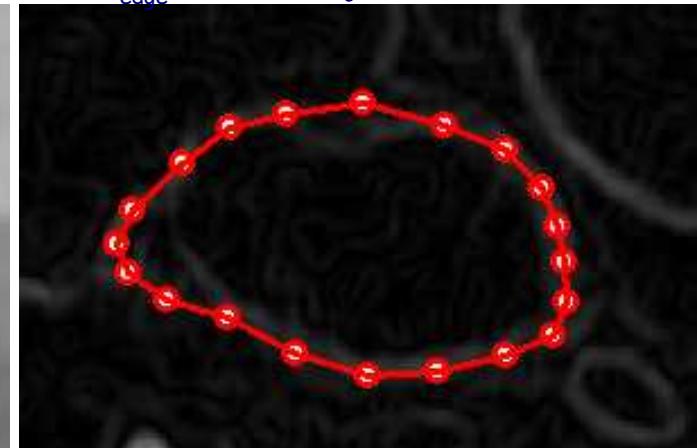
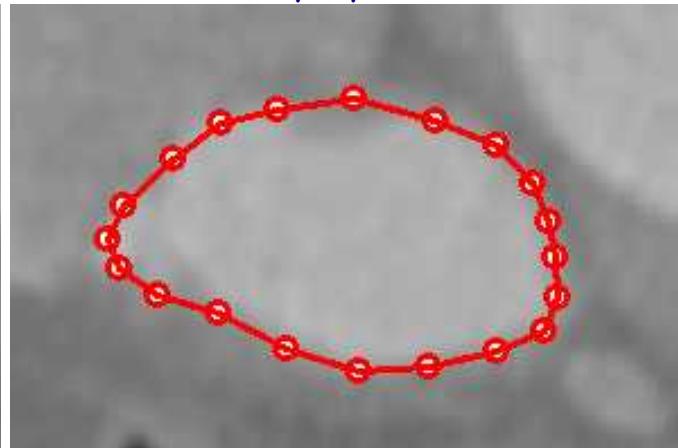


## Waż balon i obraz rzeczywisty

kontur początkowy



kontury wynikowe:  $\alpha=0.05$ ,  $\beta=0.01$ ,  $w_{edge}=5$ ,  $k=0.5$ ,  $\gamma=1$



- kontur początkowy może być odległy od granic obiektu

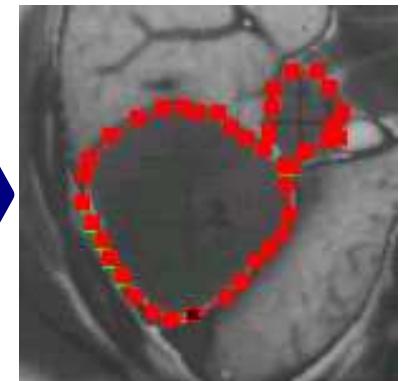
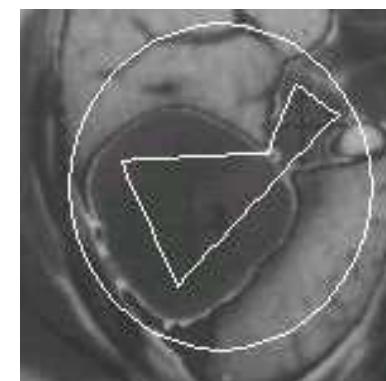
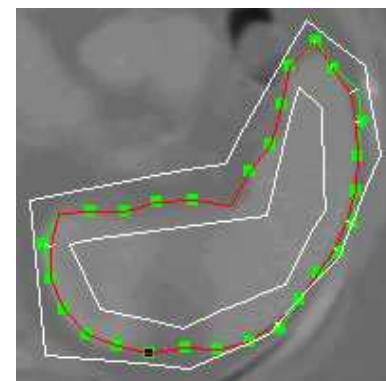
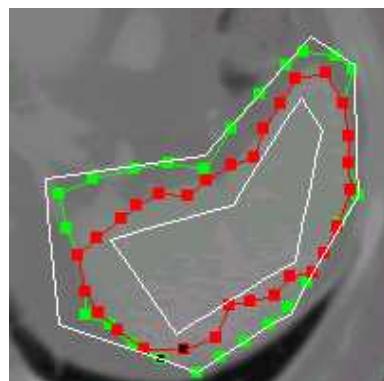


## Podwójny aktywny kontur

- ponieważ między początkową pozycją konturu a krawędzią obiektu mogą się znajdować zakłócenia, wąż może się zatrzymać przed dotarciem do tego obiektu
- lokalnie zakłócenia mogą być większe i mniejsze, co dodatkowo utrudnia dobór parametrów
- można zainicjować **2 kontury**: jeden **na zewnątrz**, a drugi **wewnątrz**

Gunn, S.R. and Nixon, M.S.: *A Robust Snake Implementation; A Dual Active Contour*, IEEE Transactions on Pattern Analysis and Machine Intelligence, 19(1):63 - 68, 1997

- zewnętrzny** kontur **kurczy się**, **wewnętrzny rozszerza się**
- jeśli oba zatrzymają się przed połączeniem się, ten o wyższej energii uzyskuje dodatkową siłę kierującą go ku konturowi o niższej energii

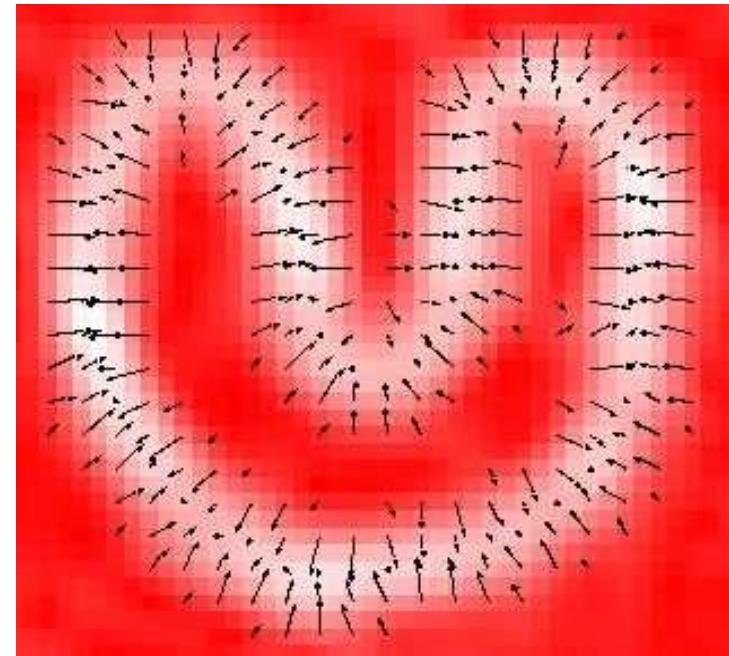




## Lokalność gradientu

- oryginalny wąż ma problem z wyodrębnianiem kształtów wklęszych
- ponadto „nie widzi” odległych obiektów
- **gradient jest lokalny w stosunku do krawędzi !!!**  
z dala od krawędzi, na jednolitych obszarach, jest zerowy

<http://iacl.ece.jhu.edu/projects/gvf/>



Xu, C. and Prince, J.L.:

*Active Contours, Deformable Models, and Gradient Vector Flow*

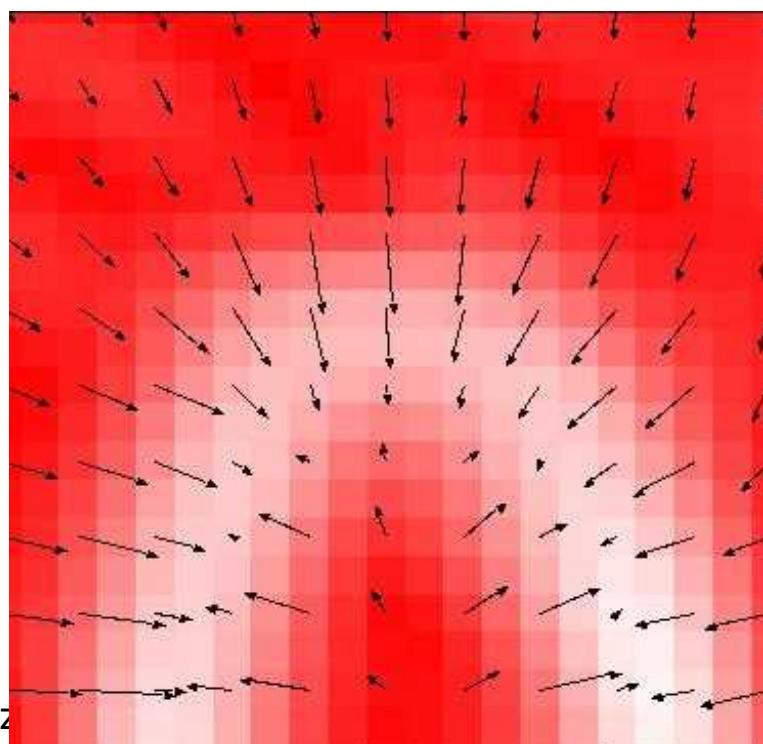
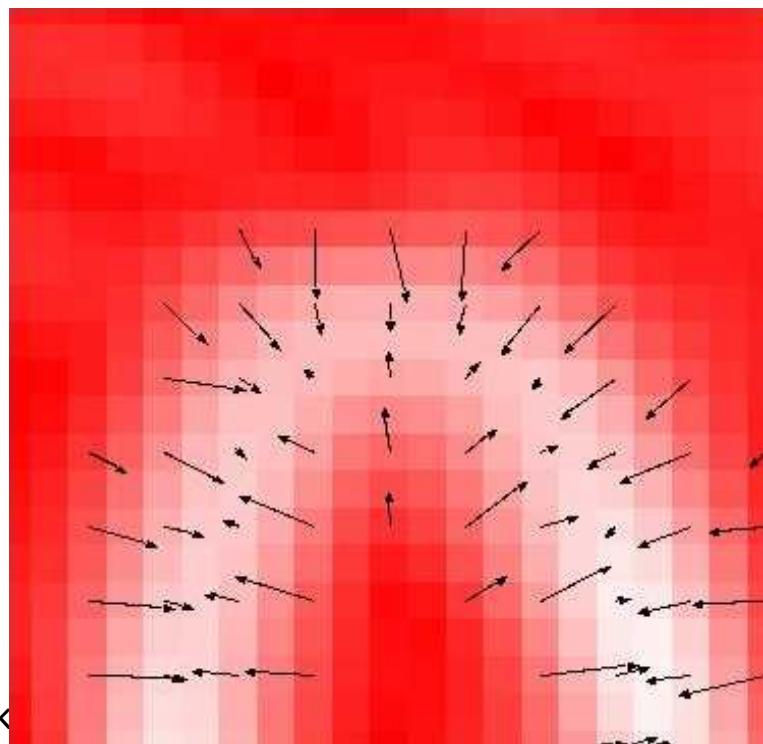
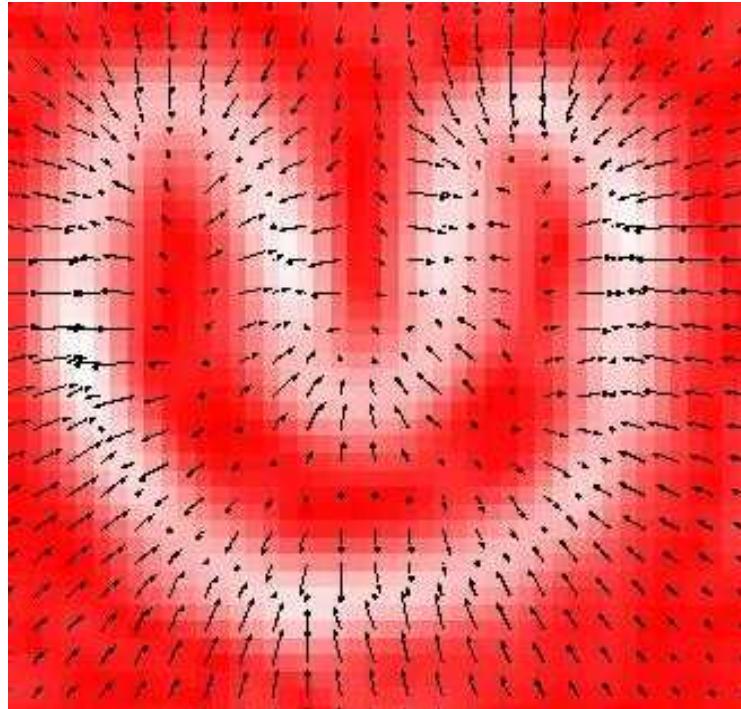
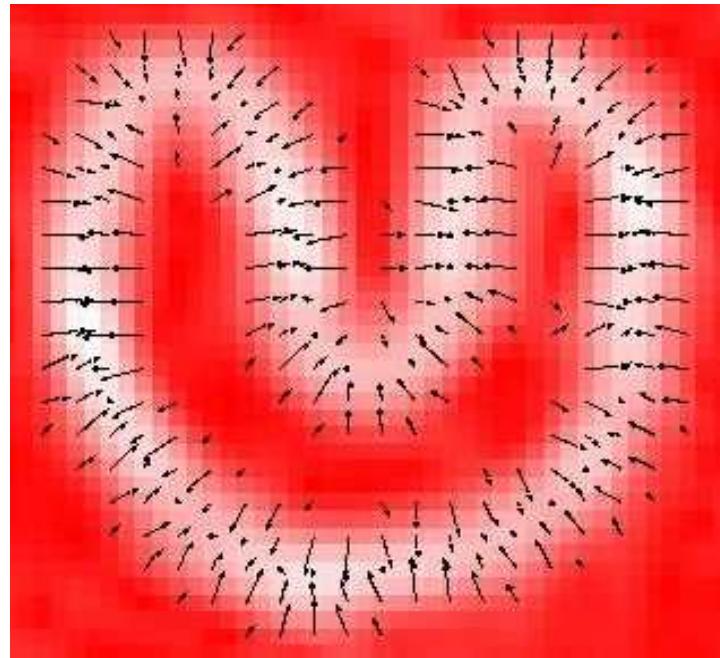


## Rozszerzenie zasięgu gradientu **(GVF - Gradient Vector Flow)**

- aktywny kontur ewoluje zgodnie z polem wektorowym  $\nabla f = \nabla E_{\text{ext}}$
- zwykle jest to gradient amplitudy gradientu obrazu  $\nabla |\nabla I|$  który jest lokalny w stosunku do krawędzi
- **GVF** jest podobnym polem wektorowym  $\mathbf{V}(x,y) = (u(x,y),v(x,y))$ , które:
  - zdefiniowane jest na podstawie obrazu i jego gradientu,
  - „propaguje” stare pole wektorowe na obszary gdzie nie jest ono zdefiniowane wzduż wektorów tego pola
  - zastępuje oryginalne pole wektorowe w procedurze ewolucji węża
  - minimalizuje następująco sformułowaną energię:

$$E_{\text{GVF}} = \int \int \mu(u_x^2 + u_y^2 + v_x^2 + v_y^2) + |\nabla f|^2 |\mathbf{V} - \nabla f|^2 dx dy$$

- w obszarach, gdzie gradient gradientu ( $|\nabla f|$ ) jest wysoki głównym minimalizowanym czynnikiem będzie  $|\mathbf{V} - \nabla f|^2$  co premiuje pole  $\mathbf{V}$  zgodne z polem oryginalnym  $\nabla f$
- w pozostałych obszarach ( $|\nabla f| \rightarrow 0$ ) minimalizowane są pochodne kierunkowe  $\mathbf{V}$  co premiuje propagowanie stałych kierunków pola zgodnie z kierunkami jego wektorów
- współczynnik  $\mu$  kontroluje zgodność nowego i starego pola wektorowego
- Xu, C. and Prince, J.L.: *Snakes, Shapes, and Gradient Vector Flow*, IEEE Transactions on Image Processing, 7(3):359-369, 1998.



Wyk

Przetwarz



## Rozwiążanie numeryczne GVF

- rozwiążanie minimalizujące energię  $E_{\text{GVF}}$  wyliczamy z równania Eulera:

$$\mu \nabla^2 \mathbf{V} - (\mathbf{V} - \begin{bmatrix} f_x \\ f_y \end{bmatrix}) \cdot (f_x^2 + f_y^2) = 0 \quad \text{dla } \mathbf{V} = \begin{bmatrix} u \\ v \end{bmatrix}$$

- zastępując prawą stronę równań pochodną pola po czasie:

$$\mathbf{V}_t(x, y, t) = \mu \nabla^2 \mathbf{V}(x, y, t) - b(x, y) \cdot \mathbf{V}(x, y, t) + \begin{bmatrix} c^1(x, y) \\ c^2(x, y) \end{bmatrix}$$

$$b(x, y) = f_x(x, y)^2 + f_y(x, y)^2, \quad c^1(x, y) = b(x, y) \cdot f_x(x, y), \quad c^2(x, y) = b(x, y) \cdot f_y(x, y)$$

pochodne po czasie:

$$\mathbf{V}_t(x, y) = \frac{1}{\Delta t} (\mathbf{V}(x, y, t+1) - \mathbf{V}(x, y, t))$$

laplasjany :

$$\nabla^2 \mathbf{V}(x, y) = \frac{1}{\Delta x \Delta y} (\mathbf{V}(x+1, y) + \mathbf{V}(x-1, y) + \mathbf{V}(x, y+1) + \mathbf{V}(x, y-1) - 4 \mathbf{V}(x, y))$$

- mamy:  $\mathbf{V}(x, y, t+1) = (1 - b(x, y) \Delta t) \cdot \mathbf{V}(x, y, t) +$

$$+ r \cdot (\mathbf{V}(x+1, y) + \mathbf{V}(x-1, y) + \mathbf{V}(x, y+1) + \mathbf{V}(x, y-1) - 4 \mathbf{V}(x, y)) + \begin{bmatrix} c^1(x, y) \\ c^2(x, y) \end{bmatrix}$$

gdzie:  $r = \frac{\mu \Delta t}{\delta x \delta y}$  a  $\Delta t \leq \frac{\Delta x \Delta y}{4 \mu}$  (warunek stabilności)



## Przykłady segmentacji z GVF

<http://iacl.ece.jhu.edu/projects/gvf/>

Xu, C. and Prince, J.L.: *Active Contours, Deformable Models, and Gradient Vector Flow*

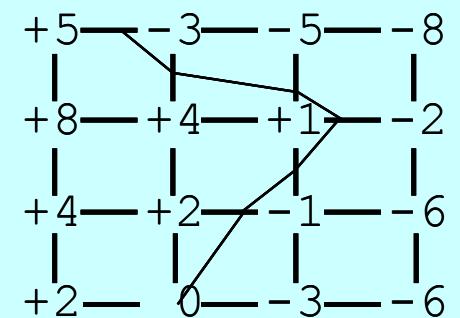


## Geodezyjny aktywny kontur - „Level Sets”

- w tej gałęzi modeli odkształcalnych **kontur** nie jest reprezentowany jawnie w postaci punktów, ale **niejawnie** w postaci funkcji  $\Phi$  o wymiarze o jeden wyższym niż wymiar problemu
- dokładniej jest to **poziom zerowy** tej funkcji (*level set zero*) określany też jako front  $\Gamma$

$$\Gamma(t) = \{X \in \Re^m \mid \phi(X, t) = 0\}$$

- w przypadku obrazu cyfrowego 2D front ten jest linią a funkcja  $\Phi$  jest określona we wszystkich pikselach obrazu
- funkcja  $\Phi$  jest zwykle **funkcją odległości od konturu ze znakiem** – w środku ujemna, na zewnątrz dodatnia
- nie jest tu narzucony żaden warunek co do kształtu i topologii konturu, możliwe są dowolne konfiguracje





## Ewolucja konturu

- ewolucji nie podlega sam kontur  $\Gamma$ , ale cała funkcja  $\Phi$
- ewolucja odbywa się zgodnie z równaniem różniczkowym Hamiltona-Jacobiego:

$$\phi_t + F |\nabla \phi| = 0, \quad \phi(X, 0) \text{ dane}$$

$\phi_t$  – pochodna po czasie,

$\nabla \phi$  – gradient funkcji

$F$  – prędkość ewolucji poziomów funkcji  $\phi$  (w tym frontu  $\Gamma$  – poziomu zerowego) w kierunku prostopadłym tych poziomów (poziomic)

- ewolucję funkcji może realizować zastępując pochodne: czasu i przestrzenną (gradient), różnicami skończonymi:

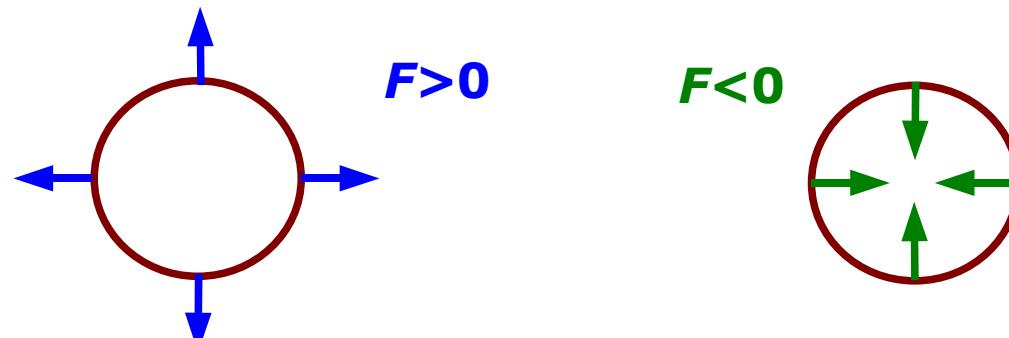
$$\phi_{ij}^{n+1} = \phi_{ij}^n - \Delta t \cdot F |\nabla \phi_{ij}^n|$$

- w praktyce liczenie gradientu jest bardziej skomplikowane: opierając się na mechanice płynów, aby wziąć pod uwagę nieciągłości funkcji, wybiera się między różnicą skońzoną lewą a prawą (górną a dolną) w zależności od prędkości ewolucji poziomic  $F$



## Prędkość ewolucji

- prędkość ewolucji  $F$  w kierunku prostopadłym do poziomie jest kluczowa do określenia zachowania się kontury
- stała prędkość większa od zera powoduje rozszerzanie się konturu  
stała prędkość mniejsza od zera powoduje kurczenie się konturu



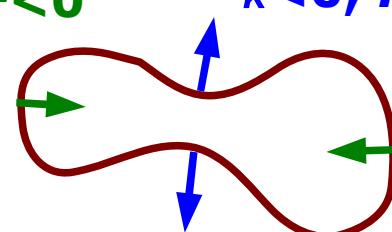
- prędkość ta może się równać ujemnej krzywiźnie (*curvature*)

$$F = -\kappa$$

wtedy każda krzywa (front początkowy) ewoluje do okręgu a następnie do punktu (znika)

$$\kappa > 0, F < 0 \quad \kappa < 0, F > 0$$

- ma to efekt wygładzania krzywej !!!





## Prędkość ewolucji w segmentacji

- w przetwarzaniu obrazów modyfikuje się równanie ewolucji, aby wziąć pod uwagę wpływ obrazu na położenie konturu:

$$\phi_t + F_s(1 - \epsilon \kappa) |\nabla \phi| - \beta \nabla P \cdot \nabla \phi = 0$$

$$F_s = \frac{1}{1 + |\nabla I|^2}$$

prędkość ewolucji ma spaść do zera kiedy gradient jest wysoki (krawędź)

$$P = -|\nabla(G_\sigma * \nabla I(x))|$$

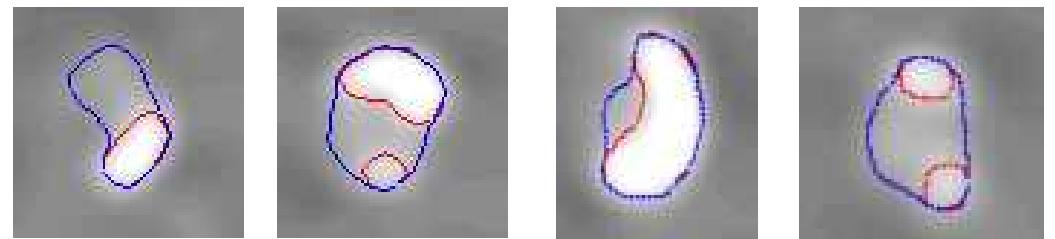
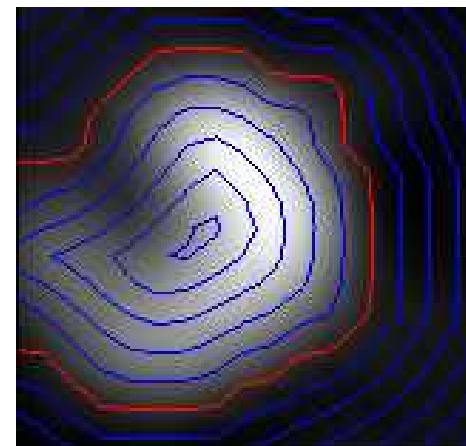
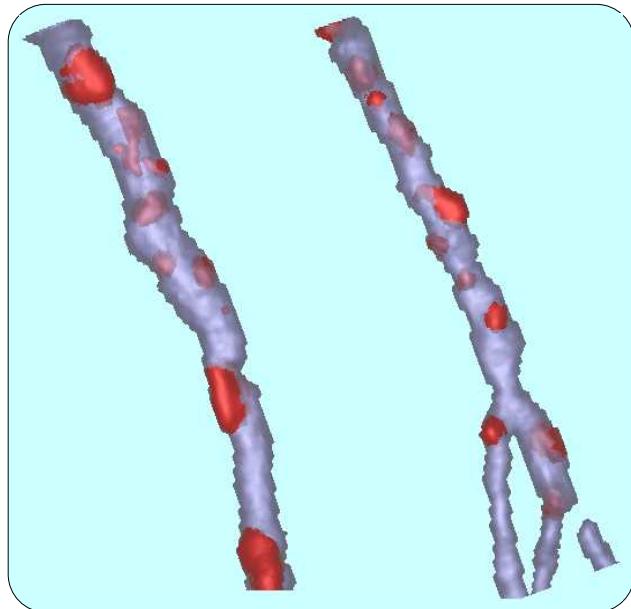
dodatkowe pole wektorowe (gradient gradientu) ma kierować węza w pobliże krawędzi

- bez obecności gradientu wąż się rozszerza (prędkość 1)
- cały czas działa czynnik wygładzający (krzywizna  $\kappa$ )
- w pobliżu krawędzi „wkracza do akcji” dodatkowe pole wektorowe  $\mathbf{P}$
- na krawędzi prędkość ewolucji ulega wyhamowaniu –  $F_s$



## Segmentacja obrazów

- krzywa konturu dopasowuje się do kształtu obiektu
- radzi sobie z formami wklęsłymi
- 2D i 3D
- metoda globalna – dosyć wolna





## Proces segmentacji

- dane wejściowe: obraz  $\mathbf{I}$  i kontur początkowy  $\Gamma_0$
- inicjalizacja  $\Phi$  jako odległość ze znakiem do  $\Gamma_0$
- ewolucja funkcji  $\Phi$  zgodnie z przyjętym równaniem/prędkością ewolucji (czyli rozwiązywanie równania różniczkowego)
- warunek stopu:
  - niezmienność funkcji  $\Phi$
  - wykonanie założonej liczby kroków iteracji
  - zatrzymanie ewolucji przez proces wyższego rzędu (operator oceniający na bieżąco wyniki segmentacji\*)
- dopiero na końcu – wyliczenie ostatecznej formy kontury jako poziomu zerowego funkcji  $\Phi$



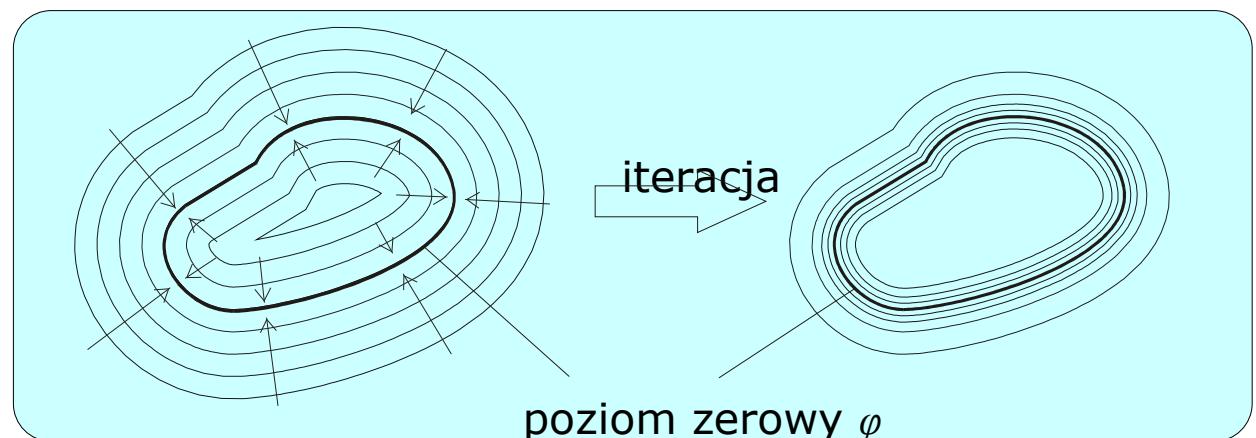
## Inicjalizacja i reinicjalizacja funkcji odległości

- inicjalizacja polega na podaniu odległości ze znakiem do konturu początkowego  $\Gamma_0$
- można to osiągnąć w czasie  $O(N)$  poprzez wykorzystanie algorytmu **Fast Marching**, które jest uproszczeniem równania Hamiltona-Jacobiego do przypadku stacjonarnego ( $F>0$ , front tylko się rozszerza lub tylko się cofa, przechodząc przez każdy piksel tylko raz):

$$|\nabla T| F = 1, \text{ dla } F = 1$$

- ponadto okresowo trzeba funkcji  $\Phi$  przywrócić charakter funkcji odległości, którą traci poprzez zbliżanie się innych poziomów do poziomu

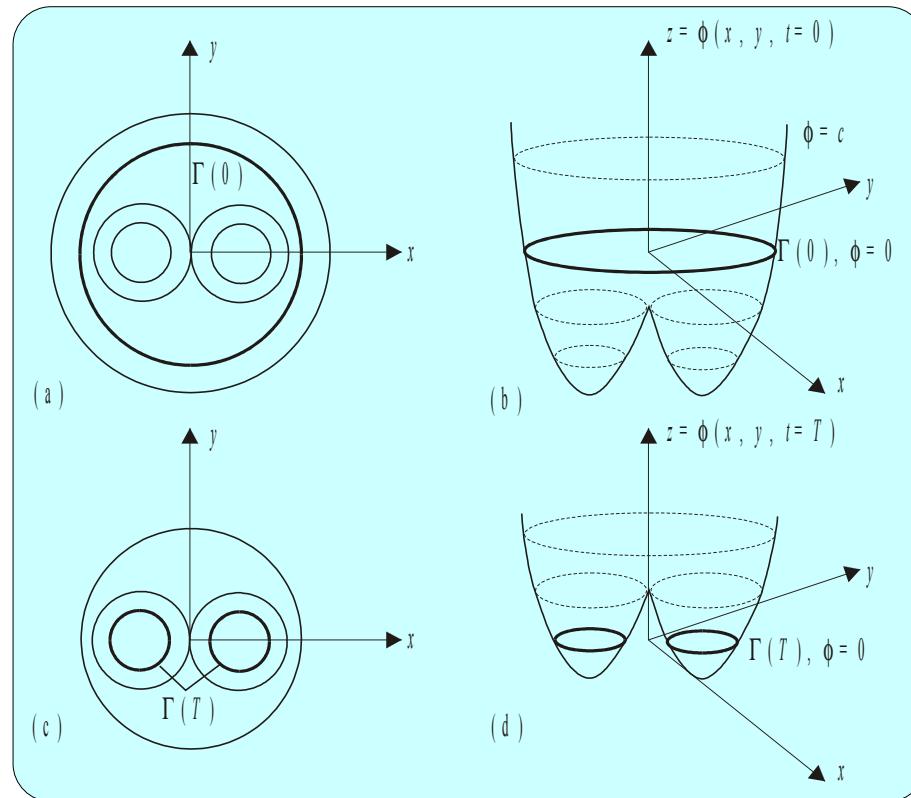
- wykonuje się to przez wyliczenie bieżącej pozycji konturu i ponowne zastosowanie algorytmu Fast Marching





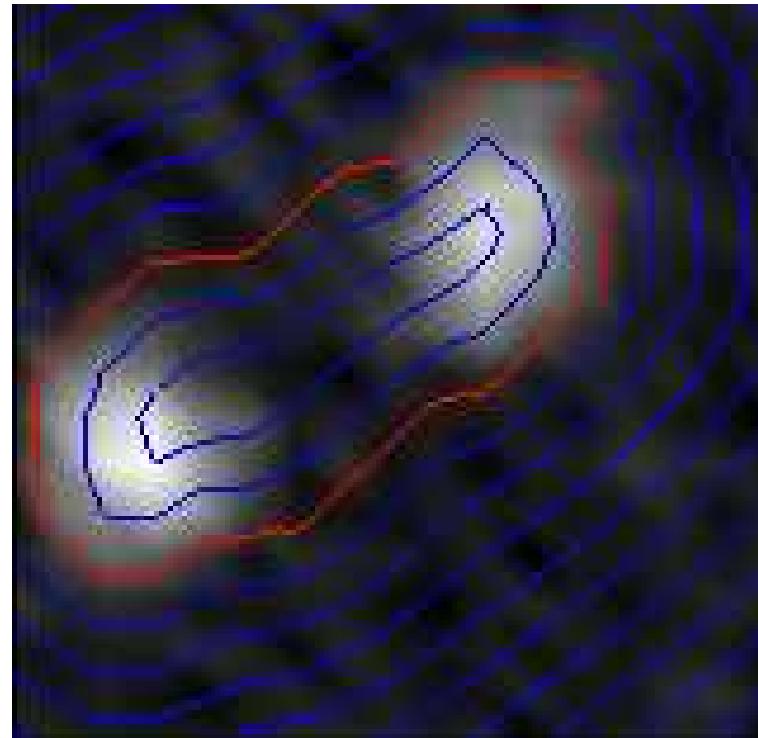
## Ewolucja konturu - zmiany topologii

- **zmiany w topologii** widziane są jako pojawianie się, łączenie, zanikanie grup ujemnych wartości  $\Phi$
- jest to naturalny proces, zarówno w 2D jak i 3D



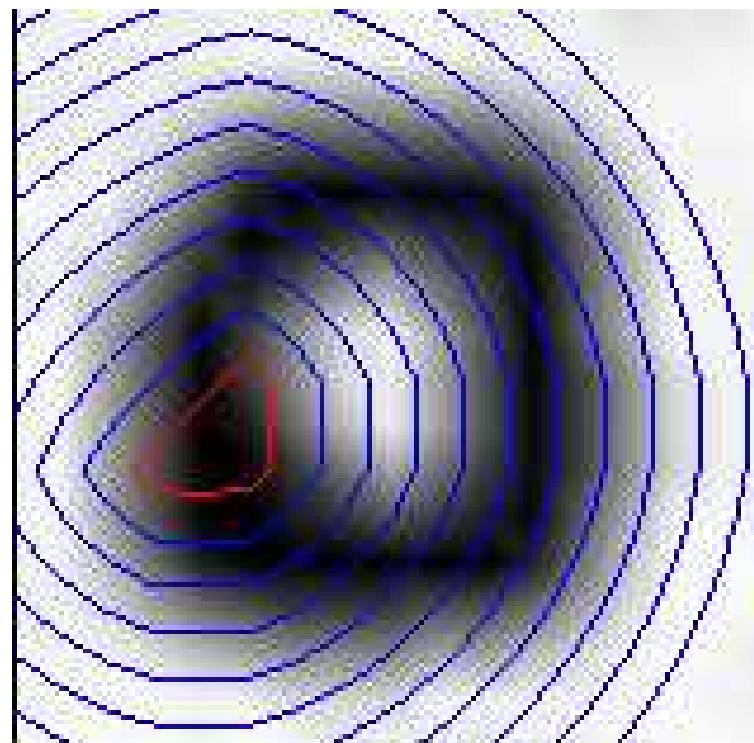


## Ewolucja konturu - zmiany topologii





## Ewolucja konturu - zmiany topologii



# GNU Free Documentation License



Version 1.2, November 2002

Copyright (C) 2000,2001,2002 Free Software Foundation, Inc.  
51 Franklin St, Fifth Floor, Boston, MA 02110-1301 USA  
Everyone is permitted to copy and distribute verbatim copies  
of this license document, but changing it is not allowed.

## 0. PREAMBLE

The purpose of this License is to make a manual, textbook, or other functional and useful document "free" in the sense of freedom: to assure everyone the effective freedom to copy and redistribute it, with or without modifying it, either commercially or noncommercially. Secondly, this License preserves for the author and publisher a way to get credit for their work, while not being considered responsible for modifications made by others.

This License is a kind of "copyleft", which means that derivative works of the document must themselves be free in the same sense. It complements the GNU General Public License, which is a copyleft license designed for free software.

We have designed this License in order to use it for manuals for free software, because free software needs free documentation: a free program should come with manuals providing the same freedoms that the software does. But this License is not limited to software manuals; it can be used for any textual work, regardless of subject matter or whether it is published as a printed book. We recommend this License principally for works whose purpose is instruction or reference.

## 1. APPLICABILITY AND DEFINITIONS

This License applies to any manual or other work, in any medium, that contains a notice placed by the copyright holder saying it can be distributed under the terms of this License. Such a notice grants a world-wide, royalty-free license, unlimited in duration, to use that work under the conditions stated herein. The "Document", below, refers to any such manual or work. Any member of the public is a licensee, and is addressed as "you". You accept the license if you copy, modify or distribute the work in a way requiring permission under copyright law.

A "Modified Version" of the Document means any work containing the Document or a portion of it, either copied verbatim, or with modifications and/or translated into another language.

A "Secondary Section" is a named appendix or a front-matter section of the Document that deals exclusively with the relationship of the publishers or authors of the Document to the Document's overall subject (or to related matters) and contains nothing that could fall directly within that overall subject. (Thus, if the Document is in part a textbook of mathematics, a Secondary Section may not explain any mathematics.) The relationship could be a matter of historical connection with the subject or with related matters, or of legal, commercial, philosophical, ethical or political position regarding them.

The "Invariant Sections" are certain Secondary Sections whose titles are designated, as being those of Invariant Sections, in the notice that says that the Document is released under this License. If a section does not fit the above definition of Secondary then it is not allowed to be designated as Invariant. The Document may contain zero Invariant Sections. If the Document does not identify any Invariant Sections then there are none.

The "Cover Texts" are certain short passages of text that are listed, as Front-Cover Texts or Back-Cover Texts, in the notice that says that the Document is released under this License. A Front-Cover Text may be at most 5 words, and a Back-Cover Text may be at most 25 words.

A "Transparent" copy of the Document means a machine-readable copy, represented in a format whose specification is available to the general public, that is suitable for revising the document straightforwardly with generic text editors or (for images composed of pixels) generic paint programs or (for drawings) some widely available drawing editor, and that is suitable for input to text formatters or for automatic translation to a variety of formats suitable for input to text formatters. A copy made in an otherwise Transparent file format whose markup, or absence of markup, has been arranged to thwart or discourage subsequent modification by readers is not Transparent. An image format is not Transparent if used for any substantial amount of text. A copy that is not "Transparent" is called "Opaque".

Examples of suitable formats for Transparent copies include plain ASCII without markup, Texinfo input format, LaTeX input format, SGML or XML using a publicly available DTD, and standard-conforming simple HTML, PostScript or PDF designed for human modification. Examples of transparent image formats include PNG, XCF and JPG. Opaque formats include proprietary formats that can be read and edited only by proprietary word processors, SGML or XML for which the DTD and/or processing tools are not generally available, and the machine-generated HTML, PostScript or PDF produced by some word processors for output purposes only.

The "Title Page" means, for a printed book, the title page itself, plus such following pages as are needed to hold, legibly, the material this License requires to appear in the title page. For works in formats which do not have any title page as such, "Title Page" means the text near the most prominent appearance of the work's title, preceding the beginning of the body of the text.

A section "Entitled XYZ" means a named subunit of the Document whose title either is precisely XYZ or contains XYZ in parentheses following text that translates XYZ in another language. (Here XYZ stands for a specific section name mentioned below, such as "Acknowledgements", "Dedications", "Endorsements", or "History".) To "Preserve the Title" of such a section when you modify the Document means that it remains a section "Entitled XYZ" according to this definition.

The Document may include Warranty Disclaimers next to the notice which states that this License applies to the Document. These Warranty Disclaimers are considered to be included by reference in this License, but only as regards disclaiming warranties: any other implication that these Warranty Disclaimers may have is void and has no effect on the meaning of this License.

## 2. VERBATIM COPYING

You may copy and distribute the Document in any medium, either commercially or noncommercially, provided that this License, the copyright notices, and the license notice saying this License applies to the Document are reproduced in all copies, and that you add no other conditions whatsoever to those of this License. You may not use technical measures to obstruct or control the reading or further copying of the copies you make or distribute. However, you may accept compensation in exchange for copies. If you distribute a large enough number of copies you must also follow the conditions in section 3.

You may also lend copies, under the same conditions stated above, and you may publicly display copies.



### 3. COPYING IN QUANTITY

If you publish printed copies (or copies in media that commonly have printed covers) of the Document, numbering more than 100, and the Document's license notice requires Cover Texts, you must enclose the copies in covers that carry, clearly and legibly, all these Cover Texts: Front-Cover Texts on the front cover, and Back-Cover Texts on the back cover. Both covers must also clearly and legibly identify you as the publisher of these copies. The front cover must present the full title with all words of the title equally prominent and visible. You may add other material on the covers in addition. Copying with changes limited to the covers, as long as they preserve the title of the Document and satisfy these conditions, can be treated as verbatim copying in other respects.

If the required texts for either cover are too voluminous to fit legibly, you should put the first ones listed (as many as fit reasonably) on the actual cover, and continue the rest onto adjacent pages.

If you publish or distribute Opaque copies of the Document numbering more than 100, you must either include a machine-readable Transparent copy along with each Opaque copy, or state in or with each Opaque copy a computer-network location from which the general network-using public has access to download using public-standard network protocols a complete Transparent copy of the Document, free of added material. If you use the latter option, you must take reasonably prudent steps, when you begin distribution of Opaque copies in quantity, to ensure that this Transparent copy will remain thus accessible at the stated location until at least one year after the last time you distribute an Opaque copy (directly or through your agents or retailers) of that edition to the public.

It is requested, but not required, that you contact the authors of the Document well before redistributing any large number of copies, to give them a chance to provide you with an updated version of the Document.

### 4. MODIFICATIONS

You may copy and distribute a Modified Version of the Document under the conditions of sections 2 and 3 above, provided that you release the Modified Version under precisely this License, with the Modified Version filling the role of the Document, thus licensing distribution and modification of the Modified Version to whoever possesses a copy of it. In addition, you must do these things in the Modified Version:

- \* A. Use in the Title Page (and on the covers, if any) a title distinct from that of the Document, and from those of previous versions (which should, if there were any, be listed in the History section of the Document). You may use the same title as a previous version if the original publisher of that version gives permission.
- \* B. List on the Title Page, as authors, one or more persons or entities responsible for authorship of the modifications in the Modified Version, together with at least five of the principal authors of the Document (all of its principal authors, if it has fewer than five), unless they release you from this requirement.
- \* C. State on the Title page the name of the publisher of the Modified Version, as the publisher.
- \* D. Preserve all the copyright notices of the Document.
- \* E. Add an appropriate copyright notice for your modifications adjacent to the other copyright notices.
- \* F. Include, immediately after the copyright notices, a license notice giving the public permission to use the Modified Version under the terms of this License, in the form shown in the Addendum below.
- \* G. Preserve in that license notice the full lists of Invariant Sections and required Cover Texts given in the Document's license notice.
- \* H. Include an unaltered copy of this License.
- \* I. Preserve the section Entitled "History", Preserve its Title, and add to it an item stating at least the title, year, new authors, and publisher of the Modified Version as given on the Title Page. If there is no section Entitled "History" in the Document, create one stating the title, year, authors, and publisher of the Document as given on its Title Page, then add an item describing the Modified Version as stated in the previous sentence.
- \* J. Preserve the network location, if any, given in the Document for public access to a Transparent copy of the Document, and likewise the network locations given in the Document for previous versions it was based on. These may be placed in the "History" section. You may omit a network location for a work that was published at least four years before the Document itself, or if the original publisher of the version it refers to gives permission.
- \* K. For any section Entitled "Acknowledgements" or "Dedications", Preserve the Title of the section, and preserve in the section all the substance and tone of each of the contributor acknowledgements and/or dedications given therein.
- \* L. Preserve all the Invariant Sections of the Document, unaltered in their text and in their titles. Section numbers or the equivalent are not considered part of the section titles.
- \* M. Delete any section Entitled "Endorsements". Such a section may not be included in the Modified Version.
- \* N. Do not retitle any existing section to be Entitled "Endorsements" or to conflict in title with any Invariant Section.
- \* O. Preserve any Warranty Disclaimers.

If the Modified Version includes new front-matter sections or appendices that qualify as Secondary Sections and contain no material copied from the Document, you may at your option designate some or all of these sections as invariant. To do this, add their titles to the list of Invariant Sections in the Modified Version's license notice. These titles must be distinct from any other section titles.

You may add a section Entitled "Endorsements", provided it contains nothing but endorsements of your Modified Version by various parties--for example, statements of peer review or that the text has been approved by an organization as the authoritative definition of a standard.

You may add a passage of up to five words as a Front-Cover Text, and a passage of up to 25 words as a Back-Cover Text, to the end of the list of Cover Texts in the Modified Version. Only one passage of Front-Cover Text and one of Back-Cover Text may be added by (or through arrangements made by) any one entity. If the Document already includes a cover text for the same cover, previously added by you or by arrangement made by the same entity you are acting on behalf of, you may not add another; but you may replace the old one, on explicit permission from the previous publisher that added the old one.

The author(s) and publisher(s) of the Document do not by this License give permission to use their names for publicity for or to assert or imply endorsement of any Modified Version.



## 6. COLLECTIONS OF DOCUMENTS

You may make a collection consisting of the Document and other documents released under this License, and replace the individual copies of this License in the various documents with a single copy that is included in the collection, provided that you follow the rules of this License for verbatim copying of each of the documents in all other respects.

You may extract a single document from such a collection, and distribute it individually under this License, provided you insert a copy of this License into the extracted document, and follow this License in all other respects regarding verbatim copying of that document.

## 7. AGGREGATION WITH INDEPENDENT WORKS

A compilation of the Document or its derivatives with other separate and independent documents or works, in or on a volume of a storage or distribution medium, is called an "aggregate" if the copyright resulting from the compilation is not used to limit the legal rights of the compilation's users beyond what the individual works permit. When the Document is included in an aggregate, this License does not apply to the other works in the aggregate which are not themselves derivative works of the Document.

If the Cover Text requirement of section 3 is applicable to these copies of the Document, then if the Document is less than one half of the entire aggregate, the Document's Cover Texts may be placed on covers that bracket the Document within the aggregate, or the electronic equivalent of covers if the Document is in electronic form. Otherwise they must appear on printed covers that bracket the whole aggregate.

## 8. TRANSLATION

Translation is considered a kind of modification, so you may distribute translations of the Document under the terms of section 4. Replacing Invariant Sections with translations requires special permission from their copyright holders, but you may include translations of some or all Invariant Sections in addition to the original versions of these Invariant Sections. You may include a translation of this License, and all the license notices in the Document, and any Warranty Disclaimers, provided that you also include the original English version of this License and the original versions of those notices and disclaimers. In case of a disagreement between the translation and the original version of this License or a notice or disclaimer, the original version will prevail.

If a section in the Document is Entitled "Acknowledgements", "Dedications", or "History", the requirement (section 4) to Preserve its Title (section 1) will typically require changing the actual title.

## 9. TERMINATION

You may not copy, modify, sublicense, or distribute the Document except as expressly provided for under this License. Any other attempt to copy, modify, sublicense or distribute the Document is void, and will automatically terminate your rights under this License. However, parties who have received copies, or rights, from you under this License will not have their licenses terminated so long as such parties remain in full compliance.

## 10. FUTURE REVISIONS OF THIS LICENSE

The Free Software Foundation may publish new, revised versions of the GNU Free Documentation License from time to time. Such new versions will be similar in spirit to the present version, but may differ in detail to address new problems or concerns. See <http://www.gnu.org/copyleft/>.

Each version of the License is given a distinguishing version number. If the Document specifies that a particular numbered version of this License "or any later version" applies to it, you have the option of following the terms and conditions either of that specified version or of any later version that has been published (not as a draft) by the Free Software Foundation. If the Document does not specify a version number of this License, you may choose any version ever published (not as a draft) by the Free Software Foundation.

## 5. COMBINING DOCUMENTS

You may combine the Document with other documents released under this License, under the terms defined in section 4 above for modified versions, provided that you include in the combination all of the Invariant Sections of all of the original documents, unmodified, and list them all as Invariant Sections of your combined work in its license notice, and that you preserve all their Warranty Disclaimers.

The combined work need only contain one copy of this License, and multiple identical Invariant Sections may be replaced with a single copy. If there are multiple Invariant Sections with the same name but different contents, make the title of each such section unique by adding at the end of it, in parentheses, the name of the original author or publisher of that section if known, or else a unique number. Make the same adjustment to the section titles in the list of Invariant Sections in the license notice of the combined work.

In the combination, you must combine any sections Entitled "History" in the various original documents, forming one section Entitled "History"; likewise combine any sections Entitled "Acknowledgements", and any sections Entitled "Dedications". You must delete all sections Entitled "Endorsements."

(not as a draft) by the Free Software Foundation.



# Techniki przetwarzania mediów cyfrowych

Wykład 1: Wybrane elementy akustyki (4 h)

# Wybrane elementy akustyki

- Natura dźwięku
- Parametry fali dźwiękowej
- Obrazy dźwięku
- Percepcja dźwięku przez człowieka



Fundusze  
Europejskie  
Wiedza Edukacja Rozwój



Politechnika Wrocławskiego

Unia Europejska  
Europejski Fundusz Społeczny



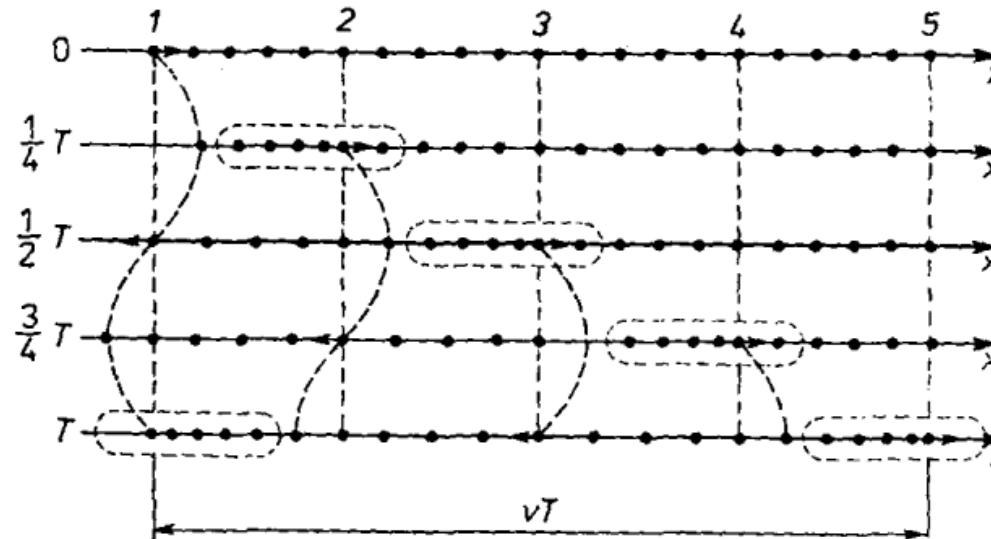
# Czym jest dźwięk?

**Dźwięk (w znaczeniu fizycznym):**

- Jest zaburzeniem mechanicznym gęstości ośrodka w postaci chwilowych zagęszczeń i rozrzedzeń, które powodują chwilowe zmiany ciśnienia w otoczeniu zaburzenia.
- Naprzemienne zagęszczenie i rozrzedzenie powoduje powstanie okresowego ciśnienia zmiennego (fali dźwiękowej).
- Pod wpływem zaburzenia cząsteczki ośrodka drgają wokół położenia równowagi w kierunku rozchodzenia się fali (w gazach i cieczach) przenosząc energię z jednego miejsca do drugiego bez przemieszczania masy.
- Docierając do ludzkiego ucha fala dźwiękowa wywołuje wrażenia słuchowe
- Dźwięk rozchodzi się w ośrodkach sprężystych (gazy, ciecze, ciała stałe)

# Źródła dźwięku

- Źródłem dźwięku są drgające ciała stałe (np. struny, membrany) oraz zawirowania powietrza.
- Drgające elementy źródła zagęszczają i rozrzedzają cząsteczki otaczającego ośrodka sprężystego, wywołując w nim falę.



# Wrażenia słuchowe

- Wrażenia słuchowe odbierane przez człowieka, czyli dźwięk (w sensie psychologicznym) charakteryzowane za pomocą cech subiektywnych, z którymi wiążą się obiektywne parametry fali dźwiękowej:

Subiektywne cechy dźwięku	Obiektywne parametry fali dźwiękowej
Wysokość	Częstotliwość
Głośność	Natężenie
Barwa	Widmo (wyższe harmoniczne)

# Wrażenia słuchowe

- Z subiektywnie odczuwaną wysokością odbieranego dźwięku wiąże się częstotliwość fali dźwiękowej.
- Im częstotliwość jest większa, tym wyższy wydaje nam się dźwięk.
- Dźwięki słyszalne przez człowieka mieszczą się w przedziale od  $16 \text{ Hz} < f < 20\,000 \text{ Hz}$
- Dolna i górna granica zakresu dźwięków słyszalnych jest ulega zmianie w zależności od wieku, stopnia zmęczenia czy stanu emocjonalnego.
- Tony o częstotliwościach:
  - poniżej 300 - 400 Hz człowiek odbiera jako niskie,
  - dźwięki z przedziału 400 - 3000 Hz - jako tony średnie,
  - powyżej 3 kHz – jako tony wysokie.
- Dźwięki niesłyszalne dla człowieka:
  - infradźwięki  $f < 16 \text{ Hz}$ ,
  - ultradźwięki  $f > 20 \text{ kHz}$ 
    - hiperdźwięki  $f > 1 \text{ GHz}$



# Wrażenia słuchowe

- Wrażenia słuchowe wywołane przez falę dźwiękową rozchodząą się w powietrzu zależą od:
  - rodzaju drgającego przedmiotu (jego kształtu, materiału, z którego został wykonany),
  - sposobu, w jaki drga źródło dźwięku,
  - własności przestrzeni, w której dźwięk się rozchodzi,
  - indywidualnych własności naszego ucha.
- Przetwarzanie dźwięku cyfrowego powoduje zmiany sposobu drgania źródła (membrany głośników lub słuchawek), co pociąga za sobą zmianę wrażeń słuchowych u odbiorcy.

# Parametry fali dźwiękowej

- Rodzaj drgań źródła można opisać w postaci funkcji wychylenia membrany (np. głośnika lub mikrofonu) z położenia równowagi w zależności od czasu.
- Najprostszym rodzajem drgań źródła dźwięku są drgania sinusoidalne. Definiuje je funkcja:

$$A(t)=A_0\sin(2\pi ft+\phi) \quad \text{lub} \quad A(t)=A_0\sin(\omega t +\phi)$$

gdzie:

$t$  – czas,

$A(t)$  – wychylenie membrany w chwili  $t$ ,

$A_0$  – maksymalne wychylenie (amplituda drgań),

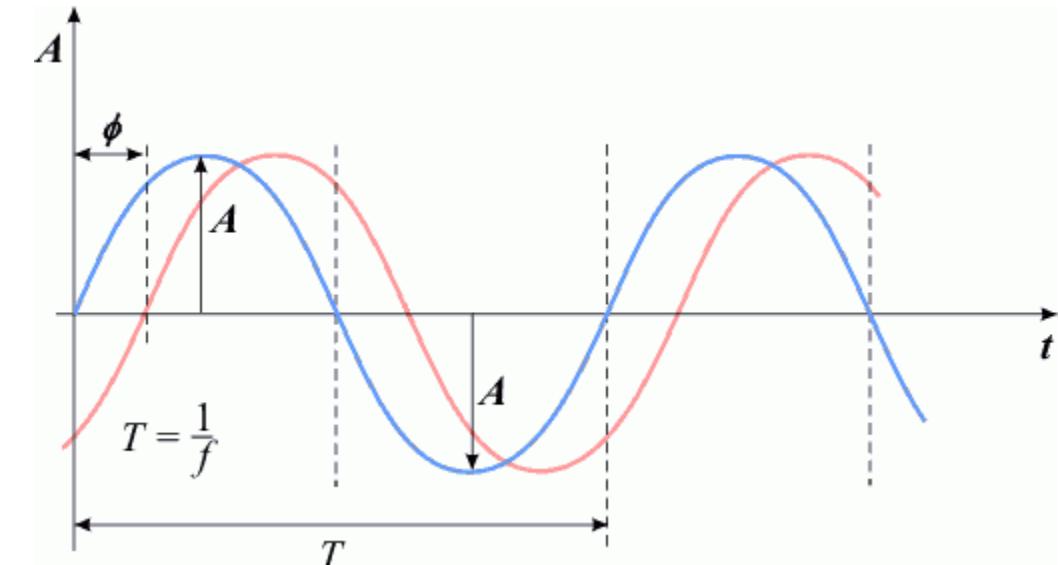
$f$  – częstotliwość drgań,

$\phi$  – faza początkowa w chwili  $t=0$ ,

$(\omega t +\phi)$  – faza ruchu w chwili  $t$ ,

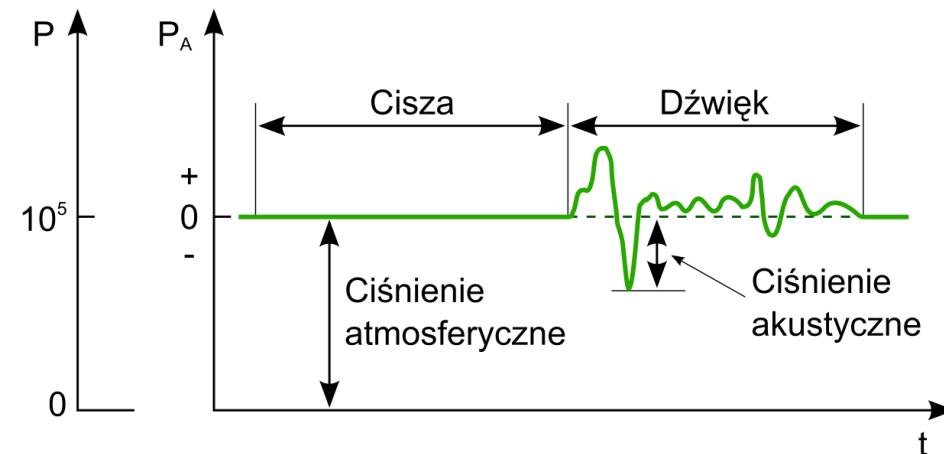
$T$  – okres drgań,  $T = 1/f$ ,

$\omega$  – częstość kołowa;  $\omega = 2\pi f = 2\pi/T$



# Parametry fali dźwiękowej

- Wychylenie membrany w danym momencie czasu A ( $t$ ) jest proporcjonalne do wartości chwilowej ciśnienia akustycznego  $p(t)$  fali wprawiającej membranę w drgania.
- Ciśnienie akustyczne jest zmiennym w czasie odchyleniem od średniej wartości ciśnienia statycznego panującego w ośrodku, występującym podczas rozchodzenia się w nim fali akustycznej.



Źródło: Wikipedia, [https://pl.wikipedia.org/wiki/Ci%C5%9Bnienie\\_akustyczne#/media/Plik:Ci%C5%9Bnienie\\_akustyczne.svg](https://pl.wikipedia.org/wiki/Ci%C5%9Bnienie_akustyczne#/media/Plik:Ci%C5%9Bnienie_akustyczne.svg)

# Parametry fali akustycznej

- Podczas przemieszczania się fal dźwiękowych mamy do czynienia z dwoma rodzajami prędkości: prędkością dźwięku oraz prędkością akustyczną.
- Prędkość dźwięku  $c$ , tj. prędkość, z jaką fale dźwiękowe rozchodzą się ośrodka, zależy od długości fali  $\lambda$  oraz jej częstotliwości  $f$  i wyraża się wzorem:  
$$c = \lambda \cdot f$$
- Prędkość akustyczna  $u$  jest średnią prędkością drgającej części, która zależy od amplitudy drgań  $A_0$  i częstości kołowej  $\omega$ :

$$u = A_0 \cdot \omega$$

- Między prędkością akustyczną  $u$  a ciśnieniem akustycznym  $p$  występuje zależność:

$$p = u \cdot \rho c,$$

Gdzie:

$\rho$  – gęstość ośrodka,

$\rho c$  – opór akustyczny



# Parametry fali dźwiękowej

- Najsłabszy szmer odbierany przez ucho ludzkie, ma ciśnienie akustyczne  $p = 2 \cdot 10^{-5}$  Pa. Jest to tzw. **Próg słyszalności**.
- Ciśnienie akustyczne dźwięku, przy którym człowiek przestaje słyszeć rozróżniając cechy takiego dźwięku, a zaczyna odczuwać ból w uszach wynosi 20 Pa i nosi nazwę **granicy lub progu bólu**.
- Ze względu na dużą rozpiętość wartości ciśnienia akustycznego stosuje się bezwymiarową wielkość zwaną **poziomem ciśnienia akustycznego**:

$$L_p = 10 \log \left( \frac{p}{p_0} \right)^2$$

Gdzie:

$p_0$  - ciśnienie odniesienia, równe co do wartości progowi słyszalności

- Wartość poziomu ciśnienia akustycznego podaje się w dB SPL (Sound Pressure Level)



# Parametry fali dźwiękowej

- Fale akustyczne przenoszą energię.
- Energia fali dźwiękowej jest proporcjonalna do kwadratu jej amplitudy:

$$E \sim A_0^2$$

- Miarą tej energii jest natężenie dźwięku wyrażone w  $\text{W/m}^2$
- Wartość natężenia dźwięku jest równa średniej wartości strumienia energii akustycznej przepływającego w czasie 1 s przez jednostkowe pole powierzchni ( $1 \text{ m}^2$ ) prostopadłej do kierunku rozchodzenia się fali.



# Parametry fali dźwiękowej

- Wartość natężenia dźwięku oblicza się z zależności:

$$I = \frac{P}{S}$$

gdzie:

$P$  – moc akustyczna źródła dźwięku (parametr źródła, np. głośników),

$S$  - pole zamkniętej powierzchni zawierającej wewnątrz źródło.

- Dźwięk emitowany przez źródła rzeczywiste jest falą kulistą lub superpozycją fal kulistych.
- Natężenie dźwięku w odległości  $r$  od źródła punktowego emitującego falę kulistą wyraża się wzorem:

$$I(r) = \frac{P}{4\pi \cdot r^2}$$



# Parametry fali dźwiękowej

- Człowiek słyszy dźwięki w bardzo szerokim zakresie natężenia od  $10^{-12}$  W/m<sup>2</sup> (natężenie progu słyszalności) do ok. 1 W/m<sup>2</sup> (natężenie progu bólu).
- Różnica między najcichszym i najgłośniejszym słyszalnym dźwiękiem wynosi  $10^{12}$ .
- Ucho ludzkie przystosowało się do odbioru dźwięków o tak dużej dynamice „logarytmując” natężenie dźwięku.
- To powoduje, że dźwięki odbierane przez człowieka jako te o nieco większym natężeniu w rzeczywistości mogą mieć natężenie kilkudziesiąt, a nawet kilkaset razy większe.
- Uwzględniając logarytmiczną naturę ludzkiego słuchu wprowadzono wielkość zwaną poziomem natężenia dźwięku.

# Parametry fali dźwiękowej

- Poziom natężenia dźwięku jest logarytmiczną miarą natężenia dźwięku w stosunku do pewnej umownie przyjętej wartości odniesienia  $I_0$ .
- Poziom natężenia dźwięku  $L$  wyznacza się ze wzoru:

$$L = 10 \log_{10} \left( \frac{I}{I_0} \right) [dB]$$

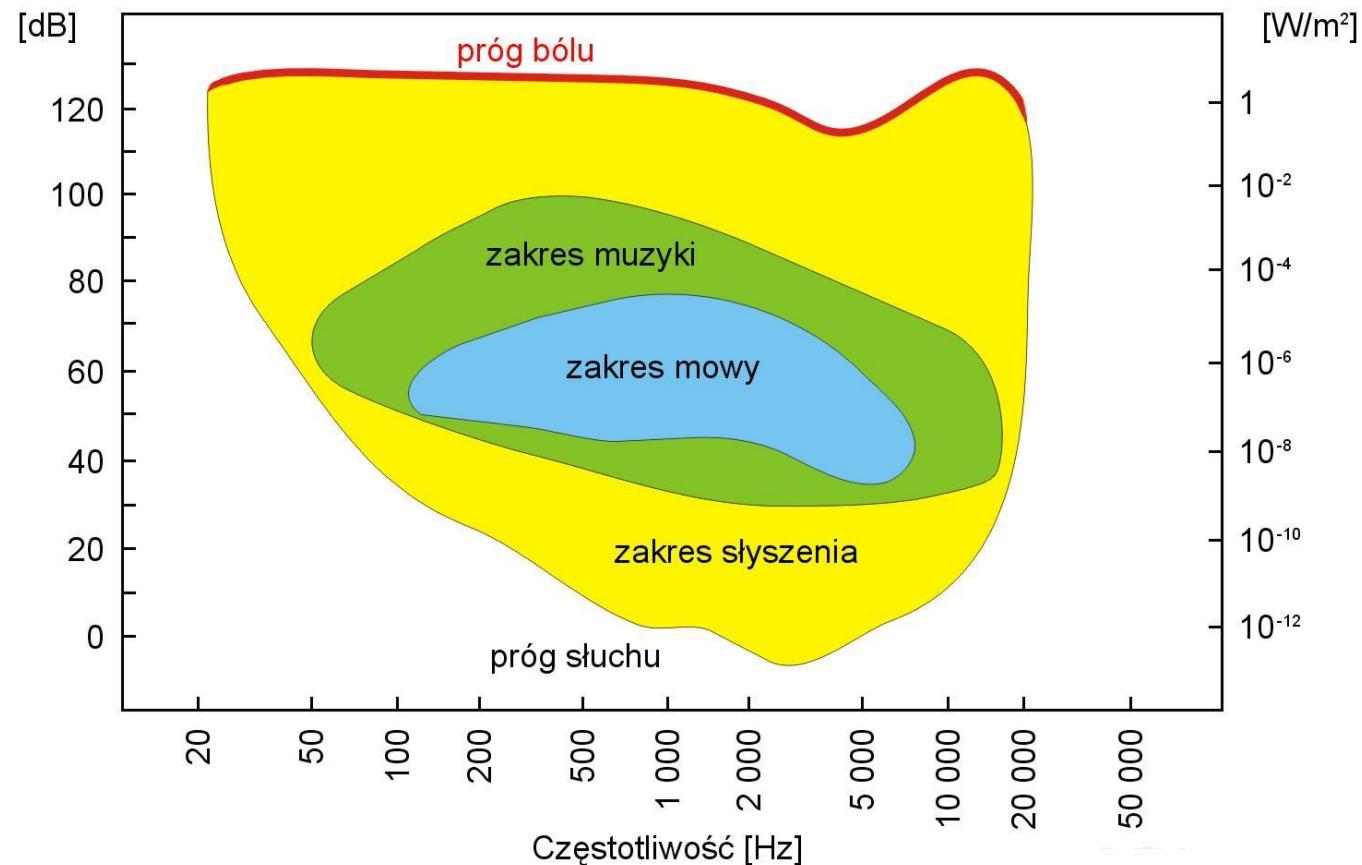
gdzie:

- $I$  – natężenie dźwięku,
- $I_0$  – natężenie dźwięku o częstotliwości 1000 Hz i mocy  $2 \cdot 10^{-5}$  Pa stanowiące wartość odniesienia,
- $I_0 = 10^{-12}$  W/m<sup>2</sup> i jest najcichszym dźwiękiem odbieranym przez człowieka.



# Percepcja dźwięku przez człowieka

## Dźwięki odbierane przez człowieka



Fundusze  
Europejskie  
Wiedza Edukacja Rozwój



Politechnika Wrocławskiego

Unia Europejska  
Europejski Fundusz Społeczny



# Percepcja dźwięku przez człowieka

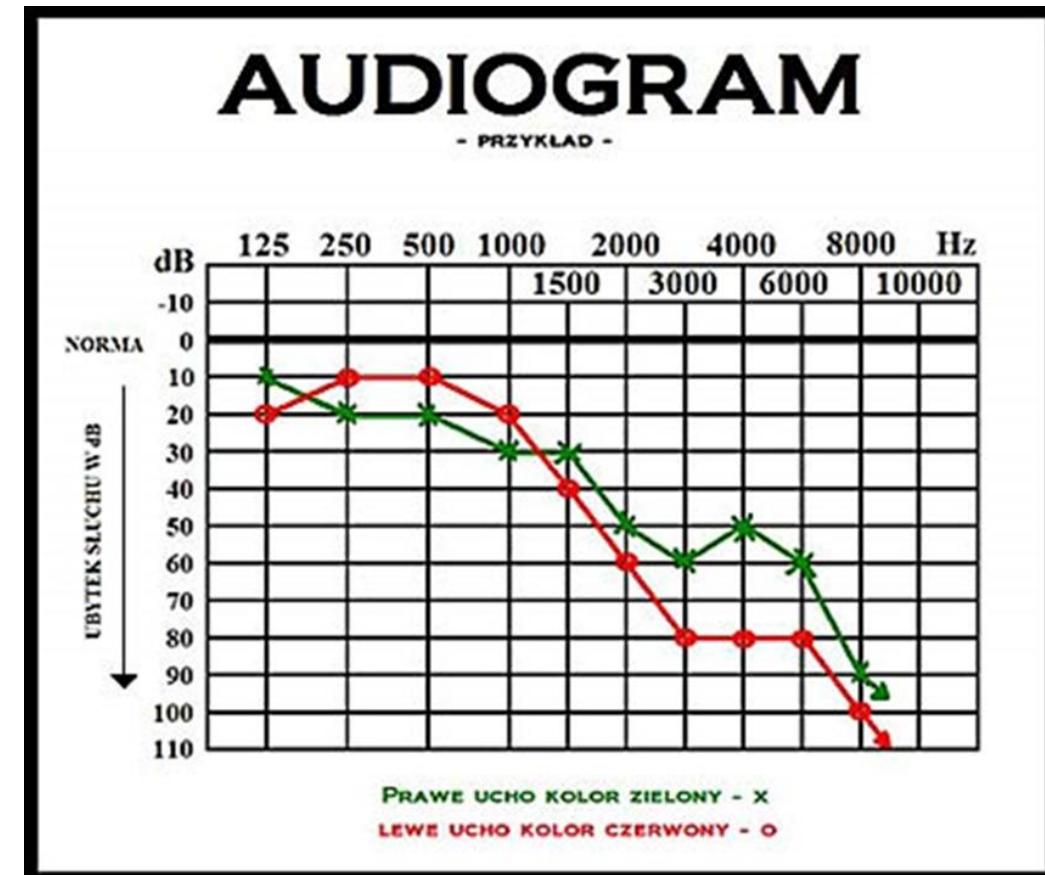
- Obiektywne parametry fali akustycznej (wielkości fizyczne) nie oddają w pełni tego, które dźwięki i w jaki sposób są odbierane przez człowieka.
- Dlatego wprowadzono parametry subiektywne (perceptualne), jak głośność i poziom głośności, wysokość oraz barwa dźwięku, które odpowiadają ludzkiemu postrzeganiu dźwięku.
- Wartości tych parametrów zostały określone na podstawie statystycznych badań odsłuchowych dużej populacji ludzi



# Percepcja dźwięku przez człowieka

## subiektywne postrzeganie głośności

- Podstawową metodą stosowaną w subiektywnych badaniach odsłuchowych jest audiometria tonalna.
- Badanie słuchu polega na pomiarze słuchu dla tonów sinusoidalnych o różnych częstotliwościach.
- Wynik badania przedstawiany jest w postaci krzywych – audiogramów, które przedstawiają próg słyszenia badanej osoby.



# Percepcja dźwięku przez człowieka

## subiektywne postrzeganie głośności

- **Głośność** jest wielkością charakteryzującą subiektywne odczuwanie natężenia dźwięku przez człowieka.
- Stanowi podstawę dla zróżnicowania dynamiki, np. siły brzmienia w utworze muzycznym.
- Głośność dźwięku odbieranego przez człowieka zależy od natężenia i częstotliwości.
- Przy stałym natężeniu:
  - jako najgłośniejsze odbierane są dźwięki o częstotliwości 3-4 kHz,
  - jako najmniej głośne - dźwięki o częstotliwościach poniżej 100 Hz oraz powyżej 10000 Hz.
- Jednostką głośności jest son.
- 1 son odpowiada głośności dźwięku o częstotliwości 1 kHz i o poziomie natężenia 40 dB.



# Percepcja dźwięku przez człowieka

## subiektywne postrzeganie głośności

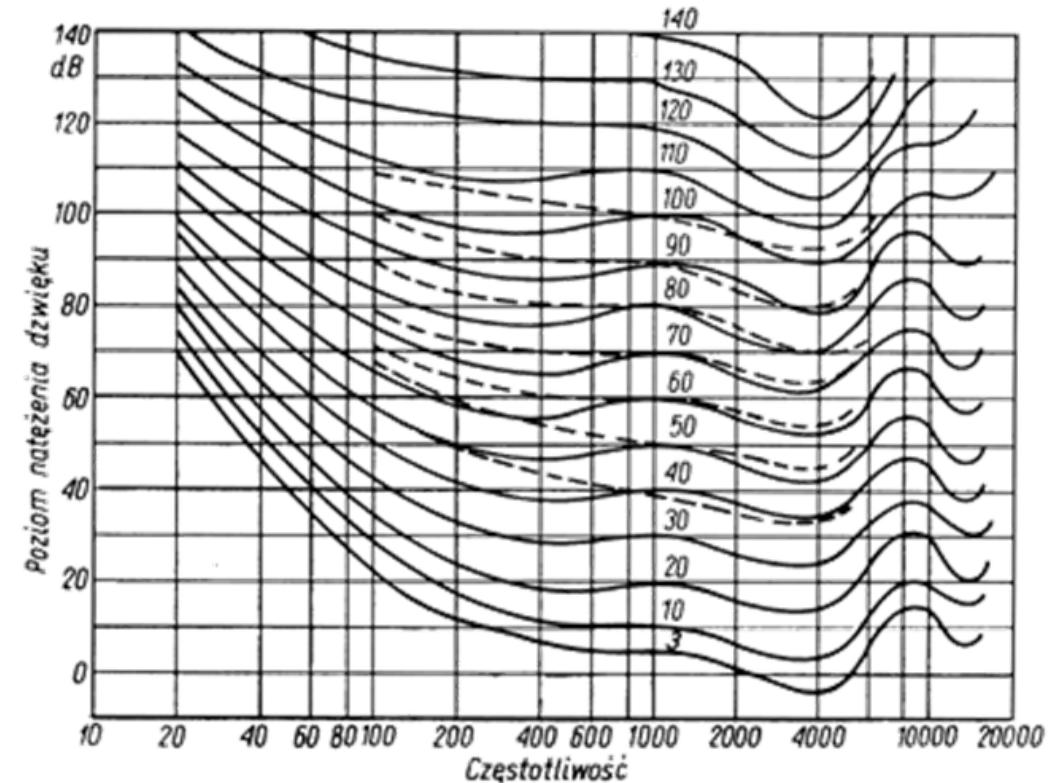
- **Poziom głośności dźwięku** jest względną miarą głośności dźwięku w odniesieniu do tonu wzorcowego.
- Wyrażany jest w **fonach**,
- Poziom głośności danego dźwięku jest liczbowo równy poziomowi ciśnienia akustycznego  $L_p$  tonu o częstotliwości 1 kHz, który brzmi jednakowo głośno jak dany dźwięk
- Dla tonu o częstotliwości 1 kHz: 1 fon = 1 dB ,
- Dźwięki o tej samej liczbie fonów wywołują to samo wrażenie głośności, ale nie muszą być to dźwięki identyczne.
- Korzystając z poziomu głośności można ocenić, czy dany dźwięk jest cichszy czy głośniejszy od innego, ale nie można stwierdzić ile razy.



# Percepcja dźwięku przez człowieka

## subiektywne postrzeganie głośności

- W zastosowaniach praktycznych korzysta się z wykresów krzywych jednakowej głośności – izofonów.
- Ze względu na subiektywny charakter, krzywe izofoniczne różnią się w zależności od metod badawczych użytych do ich wyznaczenia.
- Za standardowe przyjęto izofony wyznaczone przez Robinsona i Dadsona



# Percepcja dźwięku przez człowieka subiektywne postrzeganie głośności

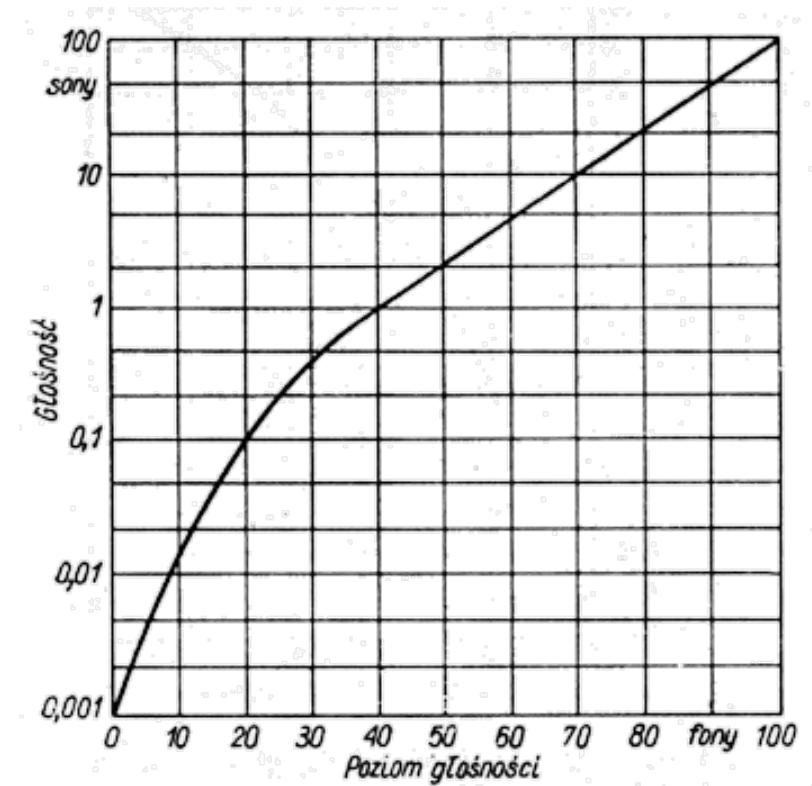
- Empirycznie wyznaczono również zależność między głośnością (w sonach) a poziomem głośności fonach (rysunek obok)
- Dla celów standaryzacyjnych przyjęto, że krzywą dobrze przybliża zależność:

$$S = 2^{(L_s - 40)/10}$$

gdzie:

S - głośność w sonach;

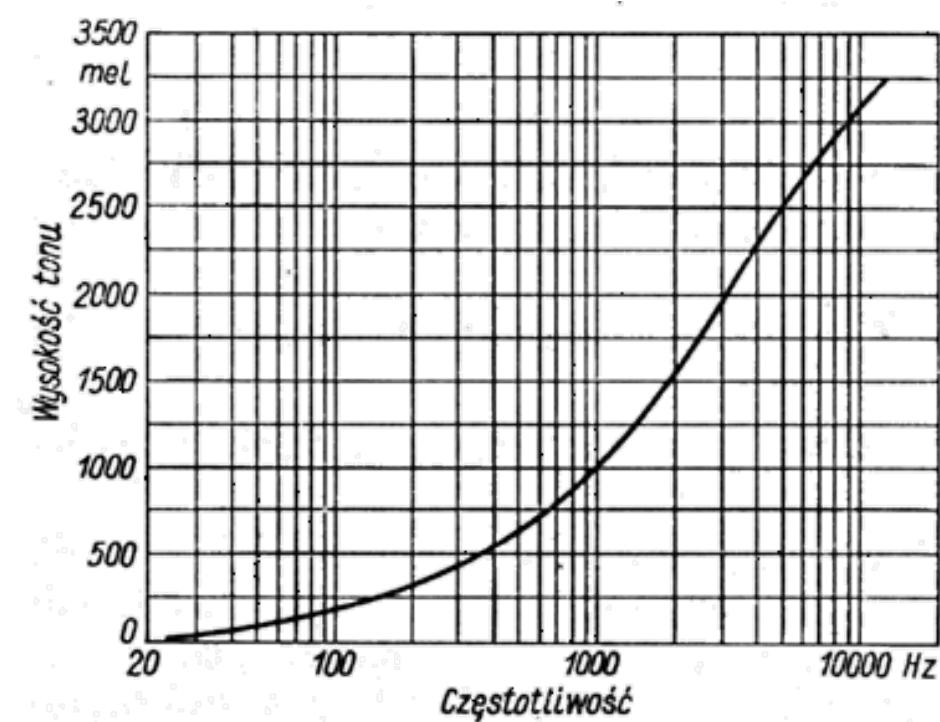
L<sub>s</sub> - poziom głośności w fonach



# Percepcja dźwięku przez człowieka

## subiektywny odbiór wysokości

- Wysokość dźwięku jest wrażeniem słuchowym, umożliwiającym określenie położenia tonu na skali częstotliwości.
- Jednostką wysokości tonu jest mel.
- Wrażenie wysokości tonu zależy głównie (choć nie tylko) od częstotliwości dźwięku.
- Na podstawie pomiarów ustalono związek między subiektywną skalą wysokości tonu a obiektywną skalą częstotliwości, przy założeniu, że wysokość tonu o częstotliwości 1000Hz i poziomie ciśnienia akustycznego równym 40 dB wynosi 1000 melów.



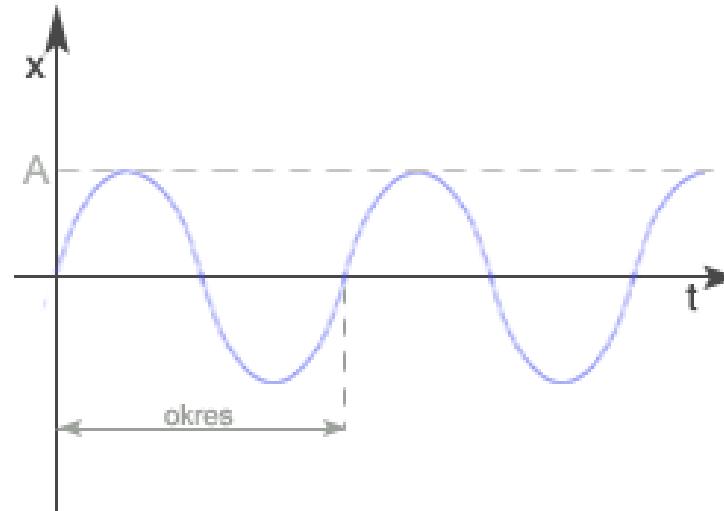
# Percepcja dźwięku przez człowieka subiektywny odbiór barwy

- Barwa dźwięku (tembr) jest cechą subiektywną, która pozwala na odróżnienie różnych głosów i brzmień instrumentów.
- Barwa zależy od widma dźwięku, na które składa się określona i różna dla różnych dźwięków liczba tonów prostych o różnych amplitudach i częstotliwościach.
- Chociaż tembr jest charakterystyczny dla danego źródła dźwięku (np. instrumentu), to może się nieznacznie zmieniać w zależności od:
  - sposobu wzbudzania drgań (pociągnięcie smyczkiem, szarpanie lub uderzenie),
  - siły wzbudzenia (a tym samym głośności dźwięku),
  - częstotliwości (różne struny mogą wydawać dźwięki nieco różniące się barwą),
  - zmian w czasie (związane z różnymi fazami obwiedni dźwięku).



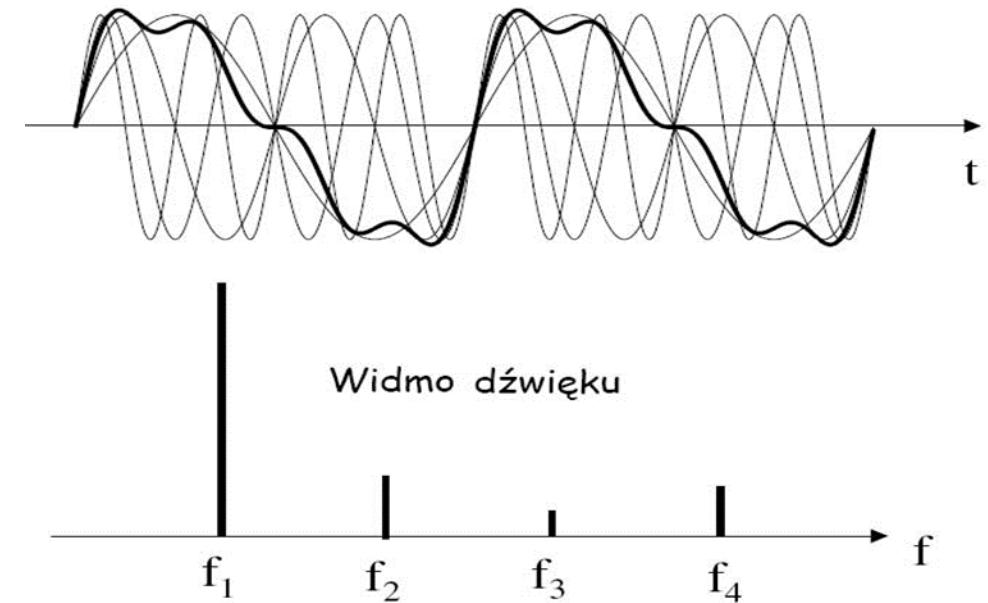
# Tony i wielotony

- Tonem lub tonem prostym nazywa się dźwięk o przebiegu sinusoidalnym i ścisłe określonej częstotliwości, amplitudzie i fazie.
- Dźwięk taki można wytworzyć przy pomocy kamertonu lub generatora elektro-akustycznego.



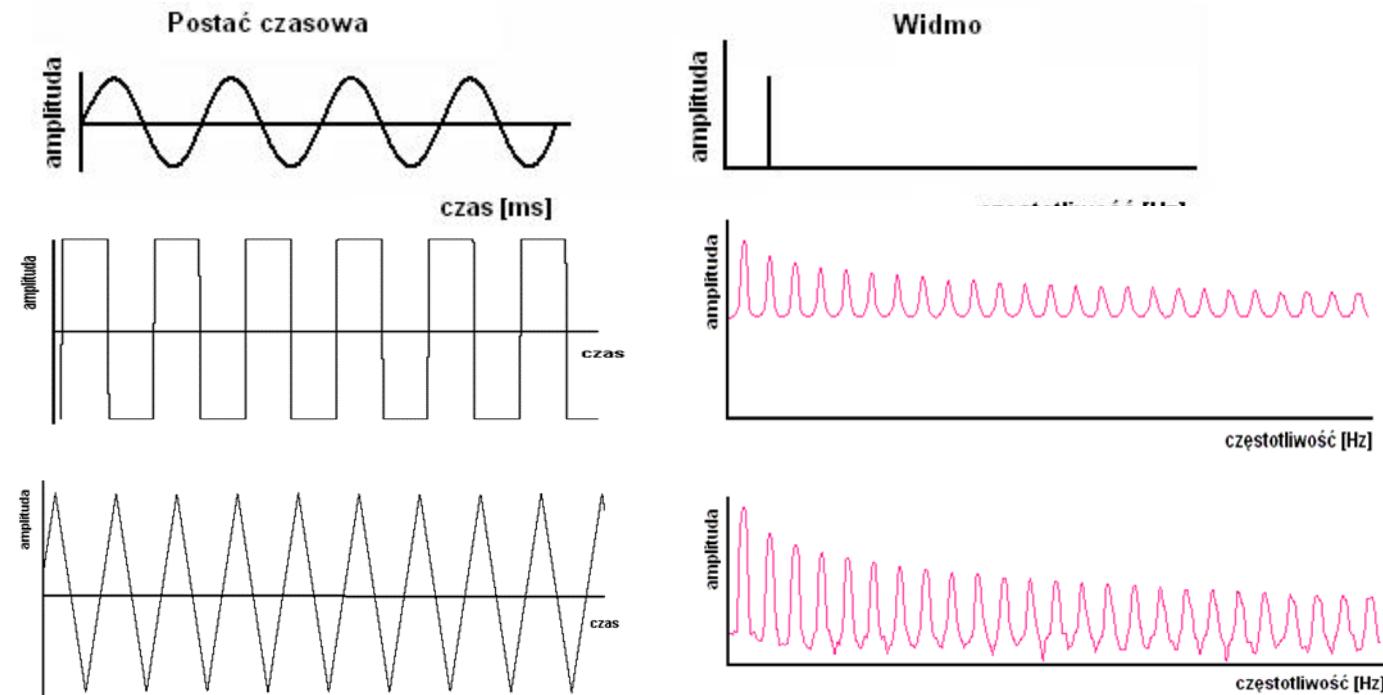
# Tony i wielotoniny

- Zdecydowana większość dźwięków spotykanych w otoczeniu to dźwięki bardziej złożone, będące efektem skomplikowanych drgań ich źródła.
- Dźwięki takie nazywamy **wielotonami**.
- Zgodnie z twierdzeniem Fouriera dowolny wieloton można przedstawić jako sumę tonów prostych o różnych częstotliwościach i amplitudach.
- Wieloton zbudowany jest z tonów składowych.



# Obrazy dźwięku – widmo (spektrum)

- Widmem akustycznym dźwięku (spektrum) nazywa się zbiór składowych sinusoidalnych dźwięku złożonego (wielotonu), określonych przez amplitudę i częstotliwość każdej składowej
- Widma uzyskuje się metodami spektroskopii lub jako wynik analizy Fouriera przebiegu falowego dźwięku.



Fundusze  
Europejskie  
Wiedza Edukacja Rozwój



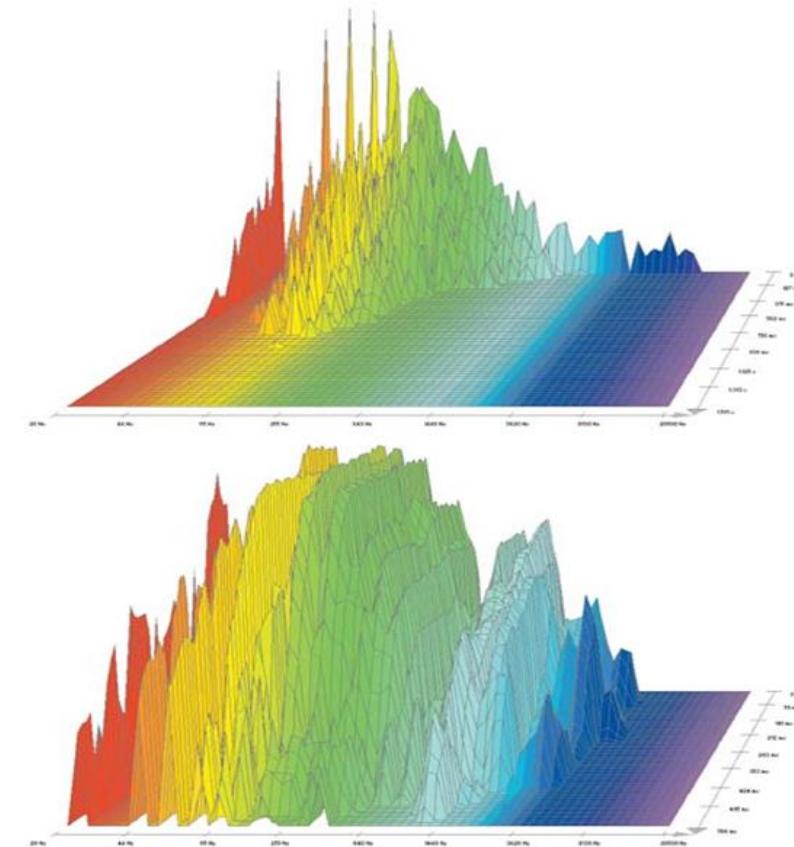
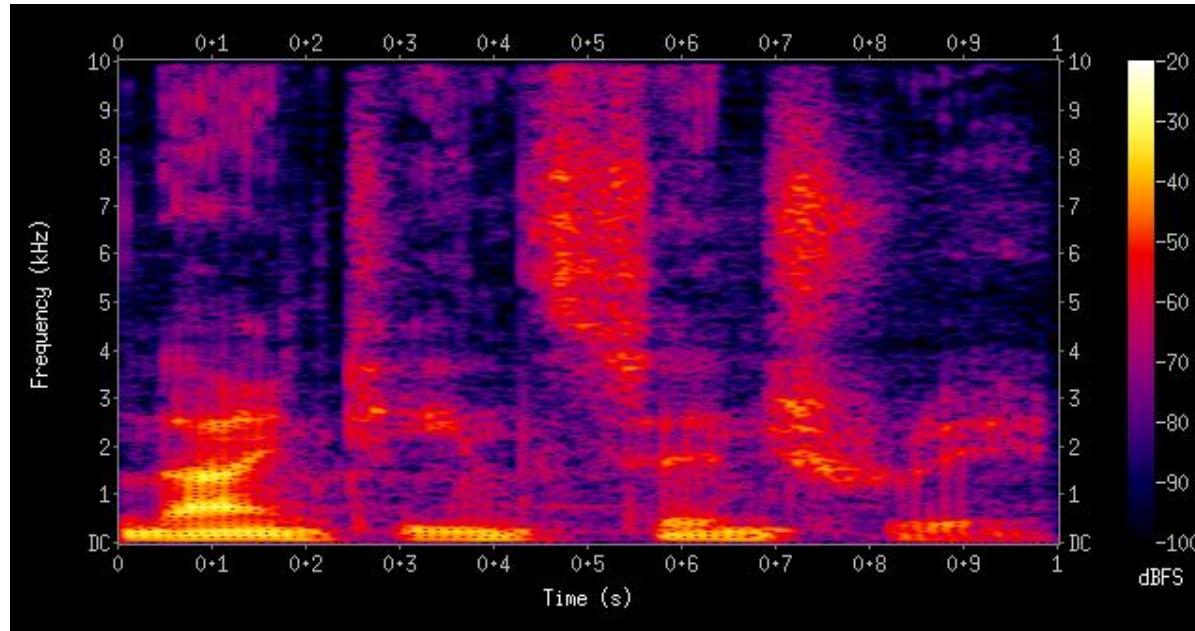
Politechnika Wrocławska

Unia Europejska  
Europejski Fundusz Społeczny



# Obrazy dźwięku - spektrogram

- Spektrogram to trójwymiarowy wykres widma amplitudowego sygnału dla każdej chwili  $t$ , dla której sygnał jest określony.
- Argumentami spektrogramu są częstotliwość i czas.



Fundusze  
Europejskie  
Wiedza Edukacja Rozwój



Politechnika Wrocławska

Unia Europejska  
Europejski Fundusz Społeczny

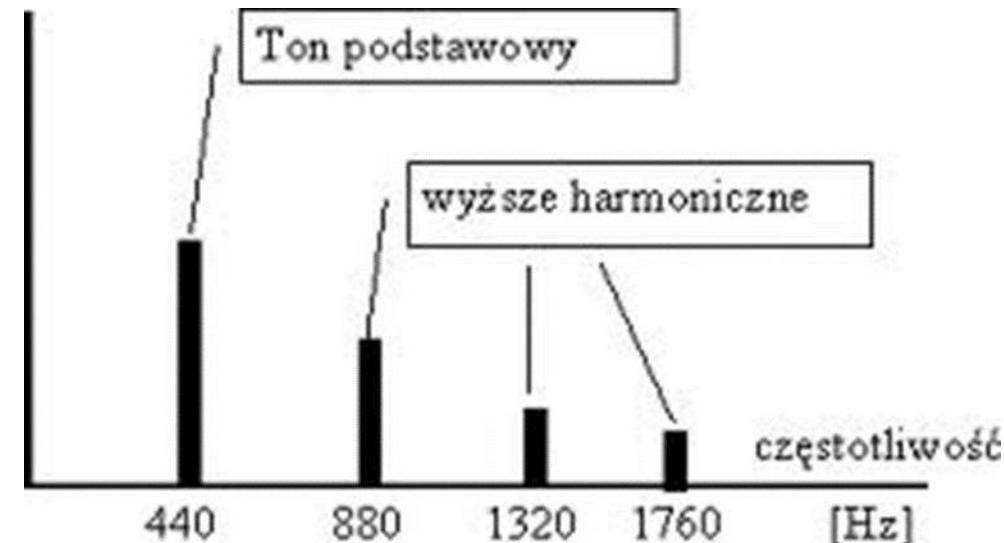


# Widmo dźwięku

- Ze względu na charakter widma wyróżnia się:
  - Dźwięki o widmie dyskretnym:
    - ✓ Tony proste, których częstotliwość i wysokość są ścisłe określone,
    - ✓ Wielotony harmoniczne, składające się z wielu tonów prostych (np. mowa, śpiew, muzyka)
  - Dźwięki o widmie ciągłym (szum, hałas).
- Widmo wielotonu harmonicznego zawiera ton podstawy i jego wielokrotności (składowe harmoniczne, alikwoty).
- Ton podstawowy jest największym wspólnym podzielnikiem składowych widma i odpowiada za postrzeganą wysokość dźwięku (nawet jeśli nie występuje w spektrum).

# Widmo harmoniczne

Na przykład, jeśli w widmie dźwięku docierającego do naszego ucha występują składowe harmoniczne o częstotliwościach: 440 Hz, 880 Hz i 660 Hz, to wysokość dźwięku słyszanego będzie odpowiadać częstotliwości 220 Hz.



# Widmo harmoniczne

- Częstotliwość tonu podstawowego – pierwszej składowej harmonicznej - oznacza się przez  $f_0$ .
- Częstotliwości wyższych składowych harmonicznych spełniają zależność:

gdzie:

$$f_n = n \cdot f_0$$

$f_0$  – częstotliwość tonu podstawowego,

$f_n$  – częstotliwość n-tej składowej harmonicznej,

$n$  – liczba naturalna większa od 1.

- Liczba wyższych składowych harmonicznych oraz ich amplitudy, czyli skład widmowy dźwięku, decydują o jego barwie.



# Szum akustyczny

- Wielotony, w których widmie nie istnieje największy wspólny podzielnik różny od jeden nazywa się wielotonami anharmonicznymi (nieharmonicznymi).
- O widmie takich dźwięków mówi się, że ma charakter ciągły.
- Przykładem tego rodzaju dźwięku jest szum akustyczny.
- Szumem akustycznym określa się dźwięk, którego widmo jest w większości zakresu słyszalności zrównoważone, tzn. nie występują w nim gwałtowne "piki", które mogłyby być odbierane jako dźwięczące rezonanse o określonej wysokości tonu.

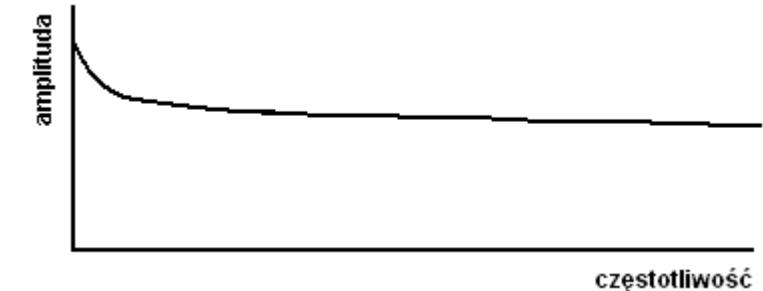
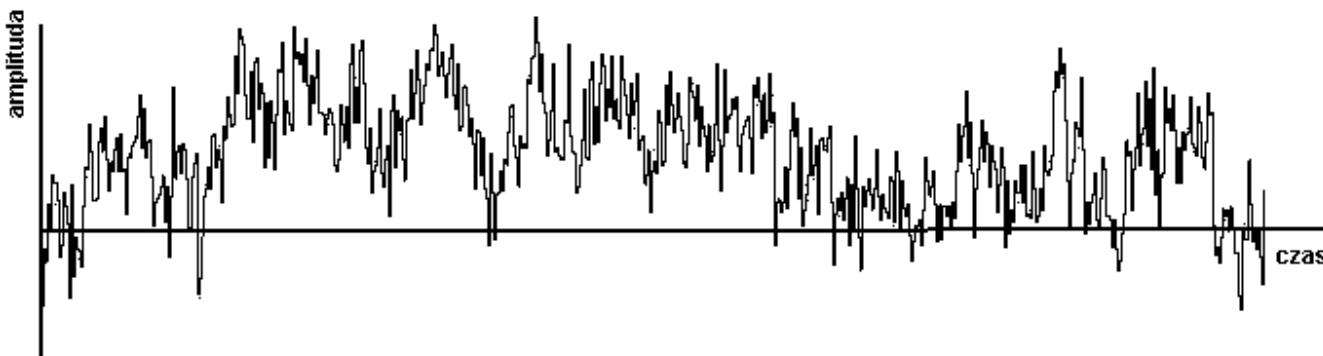
# Szum akustyczny

- Szum akustyczny o całkowicie płaskim widmie nazywa się **szumem białym**



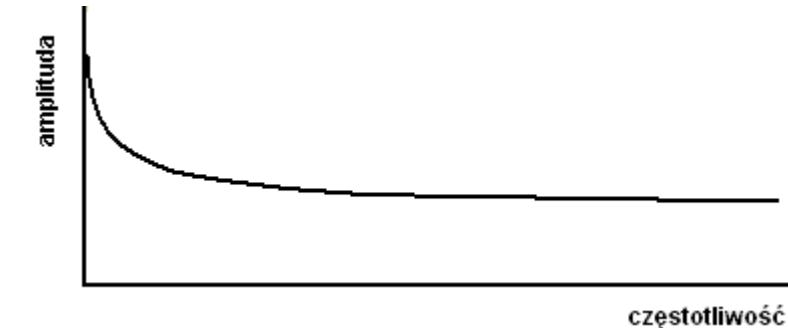
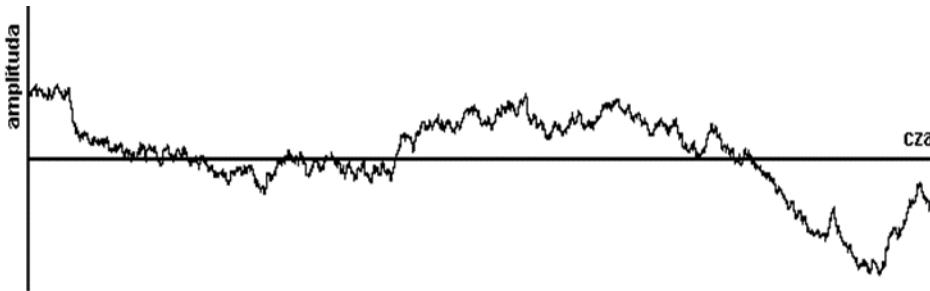
# Szum akustyczny

- Szum różowy to szum o właściwościach pośrednich między szumem białym a czerwonym.
- Widmowa gęstość mocy szumu różowego jest odwrotnie proporcjonalna do częstotliwości. Z tego względu szum różowy jest też nazywany szumem  $1/f$ .



# Szum akustyczny

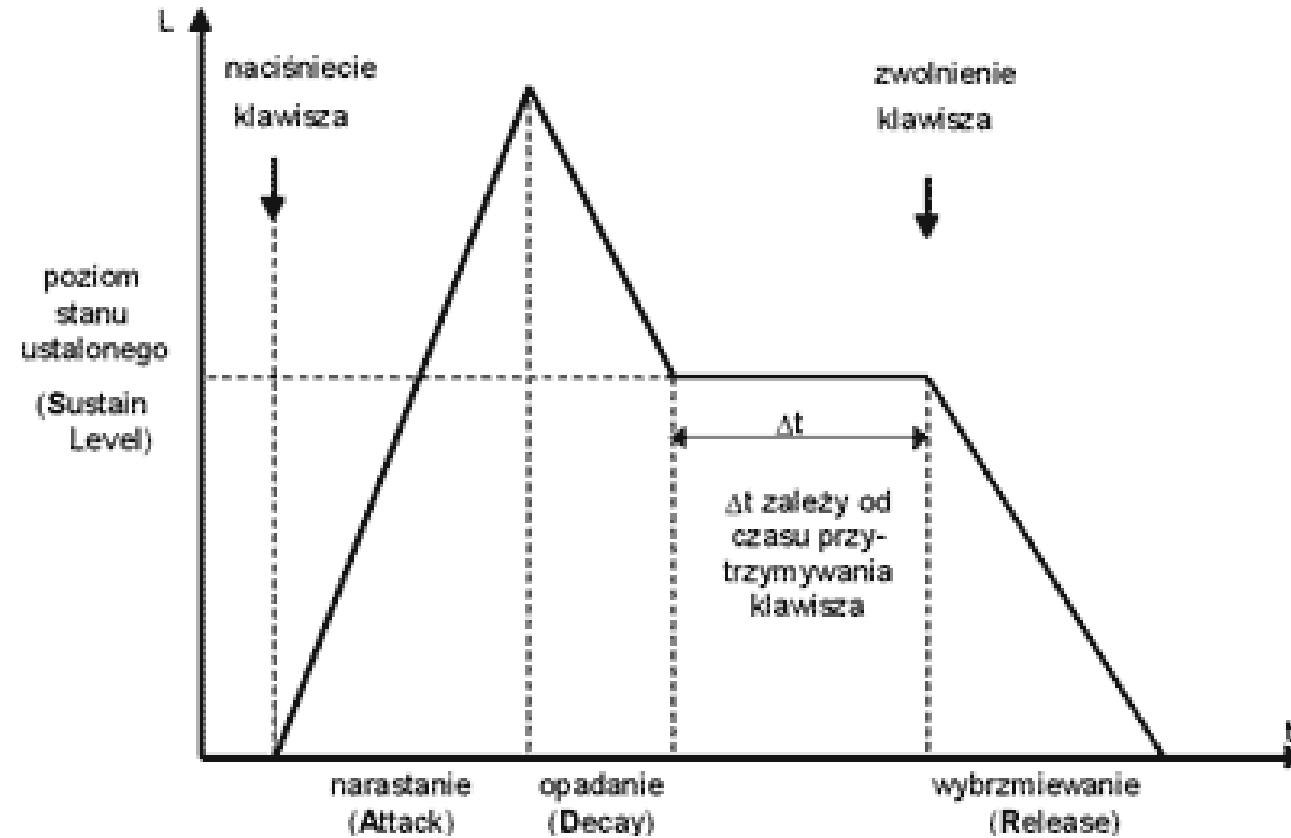
- Szum czerwony (szum brązowy, szum Browna) to rodzaj szumu wytwarzanego podczas ruchów Browna.
- Widmo szumu brązowego spada w miarę wzrostu częstotliwości odwrotnie proporcjonalnie do kwadratu tej częstotliwości tj. jego widmowa gęstość mocy proporcjonalna jest do  $1/f^2$



# Obwiednia dźwięku

- Obwiednia przedstawia zależność natężenia dźwięku (głośności) od czasu w przedziale od naciśnięcia klawisza, do jego puszczenia.
- W „typowej” obwiedni można wyróżnić 4 fazy:
  - Fazę narastania dźwięku (A - Attack), w której zwiększa on swoją głośność do maksimum,
  - Fazę opadania (D - Decay), w której dźwięk nieco zmniejsza swoją głośność,
  - Fazę ustaloną (S - Sustain), w której głośność przez pewien czas trwa na ustalonym, niezmienionym poziomie.
  - Fazę wybrzmiewania (R - Release), kiedy głośność maleje do 0.

# Obwiednia dźwięku

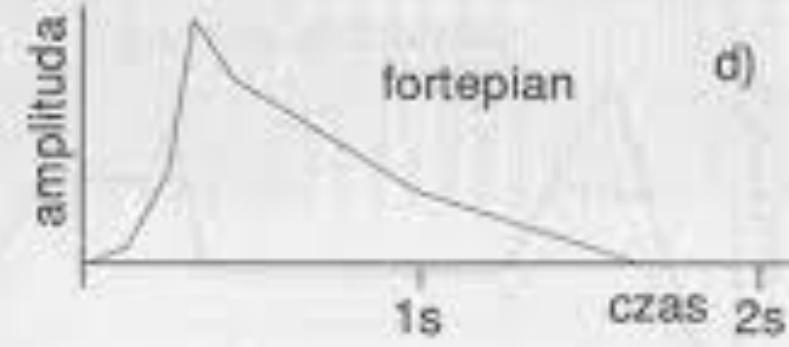
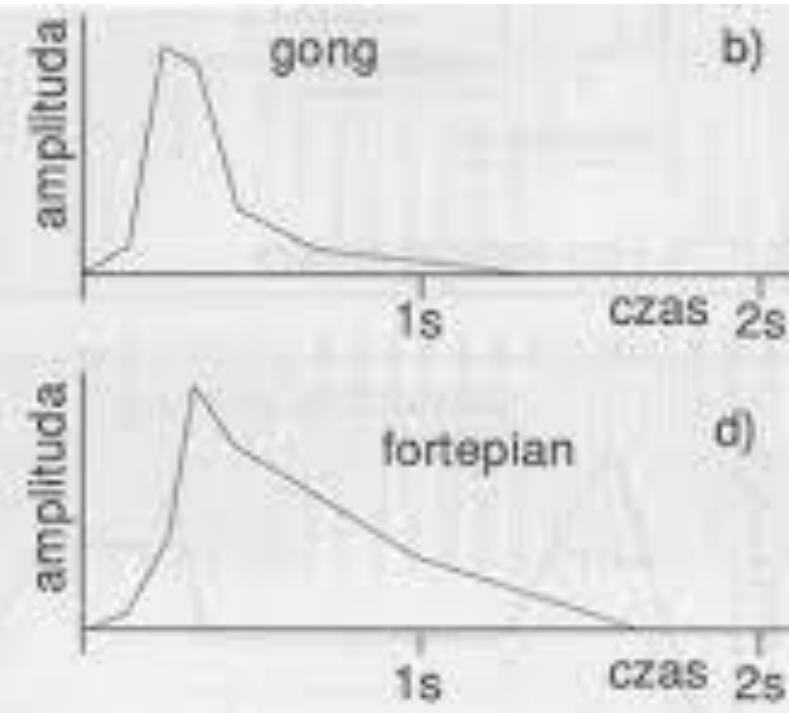
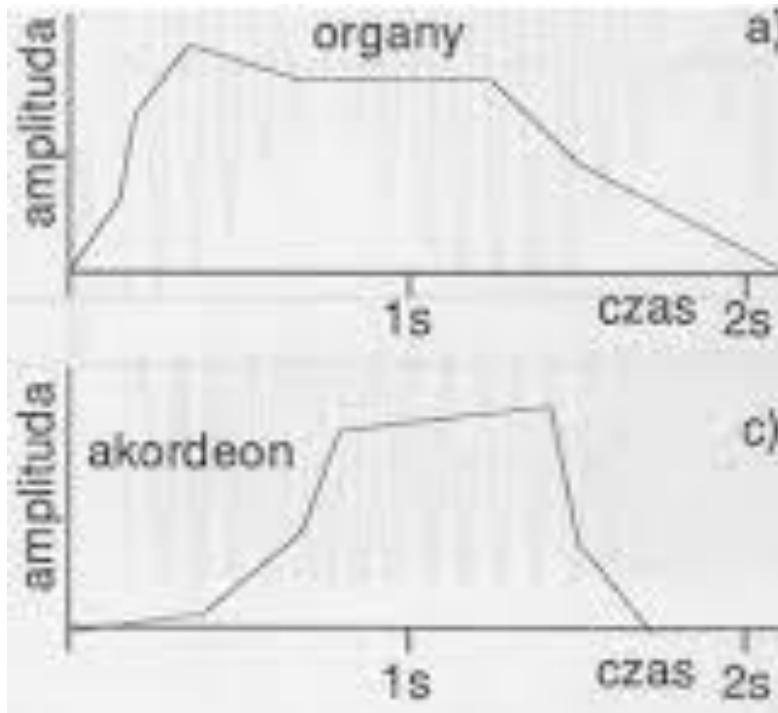


# Obwiednia dźwięku

- Czasy trwania każdej z faz obwiedni są charakterystyczne dla różnych instrumentów i dodatkowo (oprócz barwy) ułatwiają ich rozpoznawanie.
- Rozkład widmowy dźwięku dla danego instrumentu (jego barwa) może ulegać zmianie w każdej z faz ADSR.
- Jest to cecha, która w odniesieniu do dźwięków generowanych syntetycznie nadaje im dodatkowy koloryt i bardziej realistyczne brzmienie.
- Obwiednie instrumentów rzeczywistych mają różne kształty, często znacznie odbiegające od wzorcowego modelu ADSR.
- Zdarza się, że nie wszystkie fazy obwiedni ADSR występują w rzeczywistości, np. w obwiedni opadającej ADR nie występuje faza ustalona S, a dźwięk po osiągnięciu maksimum opada i wybrzmiewa.

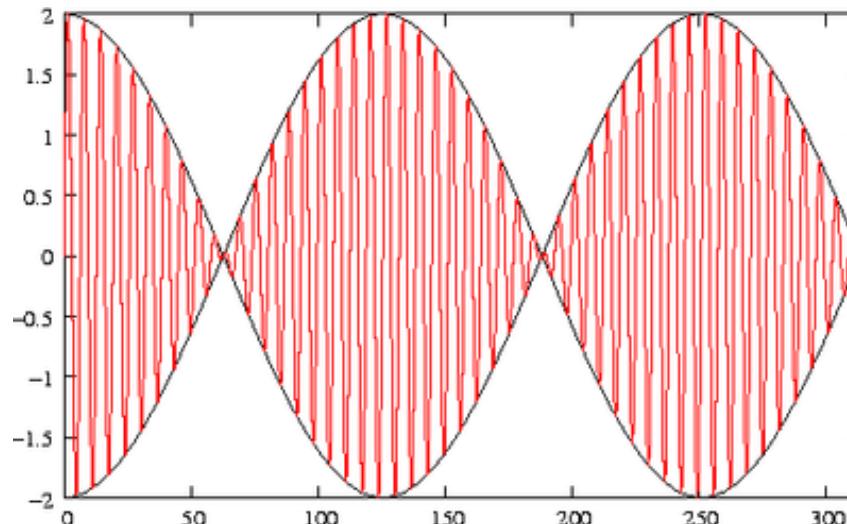
# Obwiednia dźwięku

- Przykłady obwiedni instrumentów muzycznych:



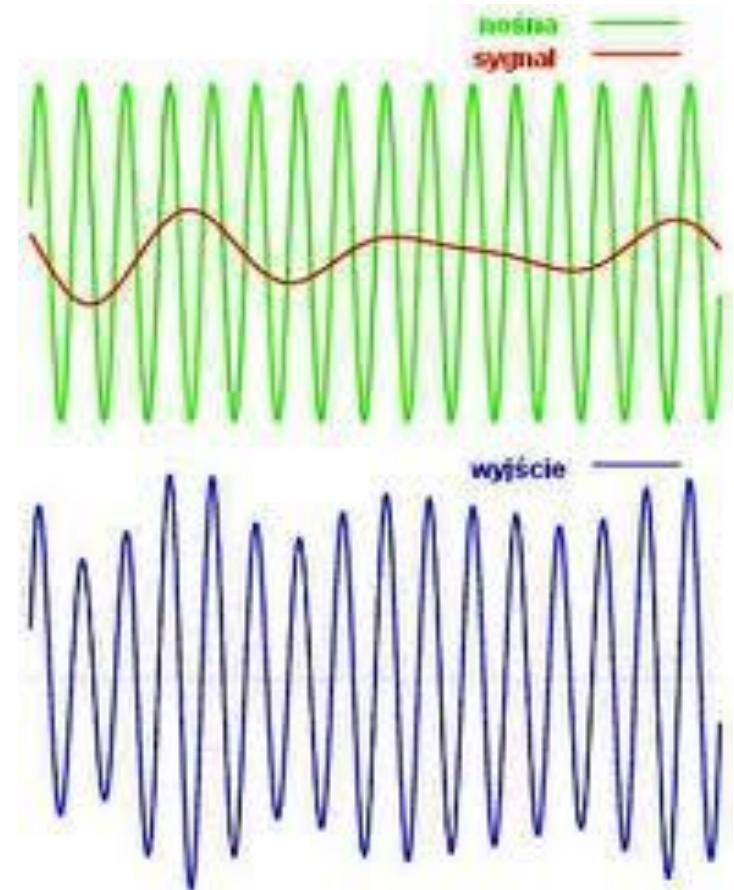
# Dudnienie

- Efekt dudnienia powstaje wówczas, gdy złożymy ze sobą dwa dźwięki o zbliżonych częstotliwościach  $f_1$  i  $f_2$ .
- Słyszmy wówczas dźwięk o niskiej częstotliwości równej różnicy  $f_1 - f_2$ .



# Modulacja amplitudy

- Jeżeli różnica między częstotliwościami nakładanych dźwięków jest nieco większa, to powstaje zjawisko modulacji amplitudy (AM) w czasie.
- Objawia się ono cykliczną zmianą głośności dźwięku.
- Wielkość zmian nazywana jest głębokością modulacji.



Fundusze  
Europejskie  
Wiedza Edukacja Rozwój



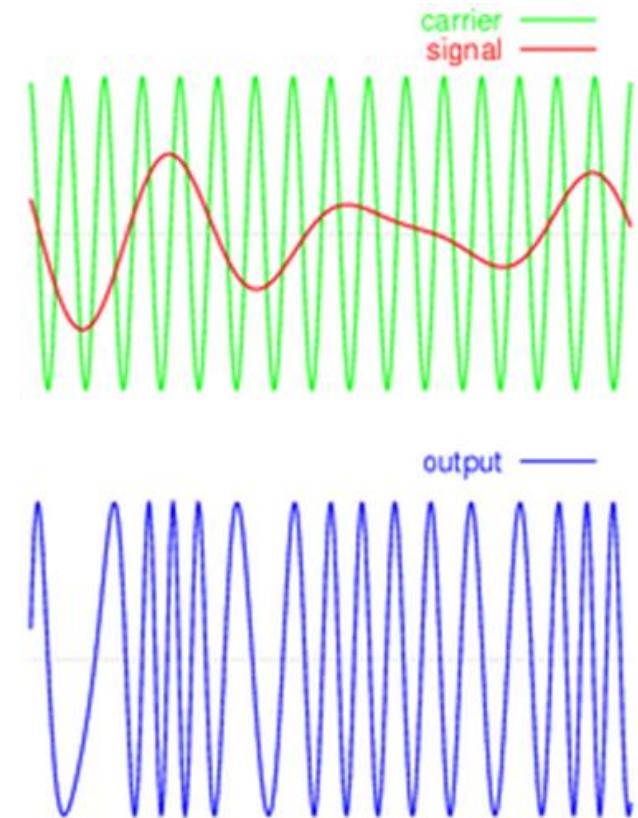
Politechnika Wrocławska

Unia Europejska  
Europejski Fundusz Społeczny



# Modulacja częstotliwości

- Modulacja częstotliwości (FM)  
polega na zmianie chwilowej  
częstotliwości sygnału nośnej  
w zależności od modulatora.
- W zakresie dźwięków słyszalnych  
modulacja częstotliwości objawia się  
charakterystyczną wibracją dźwięku





# Techniki przetwarzania mediów cyfrowych

Wykład 2: Cyfrowy zapis dźwięku (4 h)



Fundusze  
Europejskie  
Wiedza Edukacja Rozwój



Politechnika Wrocławска

Unia Europejska  
Europejski Fundusz Społeczny



# Cyfrowy zapis dźwięku

- Konwersja dźwięku analogowego na cyfrowy (dygitalizacja)
- Konwersja z sygnału akustycznego na sygnał elektryczny
- Próbkowanie
- Kwantowanie (kwantyzacja)
- Kodowanie



Fundusze  
Europejskie  
Wiedza Edukacja Rozwój



Politechnika Wrocławskiego

Unia Europejska  
Europejski Fundusz Społeczny



# Dźwięk cyfrowy

- Dźwięk cyfrowy jest sygnałem, którego dziedzina i zbiór wartości są dyskretne.
- Odpowiednikiem dźwięku cyfrowego jest dźwięk analogowy o ciągłej dziedzinie i zbiorze wartości, który może być zarówno pierwotny, jak i wtórny w stosunku do dźwięku cyfrowego.
- Dźwięk cyfrowy jest pierwotny w stosunku do analogowego, gdy jest on generowany cyfrowo. Wówczas odpowiadający mu dźwięk analogowy pojawia się w momencie odtwarzania wcześniej powstałego dźwięku cyfrowego.
- W przeciwnym przypadku, dźwięk cyfrowy powstaje w wyniku konwersji dźwięku analogowego.

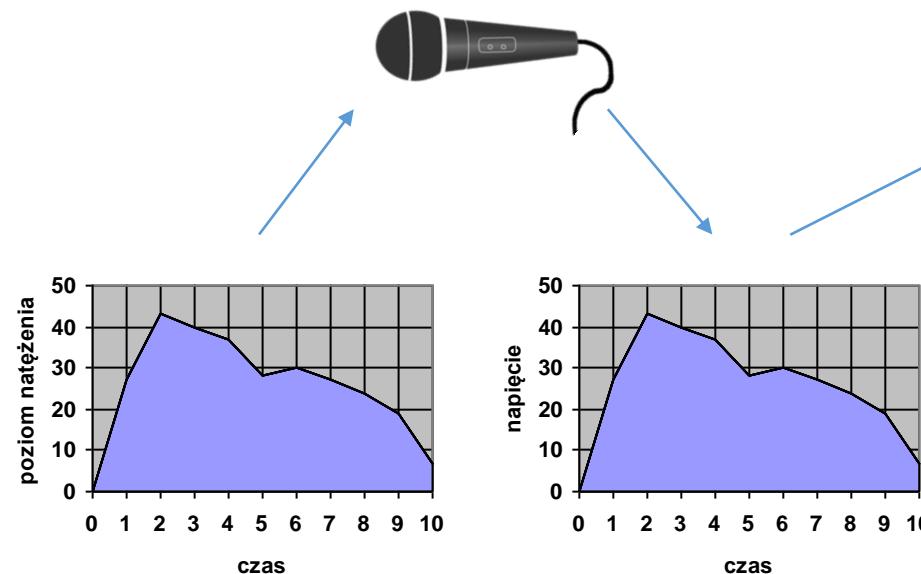
# Konwersja analogowo-cyfrowa

- Konwersja analogowo cyfrowa (A/C) polega na przekształceniu elektrycznego sygnału analogowego na jego postać cyfrową.
- Konwersja składa się z 3 podstawowych operacji:
  - ✓ Próbkowania
  - ✓ Kwantowania
  - ✓ Kodowania
- Gdy konwersji poddawany jest sygnał dźwiękowy (akustyczny), to przed dokonaniem konwersji należy go przekształcić do sygnału elektrycznego (również analogowego).
- Używa się do tego przetwornika akustyczno-elektrycznego, czyli mikrofonu.

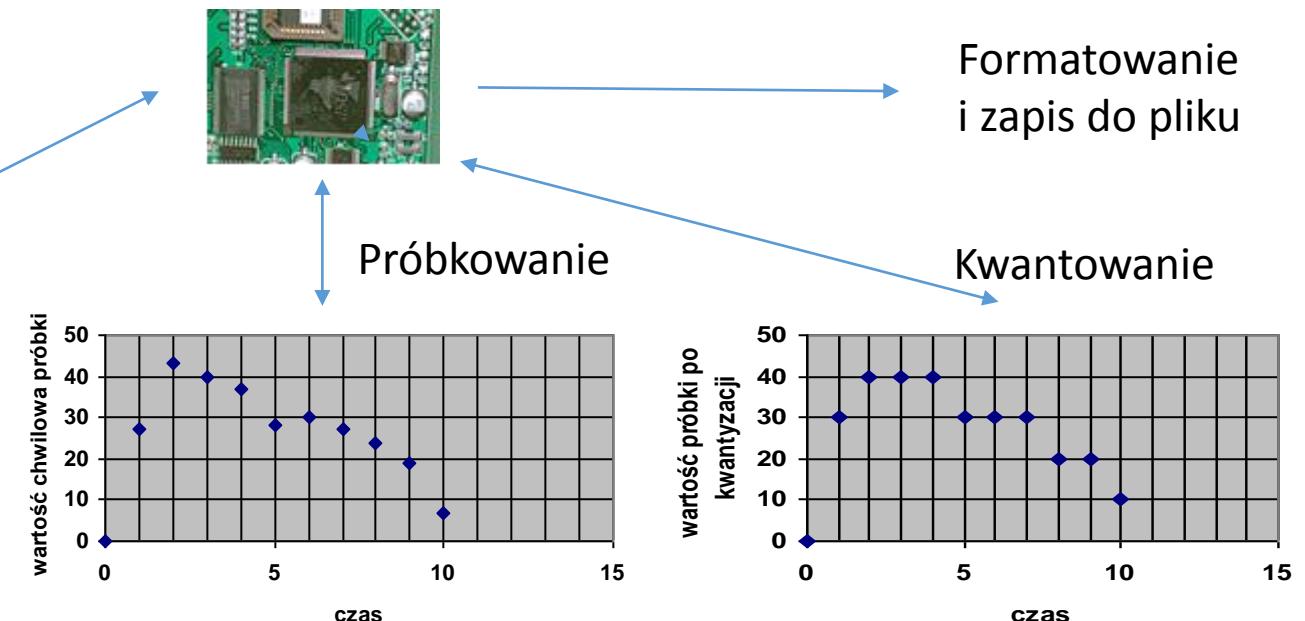


# Konwersja analogowo-cyfrowa

Konwersja akustyczno-elektryczna  
**MIKROFON**



Konwersja analogowo-cyfrowa (A/C)  
**PRZETWORNIK A/C**

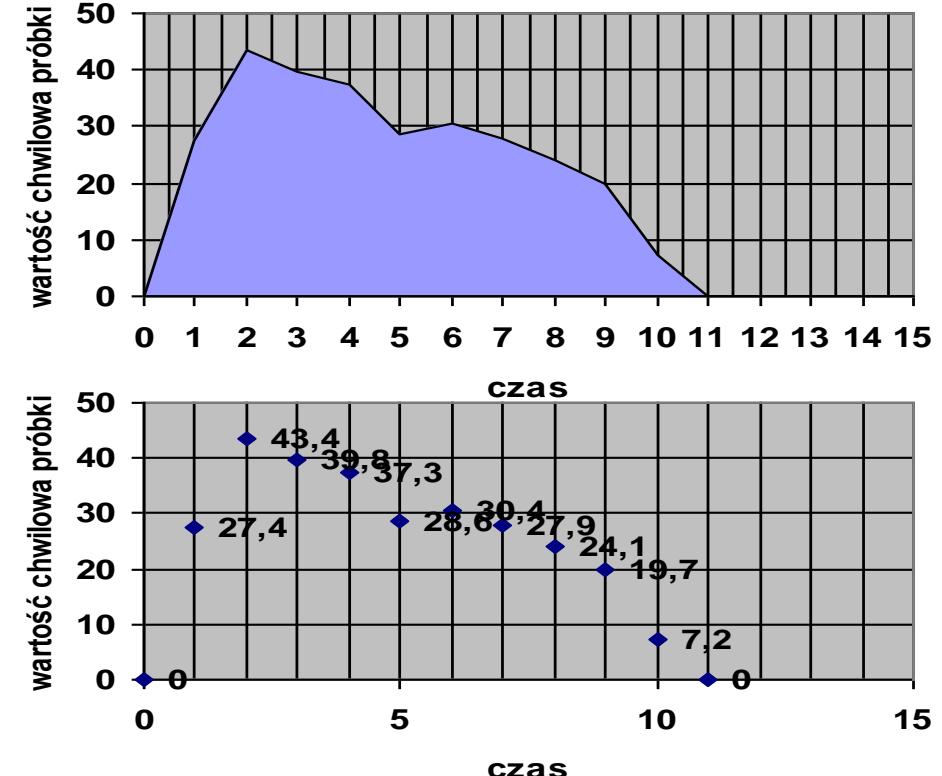


# Konwersja akustyczno-elektryczna

- Konwersja akustyczno-elektryczna odbywa się za pośrednictwem mikrofonu.
- Wejście dla mikrofonu stanowi analogowa fala dźwiękowa.
- Wyjściem mikrofonu jest analogowa „fala prądowa” reprezentowana przez zmiany natężenia lub napięcia prądu w czasie, w zależności od konstrukcji mikrofonu.
- Amplituda fali prądowej zmienia się proporcjonalnie do zmian amplitudy fali akustycznej na wejściu mikrofonu.
- Analogowy sygnał prądowy kierowany jest dalej do układu pamiętająco – próbkującego (przetwornika analogowo-cyfrowego).

# Próbkowanie

- Próbkowanie: procesem, w którym sygnał ciągły zamieniany jest na ciąg próbek o wartościach odpowiadających wartościom sygnału ciągłego w ścisłe określonych momentach czasowych.
- Momenty czasowe, w których odbywa się pomiar sygnału ciągłego określone są przez **częstotliwość próbkowania**, liczbę pomiarów wykonywanych w ciągu 1 s.



$$\begin{aligned}s(t_1) &= 0; s(t_2) = 27,4; s(t_3) = 43,4; s(t_4) = 39,8; s(t_5) = 37,3; \\ s(t_6) &= 28,6; s(t_7) = 30,4; s(t_8) = 24,1; s(t_9) = 19,7; s(t_{10}) = 7,2; \\ s(t_{11}) &= 0\end{aligned}$$

# Próbkowanie

- Z matematycznego punktu widzenia wynikowy spróbkowany sygnał  $f_s(t)$  jest iloczynem sygnału analogowego  $f(t)$  i funkcji próbkującej  $s(t)$ :

$$f_s(t) = f(t) * s(t)$$

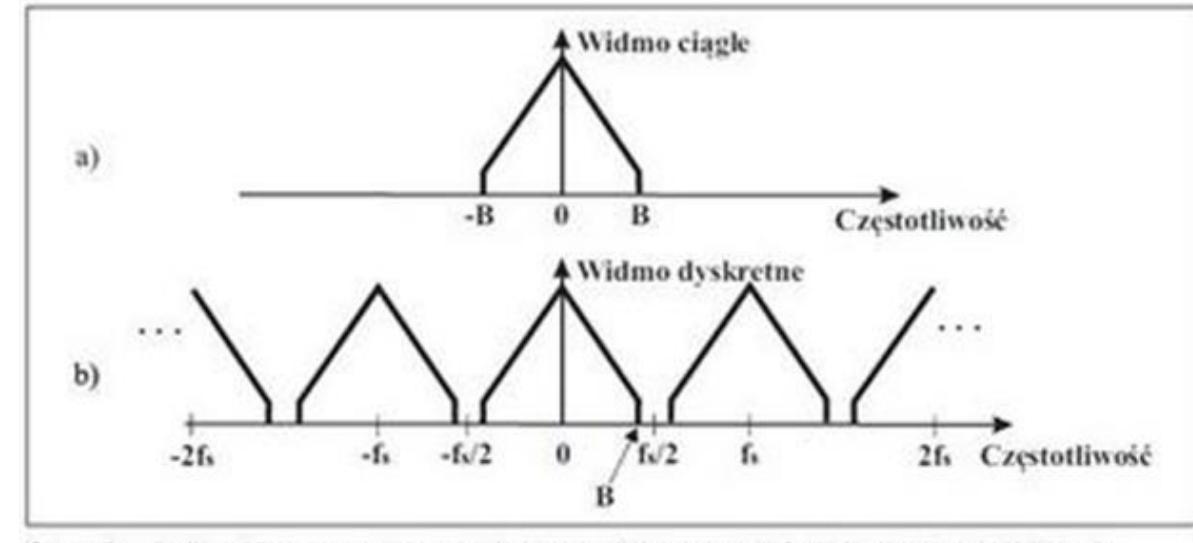
- W idealnym przypadku, gdyby pobieranie próbek odbywało się w nieskończonym krótkim czasie, funkcja próbująca byłaby ciągiem impulsów Diraca  $\delta(t)$  o okresie  $T_s$ , a wynikowy sygnał  $f_s(t)$  można wyznaczyć z zależności:

$$f_s(t) = f(t) \sum_{n=-\infty}^{n=+\infty} \delta(t - nT_s)$$

gdzie:  $n$  – indeks czasu (numer chwili czasu).

# Próbkowanie

Widma sygnału ciągłego i odpowiadającego mu sygnału spróbkowanego **nie są jednakowe**. W widmie sygnału dyskretnego oprócz widma podstawowego pojawiają się jego repliki występujące wokół wielokrotności częstotliwości próbkowania  $f_s$  (tzw. wstęgi boczne) Próbkowanie powoduje powstawanie kopii widma oddalonych od siebie o odległość równą częstotliwości próbkowania.



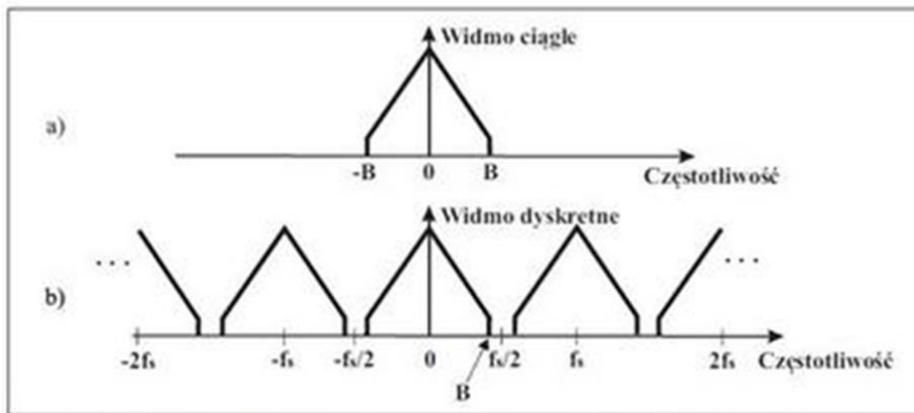
Rysunek 1. Analiza widmowa sygnału o ograniczonym paśmie: a) sygnał ciągły, b) sygnał po próbkowaniu.

# Próbkowanie

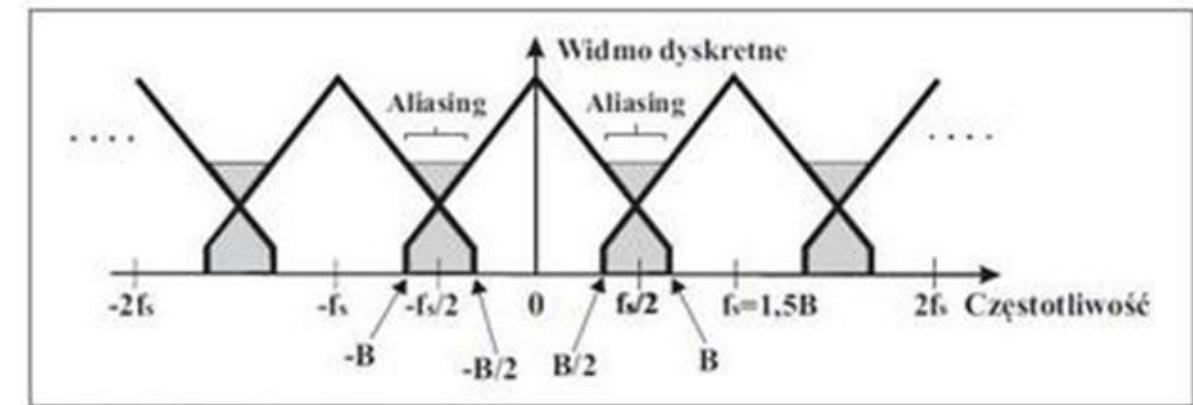
- Próbkowanie jest procesem w pełni odwracalnym pod warunkiem, że w trakcie procesu nie pojawią się błędy.
- W rzeczywistości mamy do czynienia z dwoma rodzajami błędów, które w tym procesie występują:
  - Błędy, które można wyeliminować,
  - Błędy, które można jedynie zminimalizować
- Do pierwszej grupy należy zjawisko aliasingu.
- Ogólnie mówiąc polega ono na nakładaniu widm.

# Próbkowanie

- Sygnał na rys. a) został spróbkowany z częstotliwością  $f_s$
- Sukcesywne zmniejszanie częstotliwości próbkowania spowoduje, że widma widma zaczną się ze sobą stykać.
- Dalsze zmniejszanie tej częstotliwości spowoduje, że widma zaczną nachodzić na siebie.
- W efekcie w zakresie wspólnym nie jesteśmy w stanie określić jaką wartość przyjmie sygnał.



Rysunek 1. Analiza widmowa sygnału o ograniczonym paśmie: a) sygnał ciągły, b) sygnał po próbkowaniu.

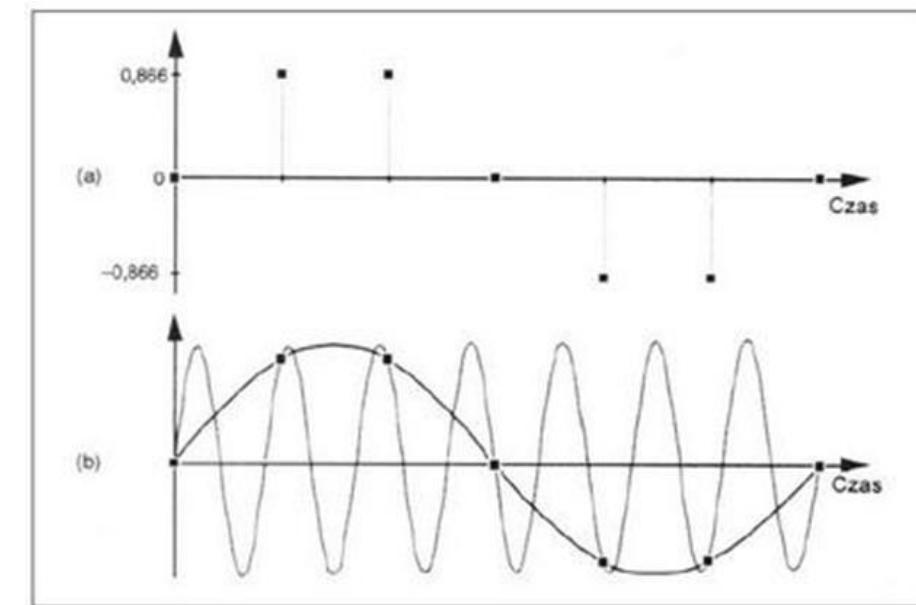


Rysunek 2. Aliasing, czyli nakładanie się widm.



# Próbkowanie

- Zjawisko alisingu można też zilustrować posługując się obrazem dźwięku w dziedzinie czasu.
- Jeśli po próbkowaniu uzyskaliśmy sygnał dyskretny jak na rys. a), to można przez nie przeprowadzić praktycznie nieskończonie wiele sygnałów sinusoidalnych (rys. b).
- Wniosek: na podstawie samych wartości próbek nie można jednoznacznie określić częstotliwości tego sygnału



Rysunek 3. Niejednoznaczność w procesie próbkowania.

# Próbkowanie

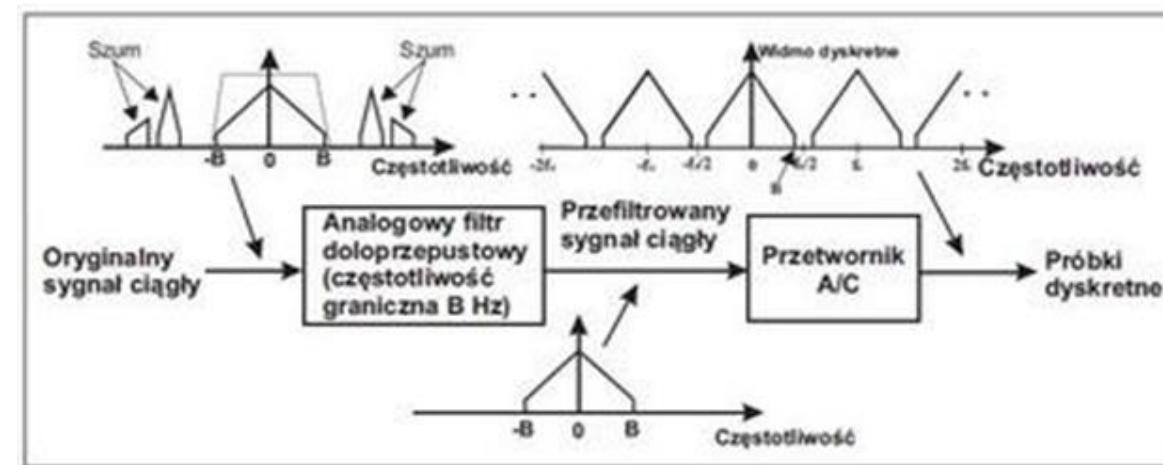
- Aby dźwięk po spróbkowaniu brzmiał czysto i bez zniekształceń, tj. wyeliminować niejednoznaczności, częstotliwość próbkowania musi spełniać warunki twierdzenia Shannona-Kotielnikowa
- Twierdzenie o próbkowaniu jest podstawowym twierdzeniem w dziedzinie teorii informacji, w szczególności w telekomunikacji i przetwarzaniu sygnałów.
- Tw. O próbkowaniu (w wersji zaproponowanej przez Claude'a Shannona):  
Jeśli funkcja  $x(t)$  nie zawiera częstotliwości wyższych niż  $B$  Hz, to można ją完全nie zdefiniować podając jej rzędne jako ciąg punktów oddalonych od siebie o  $1/(2B)$  sekund.
- $B$  nosi nazwę granicy Nyquista (częstotliwości Nyquista)

# Próbkowanie

- Twierdzenie o próbkowaniu mówi o tym, że dowolny pasmowo ograniczony sygnał analogowy może być doskonale odtworzony na podstawie nieskończego długiej sekwencji próbek, jeśli częstotliwość próbkowania przekracza  $2B$  próbek na sekundę, gdzie  $B$  jest najwyższą częstotliwością zawartą w sygnale oryginalnym.
- Nowsze sformułowania twierdzenia mówią o tym, że częstotliwość próbkowania powinna być większa lub równa  $2B$ , przy czym w sygnale oryginalnym nie może występować stała składowa sinusoidalna o częstotliwości dokładnie równej  $B$ .
- Jeśli sygnał zawiera składową o częstotliwości  $B$  Hz, to próbki pobrane dokładnie co  $\frac{1}{2B}$  sekund nie pozwolą na kompletne odtworzenie sygnału.

# Próbkowanie

- Aby nie dopuścić do powstania aliasingu, należy ograniczyć pasmo częstotliwości sygnału użytecznego stosując filtr dolnoprzepustowy, tzw. filtr antyaliasingowy
- Na przykład, jeśli mamy do przetworzenia dźwięk w zakresie słyszalnym dla ludzi, tj. 20 Hz do 20 kHz, to przed próbkowaniem należy zastosować filtr o częstotliwości odcięcia na poziomie 20 kHz, a samo próbkowanie przeprowadzić z częstotliwością większą niż 20 kHz (stąd popularna częstotliwość próbkowania 44,1 kHz)
- Taki zabieg pozwala również na wyeliminowanie z nagrania sygnałów nieużytecznych, np. szumów



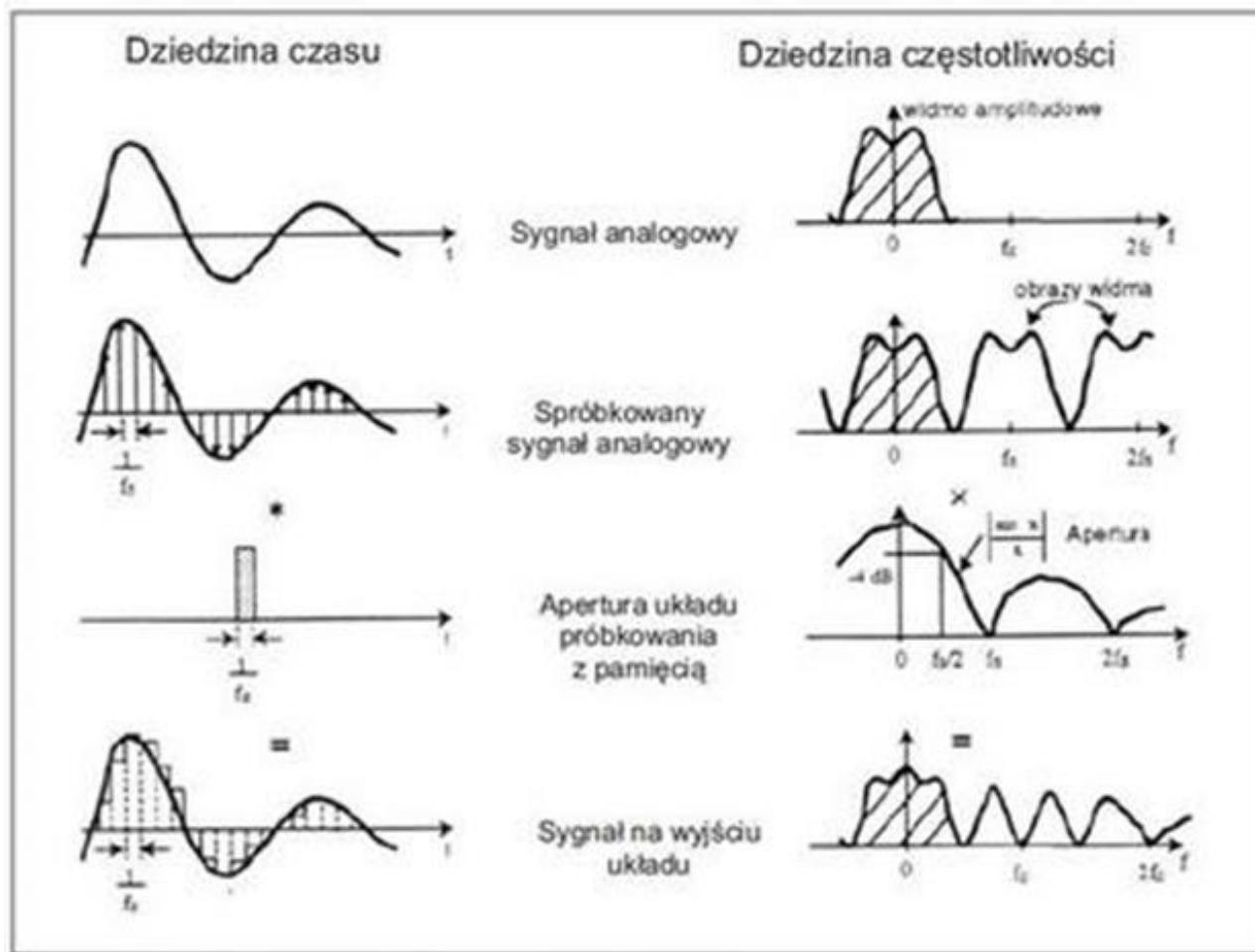
Rysunek 6. Analogowa filtracja dolnoprzepustowa poprzedzająca próbkowanie.



# Próbkowanie

- Mówiąc o błędach próbkowania, których nie można wyeliminować, należy zwrócić uwagę, że twierdzenie o próbkowaniu zakłada pewną idealizację świata rzeczywistego.
- Zakłada ono, że czas pobierania próbki jest zerowy.
- W rzeczywistości funkcja próbkująca  $s(t)$  jest falą prostokątną o niezerowym współczynniku wypełnienia.
- W efekcie w sygnale spróbkowanym na powielenia widma sygnału nakładać się będzie dodatkowo widmo okna prostokątnego, które ma postać funkcji sinc x.
- To powoduje powstawanie zniekształceń w wynikowym sygnale cyfrowym.
- Zniekształcenia te nazywa się **zniekształceniami apertury** (niezerowego czasu pobierania próbki).

# Próbkowanie



Rysunek 7. Ilustracja skutków niezerowego czasu pobierania próbek.



Fundusze  
Europejskie  
Wiedza Edukacja Rozwój



Politechnika Wrocławska

Unia Europejska  
Europejski Fundusz Społeczny



# Próbkowanie

- Zniekształcenia apretury są trudne do wyeliminowania, ponieważ nawet jeśli ograniczymy czas pobierania próbki do minimum, to i tak konieczne jest „zatrzymanie” sygnału w fazie pamiętania ze względu na prawidłową pracę kwantyzera.
- Aby zminimalizować zniekształcenia apretury należy dążyć do jak najkrótszego czasu przetwarzania sygnału przez konwerter analogowo-cyfrowy.
- Skutkiem tego błędu można również przeciwdziałać (do pewnego stopnia), wprowadzając filtr korygujący zniekształcenia typu sinc x.

# Próbkowanie

- Innym rodzajem błędów próbkowania jest efekt niestabilności częstotliwości próbkowania, czyli tzw. jitter.
- W praktycznych układach proces próbkowania jest synchronizowany zegarem kwarcowym gwarantującym stabilność częstotliwości próbkowania.
- Mimo to mogą wystąpić takie sytuacje, jak np. stany przejściowe, związane z konwersją formatów, w których trzeba uwzględnić to zjawisko.
- Słyszalnym efektem zniekształcenia jitter jest wzrost szumów, nawet o kilka dB.

# Próbkowanie

- Próbkowanie może dotyczyć również sygnałów dyskretnych, czyli takich, które zostały wcześniej spróbkowane.
- Próbkowanie sygnałów dyskretnych wykorzystuje się np. w celu zwiększenia stosunku sygnału do szum w paśmie użytecznym czy w konwersji plików w różnych standardach kodowania.
- Wyróżnia się dwie sytuacje w zależności od relacji między nową częstotliwością próbkowania a częstotliwością próbkowania sygnału pierwotnego:
  - nowa częstotliwość próbkowania < częstotliwością próbkowania sygnału pierwotnego
  - nowa częstotliwość próbkowania > częstotliwością próbkowania sygnału pierwotnego

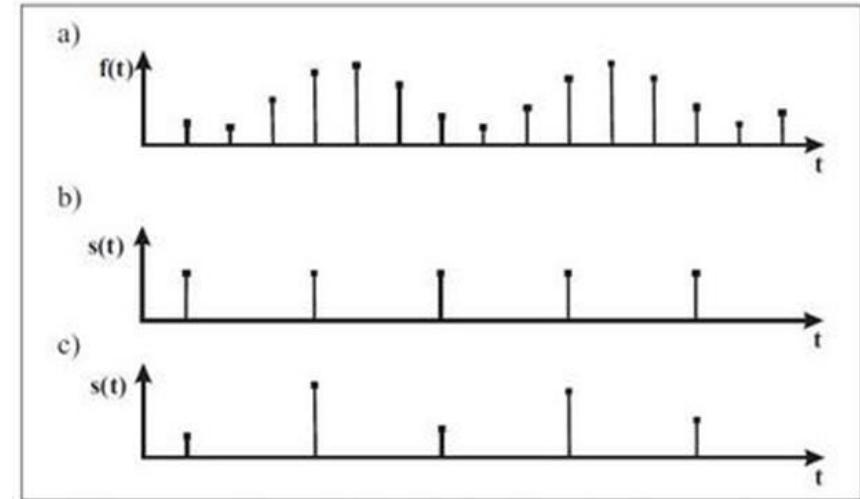
# Próbkowanie

W pierwszym przypadku mamy do czynienia ze zjawiskiem zwanym **decymacją sygnału**.

Decymacja to zmniejszenie liczby próbek w ciągu wynikowym, które osiąga się poprzez próbkowanie sygnału pierwotnego z częstotliwością niższą od pierwotnej.

Praktycznie uzyskuje się sygnał w postaci próbek równych próbkom oryginalnego sygnału w chwilach czasu będących wielokrotnościami okresu próbkowania  $T_s$ .

W pozostałych chwilach czasu sygnał będzie przyjmował wartości zerowe.



Rysunek 8. Decymacja sygnału: a) sygnał oryginalny, b) funkcja próbująca, c) wynikowy ciąg próbek.



Fundusze  
Europejskie  
Wiedza Edukacja Rozwój



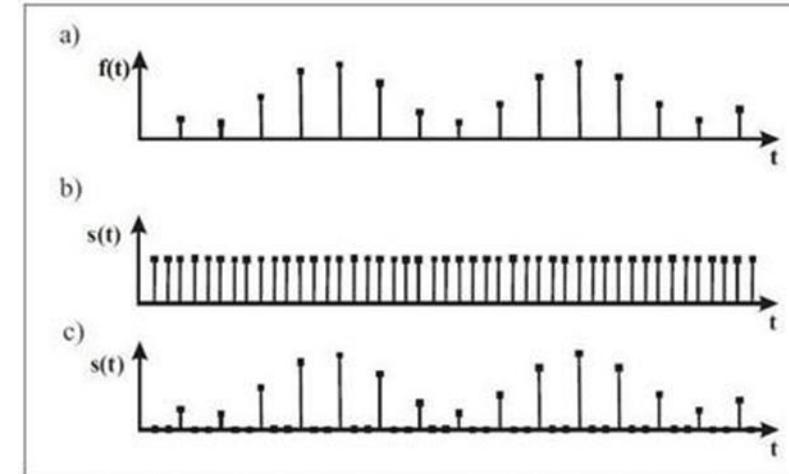
Politechnika Wrocławska

Unia Europejska  
Europejski Fundusz Społeczny



# Próbkowanie

- Do zwiększenia częstotliwości próbkowania sygnału wynikowego w stosunku do pierwotnego wykorzystuje się **interpolację sygnału**.
- W praktyce osiąga się to poprzez wstawienie pomiędzy próbki oryginalne próbek zerowych w odstępach równych okresowi próbkowania  $T_s = 1/fs$ .



Rysunek 9. Interpolacja sygnału: a) sygnał oryginalny, b) funkcja próbująca, c) wynikowy ciąg próbek.

# Próbkowanie

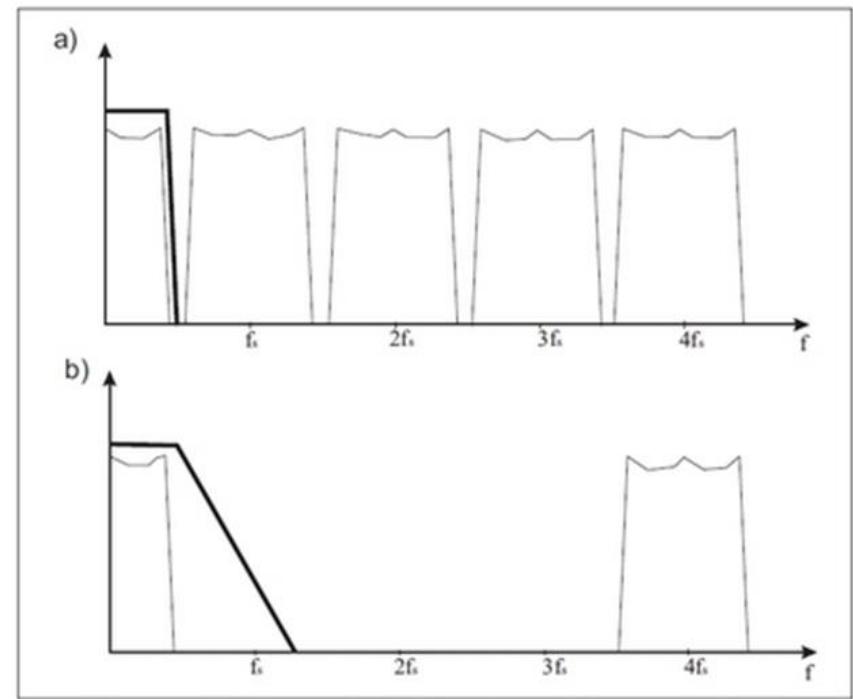
- Ze zjawiskiem interpolacji sygnału związane jest tzw. nadpróbkowanie (ang. Oversampling)
- Nadpróbkowanie to taki proces próbkowania sygnałów dyskretnych, w którym częstotliwość próbkowania  $f_s >> 2f_M$ , tj. górnej granicy częstotliwości widma sygnału.
- Nadpróbkowanie wykonuje się po przetworzeniu sygnału analogowego na cyfrowy za pomocą standardowego konwertera A/C, np. 44,1 kHz.
- Tak powstający sygnał dyskretny poddaje się interpolacji, czyli wstawieniu próbek zerowych z odpowiednim współczynnikiem interpolacji  $M$ . Współczynnik interpolacji mówi ile razy zwiększyła się częstotliwość próbkowania w stosunku do pierwotnej.
- Liczba  $M-1$  oznacza praktycznie liczbę próbek zerowych wstawionych pomiędzy próbками oryginalnymi.

# Próbkowanie

- W kolejnym kroku należy taki sygnał ze wstawionymi próbkami zerowymi, „przepuścić” przez filtr interpolacyjny, który wyliczy wartości próbek pośrednich pomiędzy tymi, które mamy już dane.
- Nadpróbkowanie stosuje się głównie po to, by można było złagodzić wymagania w stosunku do stromości zboczy filtrów dolnoprzepustowych.
- Jeśli stosujemy nadpróbkowanie widma „odsuwają” się od siebie, dzięki czemu nie musimy stosować trudnego do zrealizowania filtru o niemal prostopadłych zboczach.
- W zamian można zastosować taki, który ma o wiele łagodniejsze zbocza, a który będzie znacznie łatwiejszy do zrealizowania.

# Próbkowanie

Rysunek a) przedstawia sygnał spróbkowany z częstotliwością  $f_s$  oraz wymagany w takim przypadku filtr dolnoprzepustowy. Na rysunku b) widnieje sygnał spróbkowany z częstotliwością  $4f_s$  (tj. skonwertowany z postaci analogowej na cyfrową z częstotliwością  $f_s$  i poddany czterokrotnemu nadpróbkowaniu) oraz odpowiadający mu filtr dolnoprzepustowy.

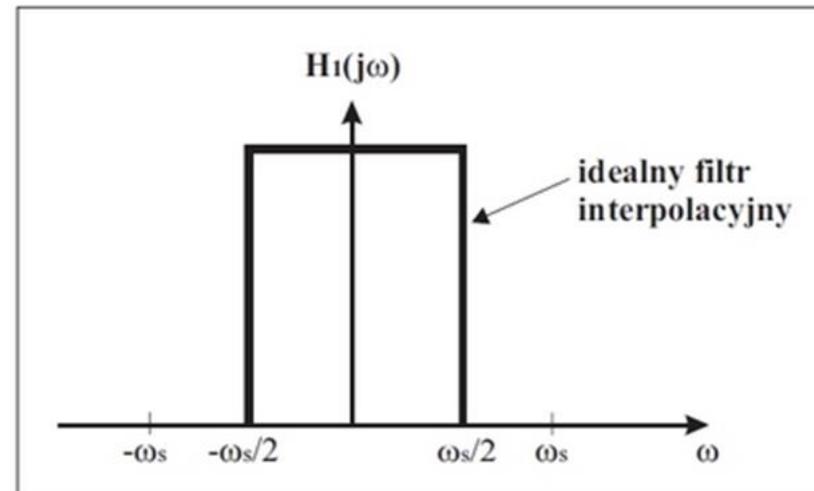


Rysunek 10. a) próbkowanie z częstotliwością  $f_s$ ; b) próbkowanie z  $4f_s$  (czterokrotny oversampling) – linią pogrubioną narysowano wymagany filtr antyalisingowy.



# Próbkowanie

- Rekonstrukcja sygnału ciągłego  $f(t)$  na podstawie ciągu próbek dyskretnych  $x(nT)$  pobranych w odstępach  $nT$  sekund odbywa się poprzez interpolację
- Aby odtworzyć sygnał ciągły na podstawie ciągu próbek należy przefiltrować sygnał spróbkowany za pomocą idealnego filtru interpolacyjnego o charakterystyce równej 1 w paśmie przepustowym i 0 poza tym pasmem.



Rysunek 11. Idealny filtr interpolacyjny.



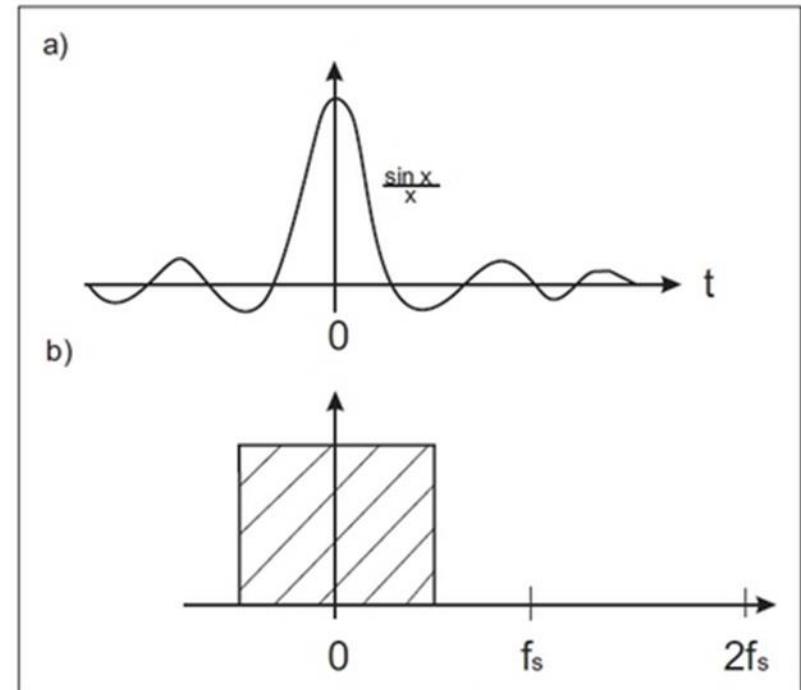
# Próbkowanie

- Filtracja powoduje interpolację przebiegu czasowego sygnału w obszarach między próbkami za pomocą funkcji  $sinc(t)$ .

- Funkcja  $sinc(t)$  wyraża się wzorem:

$$sinc(t) = \begin{cases} \frac{\sin(\pi t)}{\pi t} & \text{dla } t \neq 0 \\ 1 & \text{dla } t = 0 \end{cases}$$

- Widmem funkcji  $sinc(t)$  jest funkcja prostokątna i odwrotnie, widmem funkcji prostokątnej jest funkcja  $sinc(t)$



Rysunek 12. Funkcja  $sinc x$ : a) postać czasowa, b) widmo.

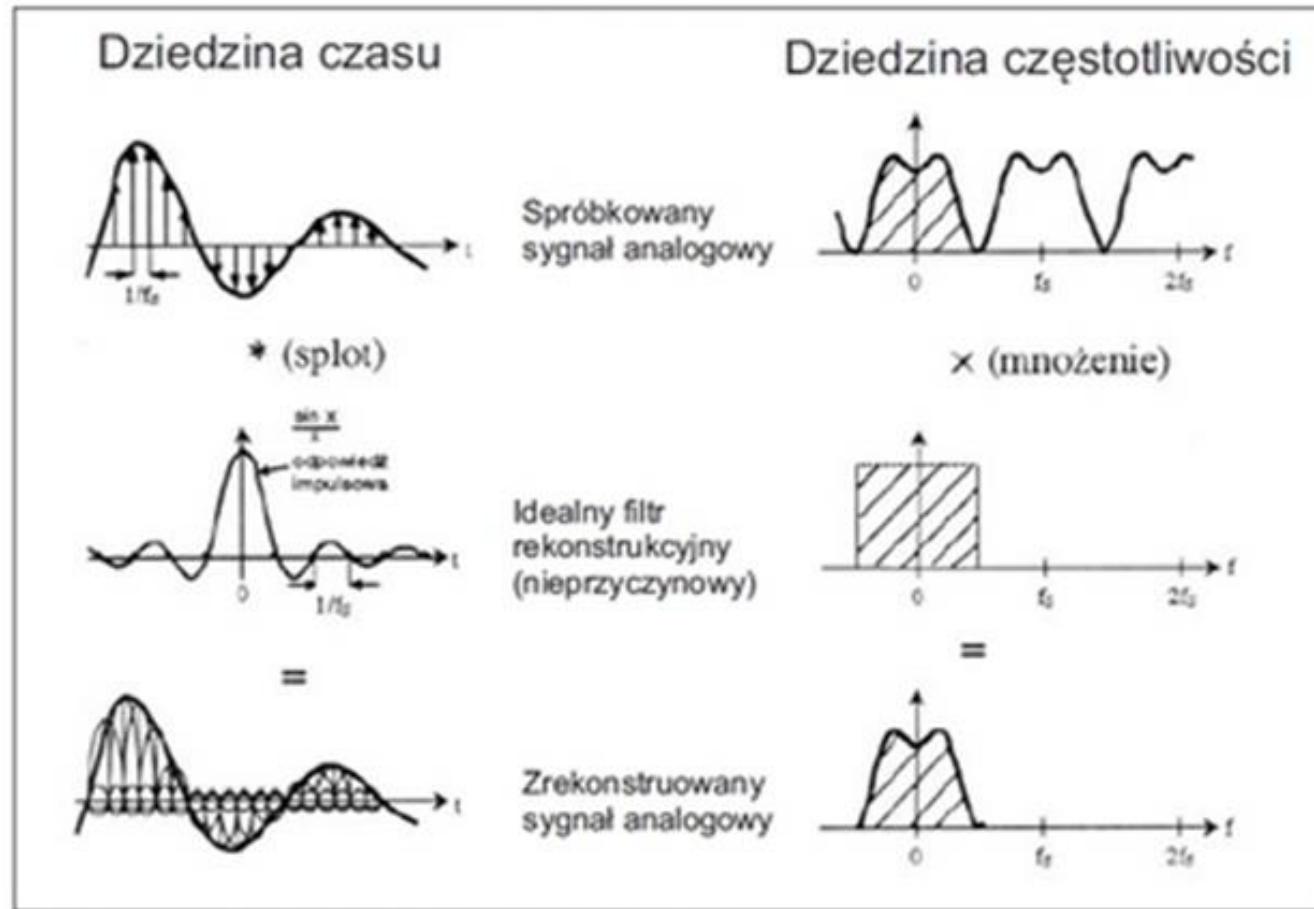
# Próbkowanie

Procedura rekonstrukcji sygnału przebiega następujących dwóch krokach:

- Wartość każdej próbki jest mnożona przez funkcję wartość funkcji  $\text{sinc}(t)$ , która jest wyskalowana w taki sposób, by jej przejścia przez zero występowały w miejscach występowania próbek, zaś maksimum funkcji jest przesunięte tak, by występowało dokładnie w czasie wystąpienia próbki  $nT$ .
- Tak wyznaczone funkcje są następnie dodawane do siebie.
- Ponieważ poszczególne funkcje mają charakter ciągły, to i wynik sumowania jest funkcją ciągłą.



# Próbkowanie



Rysunek 14. Proces interpolacji sygnału na podstawie próbek.

# Próbkowanie

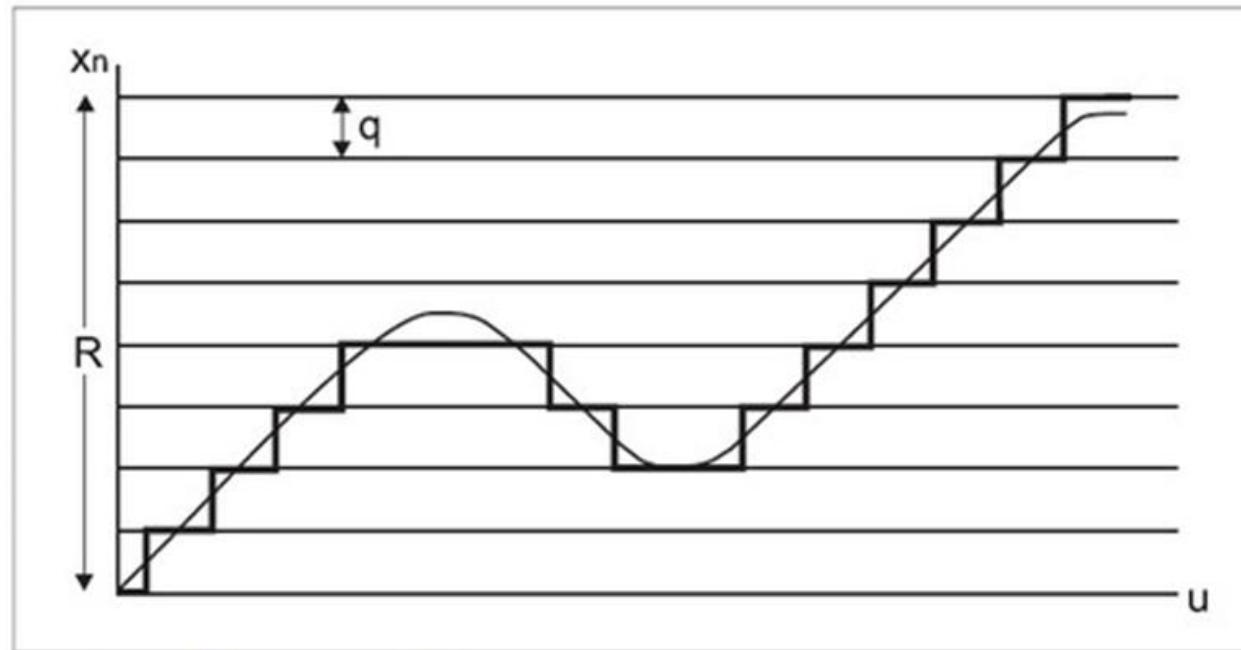
- Funkcja  $\text{sinc}(t)$  jest niemożliwa do zrealizowania w praktyce ze względu na swoją nieprzyczynowość – funkcja jest określona dla  $t < 0$ , czyli czasu ujemnego
- Dlatego zastępuje się ją inną funkcją nieidealną o wartościach określonych dla  $t \geq 0$
- Powoduje to powstawanie w wynikowym sygnale ciągłym dodatkowych zniekształceń, które częściowo można zniwelować stosując odpowiednie filtry.
- Na ogół zniekształcenia tego rodzaju są niesłyszalne.

# Kwantowanie

- Kwantowanie to proces, który polega na odwzorowaniu rzeczywistych wartości próbek pobranych podczas próbkowania na zbiór wartości dyskretnych, możliwych do zapisania na urządzeniach cyfrowych.
- Liczność zbioru wartości dyskretnych zależy od liczby bitów  $n$  przeznaczonych na zapisanie wartości pojedynczej próbki dźwięku.
- Na przykład, przy zapisie 8 bitowym wartości próbek można odwzorować na  $2^8 = 256$  liczb dyskretnych. Stosując zapis 16 bitowy liczba możliwości wzrasta do  $2^{16} = 65536$
- Liczba bitów  $n$  decyduje o tzw. rozdzielczości kwantowania (inaczej liczbie poziomów kwantowania)

# Kwantowanie

- Najstarszą metodą kwantowania jest metoda kwantowania równomiernego, zwana również modulacją kodowo-impulsową (PCM).
- Polega ona na podzieleniu całego zakresu, w jakim może zmieniać się sygnał wejściowy  $R$ , na  $k$  równomiernie rozłożonych przedziałów kwantowania o stałych szerokościach  $q$ .



Rysunek 1. Kwantowanie równomierne.



Fundusze  
Europejskie  
Wiedza Edukacja Rozwój



Politechnika Wrocławska

Unia Europejska  
Europejski Fundusz Społeczny



# Kwantowanie

- Jeśli do reprezentowania poszczególnych wartości próbek stosuje się zapis  $n$  bitowy, to szerokość przedziału  $q$  na skali kwantowania można obliczyć z zależności:

$$q = \frac{R}{2^n}$$

gdzie:  $R$  - maksymalny zakres przetwarzania kwantyzera

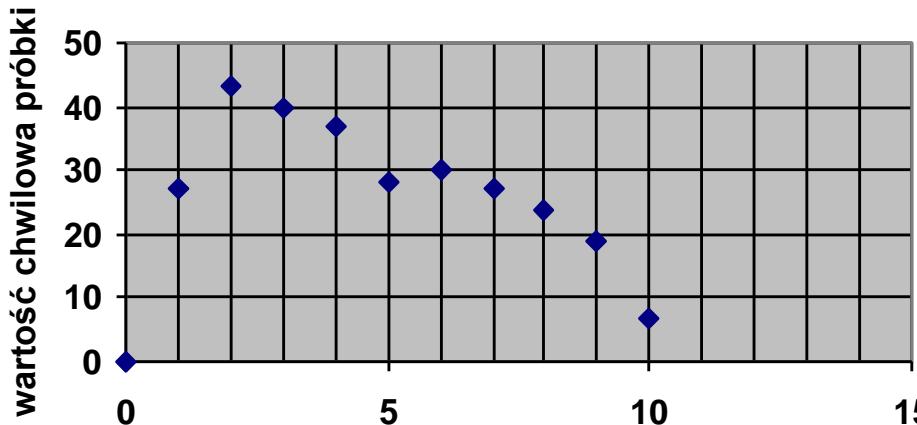
- Zakres dynamiki skwantowanego sygnału dźwiękowego, wyrażony w skali logarytmicznej, określa zależność:

$$D_Q = 20 \log \frac{R}{q} = 20 \log 2^n = 20 \cdot n \cdot \log 2 \cong 6,02 \cdot n$$

- Zakres dynamiki dźwięku zależy od rozdzielczości kwantyzera - im jest ona większa, tym większy jest zakres.

# Kwantowanie

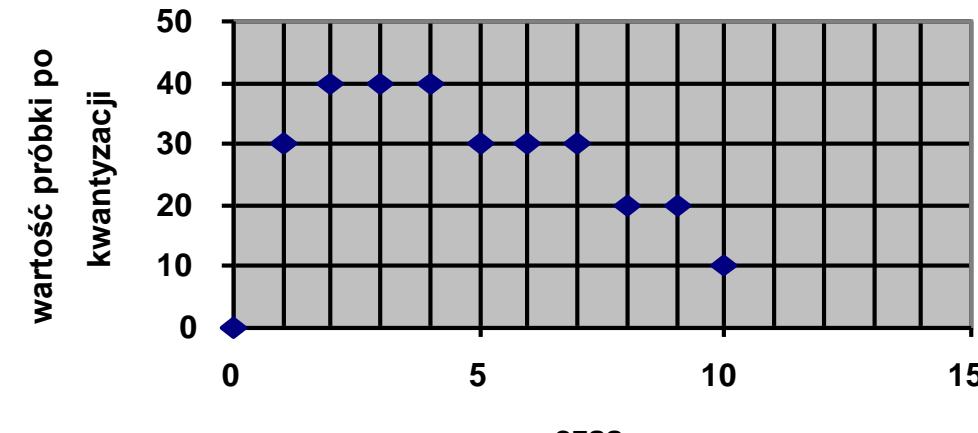
- W przeciwieństwie do próbkowania, kwantowanie jest procesem nieodwracalnym (nie jest możliwa idealna rekonstrukcja sygnału).
- Podczas kwantowania każdej wartości chwilowej próbki przypisuje się najbliższą jej liczbę dyskretną (często z zaokrągleniem w górę).
- W ten sposób pojedyncza liczba dyskretna  $x$  reprezentuje wszystkie chwilowe wartości próbek należące do przedziału  $\left(x - \frac{q}{2}, x + \frac{q}{2}\right)$ .



Fundusze  
Europejskie  
Wiedza Edukacja Rozwój



Politechnika Wrocławskiego



Unia Europejska  
Europejski Fundusz Społeczny



# Kwantowanie

- Przybliżenie wartości chwilowej liczbą dyskretną wprowadza do sygnału skwantowanego błąd zwany szumem kwantyzacji.
- Wykres sygnału skwantowanego ma kształt „schodkowy”.
- Widmo takiego sygnału zawiera wyższe składowe harmoniczne, które w praktyce powodują dodatkowe zniekształcenia nieliniowe.
- Ogólnie błędy kwantyzacji mogą być zamodelowane jako źródło szumu dodawanego do sygnału.
- Dla przyjętego modelu szumu wzór na stosunek sygnału użytecznego do szumu SNR (Signal to Noise Ratio) wyraża się wzorem:

$$\frac{S}{N} \cong 6 \cdot n + 1,8$$



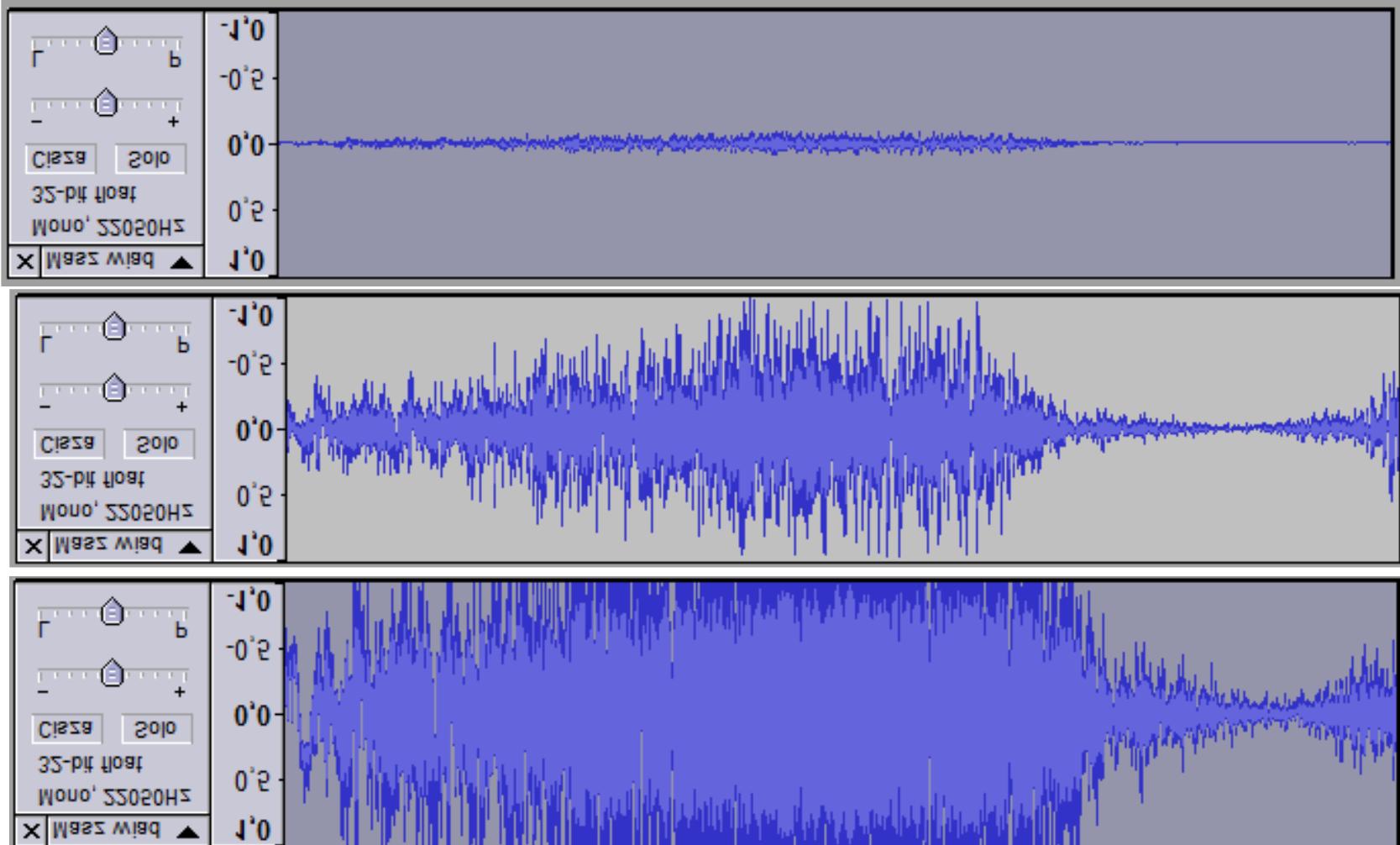
# Kwantowanie

- Wielkość zniekształceń powstających w procesie kwantowania zależy od stosunku bieżącego poziomu sygnału do maksymalnego zakresu przetwarzania.
- Jeśli bieżący poziom sygnału jest dużo mniejszy od maksymalnego zakresu przetwarzania, to mamy do czynienia z tzw. niedosterowaniem.
- Zjawisko można porównać do kwantowania tego samego sygnału z dużo mniejszą rozdzielczością bitową.
- W tym przypadku zbyt cichy sygnał sprawia, że szum kwantowania staje się wyraźnie słyszalny.
- Aby efekt niedosterowania był mniej wyraźny stosuje się tzw. Dithering – technikę polegającą na dodaniu do odtwarzanego sygnału użytecznego niskopoziomowego szumu losowego, który zagłusza zakłócenia a równocześnie jest neutralny dla ludzkiego słuchu.

# Kwantowanie

- Kiedy bieżący poziom sygnału jest większy od maksymalnego zakresu przetwarzania w sygnale wynikowym pojawia się przesterowanie (ang. Clipping).
- W rezultacie w sygnale wyjściowym pojawiają się dodatkowe składowe harmoniczne.
- Efekt przesterowania jest wyraźnie słyszalny, a samo zjawisko jest na ogół niepożądane.
- Niekiedy przesterowanie stosowane jest celowo, na przykład w elektronicznych układach modyfikujących dźwięki instrumentów muzycznych (efekt gitary fuzz, distortion czy overdrive).

# Kwantowanie



Fundusze  
Europejskie  
Wiedza Edukacja Rozwój



Politechnika Wrocławska

Unia Europejska  
Europejski Fundusz Społeczny



# Kwantowanie różnicowe

- Kwantowanie w metodzie PCM daje dobre rezultaty dla niskich i średnich rozdzielczości bitowych, natomiast przy wyższych wartościach (powyżej 16 bitów) wprowadza zniekształcenia słyszalne w sygnale wyjściowym w postaci charakterystycznych zakłocień i szumu modulacyjnego.
- Z tego powodu wprowadzono alternatywne metody kwantyzacji, do których należy **kwantyzacja różnicowa**.
- Polega ona na kodowaniu różnicy między dwoma sąsiednimi próbками sygnału wejściowego.
- Jest to możliwe, ponieważ składowe widma sygnału o małych częstotliwościach, które przenoszą dużą część energii sygnałów muzycznych, mogą posiadać dużą amplitudę, ale bardzo małą pochodną.
- Znaczy to, że sygnały posiadają wolno narastające i opadające zbocza i dlatego do ich konwersji (bez straty jakości sygnału fonicznego) wystarczy dużo mniejsza liczba bitów niż w tradycyjnej metodzie PCM.



# Kwantowanie różnicowe

W konwerterze A/C realizującym kwantowanie różnicowe wyjściowy sygnał różnicowy  $y(k)$  jest ciągiem próbek, z których każda ma wartość:

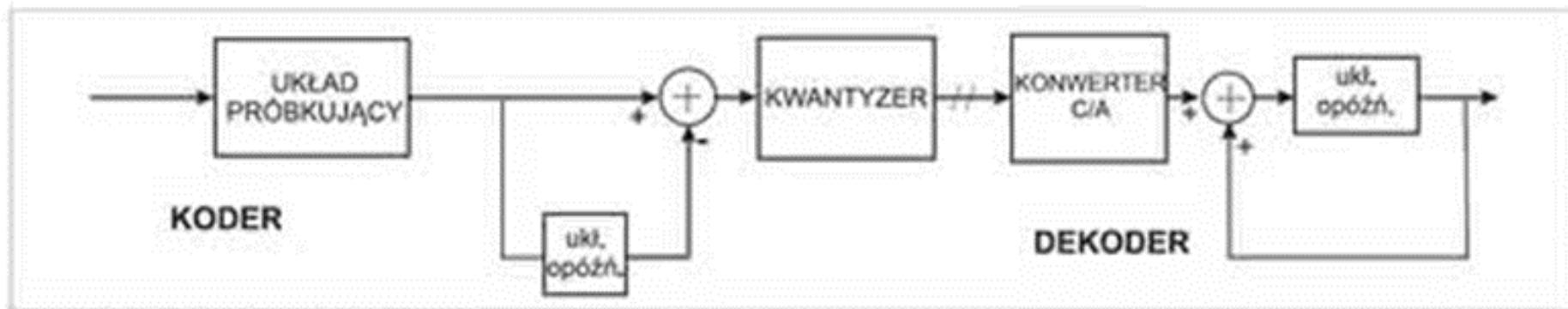
$$y_r = x_r - x_{r-1}$$

Gdzie  $x$  – jest sygnałem wejściowym, a  $r$  – oznacza numer próbki.

W dziedzinie czasu ten sam sygnał opisany jest wzorem:

$$y(t) = x(k \cdot T_s) - x(k \cdot T_s - T_s)$$

Gdzie  $T_s$  – jest okresem próbkowania, a  $k$  – liczbą całkowitą



Rysunek 3. System konwersji różnicowej.



# Kwantowanie różnicowe

- W kwantowaniu różnicowym sygnał błędu jest różnicą między sygnałem wejściowym a sygnałem wyjściowym (na wyjściu z dekodera) spełnia zależność:

$$\Delta e_r = x_r - x_r'$$

- W metodzie kwantyzacji różnicowej występuje znaczna poprawa stosunku sygnału do szumu kwantyzacji. Np. dla mowy poprawa wynosi od 3 do 11 dB
- Kwantyzacja różnicowa daje też widocznie lepsze rezultaty jeśli zwiększa się częstotliwość próbkowania.
- Im mniejszy jest okres próbkowania tym prawdopodobieństwo gwałtownej zmiany sygnału w tak krótkim czasie jest mniejsze.
- Przy bardzo wysokiej częstotliwości próbkowania maleje też sygnał błędu.



# Kwantowanie nierównomierne

- Kwantowanie może mieć również charakter nierównomierny (nieliniowy).
- Kwantowanie nierównomierne charakteryzuje się tym, że odległości między poszczególnymi poziomami kwantowania nie są jednakowe.
- Ten rodzaj kwantowania stosowany jest tam, gdzie rozkład prawdopodobieństwa wartości analogowych jest niejednostajny.
- Przy kwantowaniu nierównomiernych dąży się do tego, żeby wartościom występującym częściej niż innym przypisać większą liczbę poziomów kwantowania.
- Dzięki temu maleje błąd kwantyzacji (szum kwantyzacji).

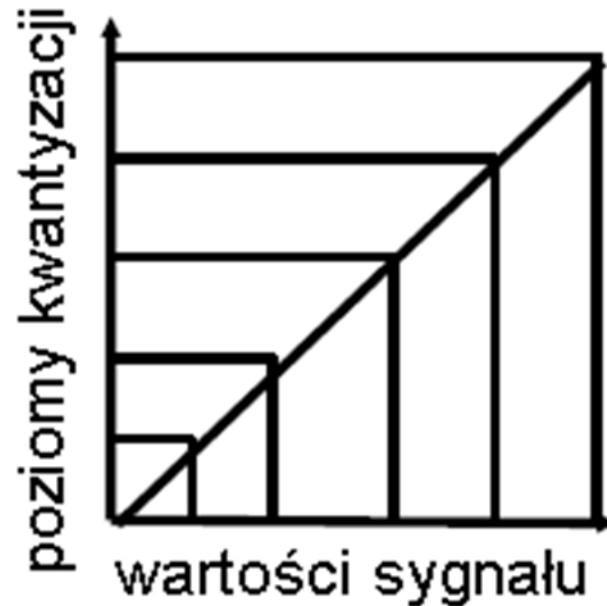
# Kwantowanie nierównomierne

- W odniesieniu do sygnału dźwiękowego kwantyzacja nierównomierna sprawdza się szczególnie dobrze w zakresie dźwięków cichych, na które ucho ludzkie jest szczególnie wyczulone i które narażone są na błędy wynikające z niedosterowania.
- W kodękach audio często stosuje się pseudologarytmiczną skalę kwantowania skorelowaną z charakterystyką ludzkiego słuchu, w której szerokości przedziałów kwantowania  $\Delta q$  są proporcjonalne do wartości kwantowanego sygnału  $x$  (tzw. zasada logarytmiczna).
- Zasady logarytmicznej nie można stosować dla  $x \sim 0$  gdyż wymagałoby to użycia nieskończonej liczby poziomów kwantyzacji.
- Z tego względu dla sygnałów o małej amplitudzie kwantyzator logarytmiczny przechodzi w równomierny. W takim przypadku stosuje się dwa algorytmy: A i  $\mu$ .
- Takie rozwiązanie powoduje spadek SNR dla cichych sygnałów.



# Kwantowanie nierównomierne

Porównanie kwantowania równomiernego i logarytmicznego



Fundusze  
Europejskie  
Wiedza Edukacja Rozwój



Politechnika Wrocławska

Unia Europejska  
Europejski Fundusz Społeczny



# Kodowanie

- Kodowanie jest ostatnim etapem konwersji analogowego sygnału dźwiękowego na sygnał cyfrowy.
- Niezależnie od tego czy kwantyzacja ma charakter liniowy czy nierównomierny, jako wartość próbki sygnału cyfrowego należy zapisać liczbę binarną reprezentującą poziom kwantowania jej przypisany.
- Przy realizacji większości operacji przetwarzania sygnałów cyfrowych zachodzi konieczność **obcinania lub zaokrąglania** reprezentacji binarnej sygnałów.
- Wynika to ze skończonej długości rejestrów używanych do przechowywania wartości ciągów próbek i współczynników filtrów.
- Przy zapisie liczb binarnych w kodzie z uzupełnieniem do dwóch, reprezentacje liczb stałopozycyjnych albo mantys liczb zmiennoprzecinkowych wygodnie jest rozważać jako  $(n + 1)$  bitowych ułamków, w których przecinek binarny leży bezpośrednio na prawo od najstarszego bitu.
- W takiej reprezentacji  $n$  oznacza liczbę bitów po przecinku (na prawo od przecinka binarnego).
- Wartość najmłodszego bitu wynosi  $2^{-n}$ ; określa ona równocześnie szerokość przedziału kwantyzacji, ponieważ liczby kwantowane są z krokiem  $2^{-n}$ .

# Kodowanie

- Do zapisywania liczb w rejestrze o skończonej długości najczęściej stosuje się zaokrąglanie.
- Polega ono na wyborze najbliższego poziomu kwantowania. W takim przypadku maksymalny błąd kwantowania wynosi:

$$\left| \frac{1}{2} 2^{-n} \right|$$

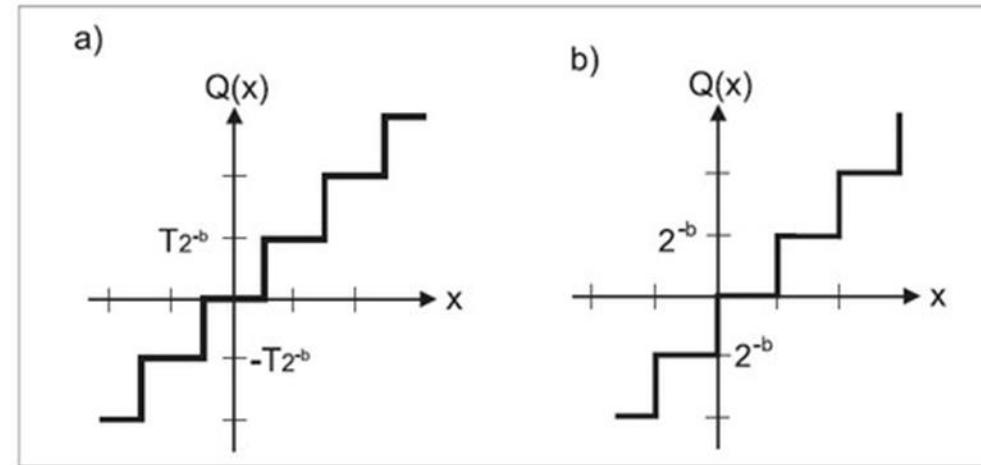
a błąd zaokrąglenia

$$\delta \in \left( -\frac{1}{2} 2^{-n}, \frac{1}{2} 2^{-n} \right)$$

- Błąd zaokrąglenia nie zależy od sposobu, w jaki zostanie przedstawiona liczba binarna, ponieważ samo zaokrąglenie dotyczy modułu liczby.

# Kodowanie

- Alternatywnym sposobem zapisu w rejestrze o stałej długości jest obcinanie najmłodszego bitu (który w rejestrze się nie mieści)
- Błąd spowodowany obcinaniem wynosi:  $\delta \in (-2^{-n}, 0)$
- Rysunek poniżej przedstawia efekty zaokrąglania (a) oraz obcinania (b);  $x$  – liczba przed zaokrągleniem (obcięciem),  $Q(x)$  – liczba zaokrąglona (obcięta)



Rysunek 2. Kwantyzacja przez: a) zaokrąglanie, b) obcinanie.



# Techniki przetwarzania mediów cyfrowych

Wykład 4: Kompresja danych dźwiękowych (8 h)

# Kompresja danych dźwiękowych

- Kompresja bezstratna:
  - DPCM
  - ADPCM
  - Metody logarytmiczne (A-Law,  $\mu$ -Law)
- Kompresja stratna:



# Kompresja danych dźwiękowych

- Najprostszym sposobem kodowania dźwięku jest metoda PCM.
- Dźwięk cyfrowy uzyskany tą metodą uważany jest za oryginalne, nieskompresowane dane cyfrowe (ang. raw data).
- Cyfrowe dane dźwiękowe zajmują bardzo dużo pamięci, np.
- 1 sekunda dźwięku monofonicznego (jeden kanał), zakodowanego z częstotliwością CD 44,1 kHz i rozdzielczością 16 bitów zajmuje:  
$$44100 \cdot 16 \text{ bitów} = 705600 \text{ bitów} = 88200 \text{ B} = 88,2 \text{ kB}$$
- A 1 minuta dźwięku stereo zakodowanego z tymi samymi parametrami (kanały stereo kodowane niezależnie), to:  
$$88,2 \text{ kB} \cdot 2 \cdot 60 \text{ s} = 10,584 \text{ MB}$$
- Stąd oczywiste stało się poszukiwanie skutecznych metod kompresji

# Kompresja danych dźwiękowych

- Kompresja to taki sposób zmiany zapisu informacji, że w efekcie zajmuje ona mniej bitów niż pierwotnie (plik wyjściowy ma mniejszą objętość).
- Metody kompresji dzieli się na:
  - Bezstratne – algorytmy kompresji i dekompresji są w pełni odwracalne a dane zdekompresowane są identyczne jak dane pierwotne
  - Stratne – algorytmy kompresji i dekompresji nie gwarantują pełnej rekonstrukcji danych wejściowych, tzn. dane wyjściowe różnią się od wejściowych
- Jako kryterium podziału algorytmów kompresji przyjmuje się również typ danych, którego one dotyczą:
  - Algorytmy ogólnego zastosowania
  - Algorytmy dedykowane dla konkretnego typu danych (np. obrazu binarnego, dźwięku)



# Kompresja danych dźwiękowych

- Metody kompresji bezstratnej są najczęściej algorytmami ogólnego zastosowania i dotyczą danych binarnych
- Ich działanie polega na wyszukiwaniu wzorców w ciągach danych binarnych. Każdy z wyszukanych wzorców zastępowany jest kodem jednoznacznie go identyfikującym, kompresując w ten sposób ciąg wejściowy.
- Dźwięk jest szczególnym rodzajem danych, ponieważ nie zawiera wielu długich, powtarzających się sekwencji. Powodem tego jest głównie szum.
- Dowolne dane analogowe zawierają zawsze pewną dozę błędu (szumu) losowego. Podczas cyfrowej rejestracji dźwięku analogowego błąd ten pojawia się na najmniej znaczących bitach.
- Ponieważ błąd ma charakter losowy, to nawet teoretycznie te same fragmenty dźwięku będą reprezentowane przez różne wzorce.

# Kompresja danych dźwiękowych

- Inne algorytmy ogólnego przeznaczenia, np. metoda Huffmana, dopasowują długości słów kodowych do częstości występowania poszczególnych wartości w ciągu danych binarnych.
- W przypadku danych dźwiękowych dają one niewielkie współczynniki kompresji w porównaniu z rozmiarami rzeczywistych plików (są zatem mało efektywne).
- Algorytmy dedykowane dla konkretnych typów danych uwzględniają ich specyfikę, np. model psychoakustyczny w kompresji dźwięku.
- Algorytmy te często mają charakter hybrydowy i jako ostatni krok stosują metody kompresji ogólnego zastosowania, np. algorytm Huffmana w kompresji mp3 czy JPEG
- Pierwsze fazy algorytmów dedykowanych można traktować jako wstępne przetwarzanie i kodowanie danych, dzięki czemu możliwa jest ich efektywna kompresja w ostatnim etapie.

# Różnicowa modulacja kodowo-impulsowa DPCM (Differential Puls Code Modulation)

- Różnicowa modulacja kodowo-impulsowa jest metodą kompresji bezstratnej stosowaną głównie do danych dźwiękowych.
- Kodowanie realizowane jest analogicznie jak w metodzie PCM z tym, że słowa kodowe w PCM reprezentują wartości próbek, natomiast słowa kodowe DPCM – różnice między dwoma kolejnymi próbками sygnału.
- Kodowanie różnicowe DPCM realizowane jest na dwa sposoby:
  - kodowanie (zwykle nieliniowe) różnicy między bieżącą a poprzednią wartością sygnału;
  - kodowanie różnicy między rzeczywistą wartością próbki a wartością estymowaną (przewidywaną) za pomocą pewnego modelu



# Różnicowa modulacja kodowo-impulsowa DPCM (Differential Puls Code Modulation)

Kodowanie różnicy między dwiema kolejnymi próbkami:

- Jeśli częstotliwość próbkowania dźwięku jest dostatecznie wysoka, to różnice między kolejnymi próbkami będą prawdopodobnie niewielkie.
- Jeżeli będziemy zapisywać (kodować) tylko różnice między wartościami próbek a nie wartości samych próbek, to możliwe jest przechowywanie tego samego sygnału dźwiękowego na mniejszej liczbie bitów przypadających na próbę. dwa następujące przykłady:
- Przykład 1
  - Przypuśćmy, że chcemy zredukować 8-bitowe dane PCM do ciągu 4-bitowych różnic. Z przyjętych założeń wynika, że różnice między kolejnymi próbkami mogą przyjmować wartości od -8 do 7. Niech kolejno zmierzone próbki mają wartości: 117, 120 i 124. Stosując DPCM zapamiętujemy jedynie liczby 3 i 4, które z powodzeniem można zapisać na 4 bitach. Proces dekompresji sprowadza się do odtworzenia kolejnych wartości próbek poprzez zsumowanie poprzedniej wartości próbki i różnicy odczytanej z pliku. Zysk z tytułu kompresji jest w tym przypadku ewidentny.

# Różnicowa modulacja kodowo-impulsowa DPCM (Differential Puls Code Modulation)

- Rozpatrzmy następujące przykłady:
- Przykład 1
- Założymy, że chcemy zredukować 8-bitowe dane PCM do ciągu 4-bitowych różnic.
- Z przyjętych założeń wynika, że różnice między kolejnymi próbками mogą przyjmować wartości od –8 do 7.
- Niech kolejno zmierzone próbki mają wartości: 117, 120 i 124.
- Stosując metodę DPCM zapamiętujemy jedynie liczby 3 ( $120-117=3$ ) i 4 ( $124-120=4$ ), które z można zapisać na 4 bitach.
- Proces dekompresji sprowadza się do odtworzenia kolejnych wartości próbek poprzez zsumowanie poprzedniej wartości próbki i różnicy odczytanej z pliku, tj. **117,  $117+3=120$ ,  $120+4=124$** .
- Zysk z tytułu kompresji jest w tym przypadku ewidentny.



# Różnicowa modulacja kodowo-impulsowa DPCM (Differential Puls Code Modulation)

## Przykład 2

- Założenia jak w poprzednim przykładzie.
- Niech teraz kolejne próbki przyjmują wartości: 17, 28 i 30.
- Różnice jakie należy w tym przypadku zapamiętać wynoszą odpowiednio: 11 i 2.
- Ponieważ największą wartością, jaką możemy zakodować w formacie 4 bitowym jest 7, a wartość pierwszej różnicy znacznie wykracza poza ten zakres, to możemy w tym przypadku zastosować dwa podejścia:
  - Liczbę 11 można zaokrąglić i zapamiętać ją jako 7, natomiast druga różnica zapisana jest zgodnie z wartością oryginalną, tj. 2. W tym przypadku, po dekompresji uzyskujemy próbki o wartościach: 17, 24 i 26. Jak widać błąd wkrada się do dwóch (i zapewne następnych) kolejnych wartości.
  - Przy drugim podejściu zaokrąglenie pierwszej wartości kompensujemy wartością drugiej różnicy i efekcie zapamiętujemy wartości: 7 i 6. Przy odtwarzaniu dźwięku wartości kolejnych próbek będą następujące: 17, 24 i 30.

# Różnicowa modulacja kodowo-impulsowa DPCM (Differential Puls Code Modulation)

- Wadą metody DPCM jest to, że liczba bitów przeznaczonych do zakodowania różnicy powinna być tak duża, aby można było na nich zapisać największą możliwą zmianę, w przeciwnym przypadku metoda prowadzi do przekłamań (jak to pokazano na przykładzie).
- Z drugiej zaś strony, w przypadku gwałtownych zmian poziomu sygnału słowo kodowe nadmiernie się rozrasta i zysk z tytułu kompresji jest niewielki.
- Zaprojektowanie dobrego algorytmu wykorzystującego metodę DPCM wiąże się ze spełnieniem dwóch wymagań:
- Z jednej strony, chcielibyśmy móc dokładnie zapamiętać każdą różnicę, zwłaszcza najmniejsze wartości.
- Z drugiej strony, duży zakres prawdopodobnych różnic oznacza też, że od czasu do czasu trzeba będzie (równie dokładnie) przechowywać wartości duże.



# Różnicowa modulacja kodowo-impulsowa DPCM (Differential Puls Code Modulation)

- Dlatego rzeczywiste techniki DPCM wykorzystują nieliniowy format kodowania.
- Sposób kodowania dla dwóch metod kompresji (8-bitowych próbek do 4-bitowych różnic) wykorzystywanych w komputerach klasy AMIGA:

Wartość kodu	-8	-7	-6	-5	-4	-3	-2	-1	0	1	2	3	4	5	6	7
Liczby	-34	-21	-13	-8	-5	-3	-2	-1	0	1	2	3	5	8	13	21
Fibonacciego																
Wartości eksponencjalne	-128	-64	-32	-16	-8	-4	-2	-1	0	1	2	4	8	16	32	64

# Różnicowa modulacja kodowo-impulsowa DPCM (Differential Puls Code Modulation)

- Obie metody kodowania dają znacznie większe możliwości w zakresie kodowania małych wartości różnic a tylko kilka dla wartości dużych
- Rzeczywisty proces dekompresji jest prosty:
- Dla każdej nowej wartości próbki przeglądamy tablicę kodową i odczytujemy z niej różnicę odpowiadającą zakodowanemu słowu a następnie dodajemy ją do wartości poprzedniej próbki.
- W praktyce metoda DPCM z kodowaniem rzeczywistych różnic jest rzadko stosowana, ponieważ aby utrzymać małe różnice trzeba zwiększyć częstotliwość próbkowania, co neguje korzyści wynikające z mniejszych rozmiarów próbek.

# Różnicowa modulacja kodowo-impulsowa DPCM (Differential Puls Code Modulation)

- Druga metoda, stosowana w kodowaniu DPCM znacznie częściej, wykorzystuje predykcję.
- Idea leżąca u podstaw „predykcyjnych” metod kompresji polega na znalezieniu sposobu przewidywania następnej wartości elementu danych na podstawie tych danych, które dotychczas się pojawiły.
- Jeśli możemy poprawnie przewidzieć następny element, to nie musimy zapamiętywać jego wartości. Idea ta zwana jest też często modelowaniem danych.
- Można zbudować prosty kompresor na podstawie dowolnej funkcji predykcji. Np. niech kompresor porównuje przewidywaną wartość z wartością rzeczywistą. Jeśli predykcja jest poprawna, kompresor przesyła jedynkę. W przeciwnym przypadku – wysyła zero i rzeczywistą wartość. Jeśli przez większość czasu predyktor działa poprawnie, uzyskujemy skompresowane dane na wyjściu.
- Ze względu na błędy losowe oraz inne czynniki, trudno jest dokładnie przewidywać kolejne wartości. Dlatego, większość metod wykorzystujących predykcję przesyła różnicę między wartością przewidywaną a rzeczywistą (tzw. błąd predykcji).
- Jeśli predykcja jest w miarę dokładna, to różnice te będą niewielkie. Każda z różnic może być więc zapisana na kilku bitach.

# Adaptacyjna różnicowa modulacja kodowo-impulsowa ADPCM (Adaptive Differential Puls Code Modulation)

- Rozszerzeniem metody kompresji DPCM jest adaptacyjna różnicowa modulacja kodowo-impulsowa (ADPCM).
- Modyfikacja wprowadzona w stosunku do metody DPCM polega na dopasowaniu długości słowa kodowego do aktualnej wartości różnicy.
- W tym celu stosuje się zmienny współczynnik tzw. rozmiar kroku.
- Rozmiar kroku jest tak dobierany, aby rzeczywistą wartość różnicy po podzieleniu przez rozmiar kroku można było dokładnie zapisać na 4 bitach (w algorytmie IMA ADPCM).
- Rozmiary kroku są różne dla różnych implementacji algorytmu, np. dla IMA rozmiar kroku wyraża się wzorem  $1,1^n$ ,  $n = 0, 1, 2 \dots$ , a jego wartości są stablicowane.
- Metoda ADPCM jest często używana do kodowania mowy i wykorzystywana w standardzie kompresji dźwięku G.726 w telefonii internetowej - VoIP.

# Logarytmiczna metoda PCM (log PCM)

- Standardowa metoda PCM wykorzystuje liniową skalę kwantowania, w której szum kwantowania jest stały i jednakowy zarówno dla dźwięków głośnych, jak i cichych.
- W konsekwencji dźwięki ciche (typowe dla dźwięków mowy) są „zagłuszane” przez szum.
- W logarytmicznej metodzie PCM zastosowano rozwiązańe, które polega na tym, by utrzymać stałą wielkość stosunku mocy sygnału użytecznego do mocy szumu, tak aby był on możliwie niezależny od sygnału wejściowego ( $\text{SNR} = \text{constans}$ , gdzie SNR – Signal to Noise Ratio)



# Logarytmiczna metoda PCM (log PCM)

- Utrzymanie stałej wartości współczynnika SNR w całym zakresie słyszalności jest możliwe poprzez zastosowanie logarytmicznej skali kwantowania.
- Dzięki temu możliwe jest „zagęszczenie” sygnału wejściowego przed przekazaniem do dalszego przetwarzania i „rozszerzenie” go po ponownym przyjęciu.
- Z tego względu logarytmiczna metoda PCM określana jest też często metodą dynamiczną.
- Zagęszczanie i rozszerzanie sygnału realizowane jest przez urządzenie zwane kompanderem (lub kompandorem)
- Kompander jest połączeniem kompresora i ekspandera (kompander = kompresor + ekspander).

# Logarytmiczna metoda PCM (log PCM)

- Zadaniem kompresora jest zmniejszenie zakresu dynamiki przed np. transmisją.
- Ekspander wykonuje proces odwrotny, czyli poszerza zakres dynamiki, tak aby przywrócone zostały pierwotne relacje pomiędzy składowymi sygnału.
- Na przykład, kompresor zmniejszy dynamikę z 60 do 40dB, to ekspander powinien ją zwiększyć z 40 do 60dB.
- Zasada działania kompandera wymaga aby wzmacnienie sygnału było zmienne w zależności od jego poziomu.
- W kompresorze silniej wzmacniane są sygnały cichsze.
- W ekspanderze odwrotnie, sygnały cichsze są wzmacniane słabiej.
- Dzięki temu słabiej wzmacniane są też szумy .
- Wzmocnienie układu jest więc płynnie zmieniane i niejako 'śledzi' sygnał.



# Logarytmiczna metoda PCM (log PCM)

- W pełni logarytmiczne kwantowanie jest w praktyce niemożliwe do osiągnięcia.
- Dlatego opracowano dwie metody umożliwiające osiągnięcie podobnych efektów.
- Zastosowano je odpowiednio w kompanderach a-Law oraz  $\mu$ -Law.
- Są to standardowe algorytmy kompresji ukierunkowane na komunikację telefoniczną włączone do standardu ITU G.711.
- Standard A - Law wykorzystywany jest głównie w Europie, zaś  $\mu$  - Law (zwany również u - Law) w Stanach Zjednoczonych i Japonii.
- Oba algorytmy umożliwiają osiągnięcie współczynników kompresji rzędu 2:1.
- Dzięki wprowadzeniu metod nieliniowych kod 8 bitowy może z powodzeniem zastępować kod 16 bitowy PCM.



# Logarytmiczna metoda PCM (log PCM)

- Podejście stosowane w metodach nieliniowych sprawdza się z wielu powodów:
- Po pierwsze, większość dźwięków składa się głównie z próbek o małych wartościach, a ten sposób kodowania zapewnia większą dokładność dla małych wartości, kosztem mniejszej dokładności dla rzadziej występujących dużych próbek.
- Po drugie, ludzkie słyszenie jest logarytmiczne. Zmiana głośności dźwięków cichych jest bardziej zauważalna niż w przypadku dźwięków głośnych.
- Zasadniczo, kodowanie logarytmiczne daje większą dokładność wszędzie tam, gdzie jest ona bardziej uchwytna dla ucha.



# Logarytmiczna metoda PCM (log PCM)

- Algorytmy  $\mu$  - Law i A - Law różnią się sposobem realizacji kwantyzacji logarytmicznej. Obie funkcje zostały zdefiniowane w postaci ciągłej i cyfrowej.
- W wersji ciągłej algorytm  $\mu$  - Law zakłada, że wartości wszystkich próbek należą do przedziału  $(-1, 1)$ .
- Dla danej próbki sygnału  $s$  odpowiadająca jej próbka  $\mu$  - Law przyjmuje wartość:  
$$s_\mu = sgn(s) \frac{\ln(1 + \mu|s|)}{\ln(1 + \mu)}$$
- Gdzie  $sgn(s)$  – znak próbki  $s$ ,  $\mu = 255$  jest wartością znormalizowaną dla systemów amerykańskich i japońskich
- Aby przekształcić  $s_\mu$  w pierwotne wartości  $s$  stosuje się zależność:

$$s = sgn(s_\mu) \frac{1}{\mu} ((1 + \mu)^{|s_\mu|} - 1)$$



# Logarytmiczna metoda PCM (log PCM)

- W algorytmie A – Law definiuje kolekcję kodów zależnych od dokładnych wartości numerycznych współczynników A. W Europie A=87,6.
- Jeśli mamy próbę  $s$  o wartości z przedziału  $<0, 1>$ , to wartość próbki A-Law wylicza się następująco:

$$s_A = sgn(s) \begin{cases} \frac{A|s|}{1 + \ln(A)}, & |s| < \frac{1}{A} \\ \frac{1 + \ln(A|s|)}{1 + \ln(A)}, & A \leq |s| \leq 1 \end{cases}$$



# Logarytmiczna metoda PCM (log PCM)

- Żeby wrócić do formatu liniowego z formatu A-Law stosuje się zależność:

$$s = \text{sgn}(s_A) \begin{cases} \frac{|s_A|(1 + \ln(A))}{A}, & |s_A| < \frac{1}{1 + \ln(A)} \\ \frac{e^{|s_A|(1+\ln(A))} - 1}{A}, & \frac{1}{1 + \ln(A)} \leq |s_A| < 1 \end{cases}$$

- Ze względu na to, że operacje potęgowania i logarytmowania są czasochłonne zarówno w algorytmie  $\mu$  - Law, jak i A – Law korzysta się ze szczególnego rodzaju aproksymacji, zdefiniowanej przez standard ITU G.711, która może być łatwo zrealizowana sprzętowo i algorytmicznie.
- Aproksymacja ta jest traktowana jako cyfrowa postać algorytmów  $\mu$  - Law i A – Law.

# Kodowanie i kompresja dźwięków mowy

- Mowa ludzka jest specyficznym rodzajem dźwięku.
- Charakteryzuje się licznymi pauzami między poszczególnymi słowami czy zdaniami. W tych miejscach w nagraniu występuje cisza.
- Próbki pobierane w okresach ciszy mają wartości zerowe lub zbliżone do zera (ze względu na szum).
- Najprostszy algorytm kompresji mowy – Kodowanie ciszy – polega na zamianie ciągów próbek „zerowych” pojedynczą wartością „0”, po której zapisywana jest liczba takich próbek.
- Tego rodzaju kodowanie może być stosowane wyłącznie do plików zawierających nagrania mowy i daje współczynnik kompresji rzędu 2:1.

# Kodowanie i kompresja dźwięków mowy

- Bardziej wyrafinowane, specyficzne metody kompresji mowy składają się z dwóch części:
  1. Matematyczny model ludzkiego traktu głosowego - ciąg równań wykorzystujących wiedzę z dziedziny biologii (krtań tchawica, język, itp.) do wytworzenia dźwięków wydawanych przez człowieka.
  2. Program analityczny (analysis engine), który pobiera dźwięk z wejścia i na jego podstawie ustala parametry modelu.
- Aby skompresować mowę wykorzystuje się analizator do wyodrębnienia parametrów modelu.
- Aby ją zdekompresować, podaje się do modelu traktu głosowego parametry, wyodrębnione w fazie kompresji, które pozwalają na zrekonstruowanie dźwięku (resynteza).
- Techniki te umożliwiają uzyskanie współczynnika kompresji rzędu 100:1, zapewniając dobrą jakość dźwięku.
- Dodatkowo, metoda ta poprzez lekką modyfikację modelu odtwarzania, pozwala modyfikować pewne charakterystyki głosu. W efekcie metody kompresji mowy często tworzą podstawy dla algorytmów modyfikacji głosu.



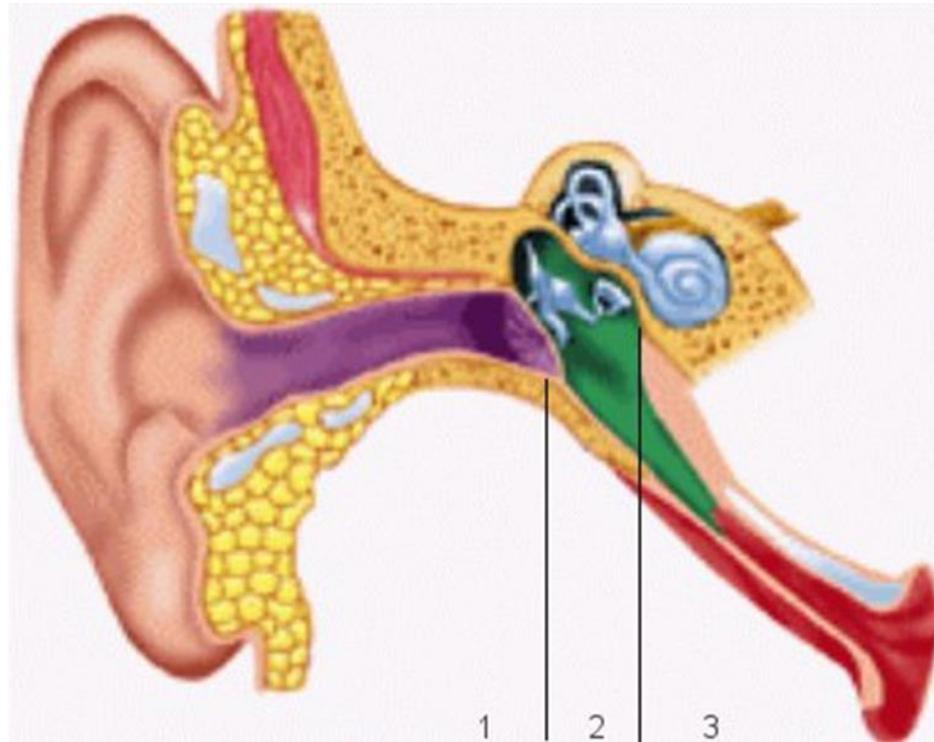
# Kodowanie perceptualne

- Bardziej zaawansowane, stratne algorytmy kompresji danych dźwiękowych wykorzystują niedoskonałości ludzkiego słuchu.
- Na ich potrzeby opracowano perceptualny model słuchu, który opisuje sposób odbierania dźwięków przez człowieka.
- Model uwzględnia anatomię i fizjologię słuchu, a w szczególności logarytmiczną charakterystykę słuchu, próg słyszenia, zjawisko maskowania dźwięków czy analizę pasm krytycznych.



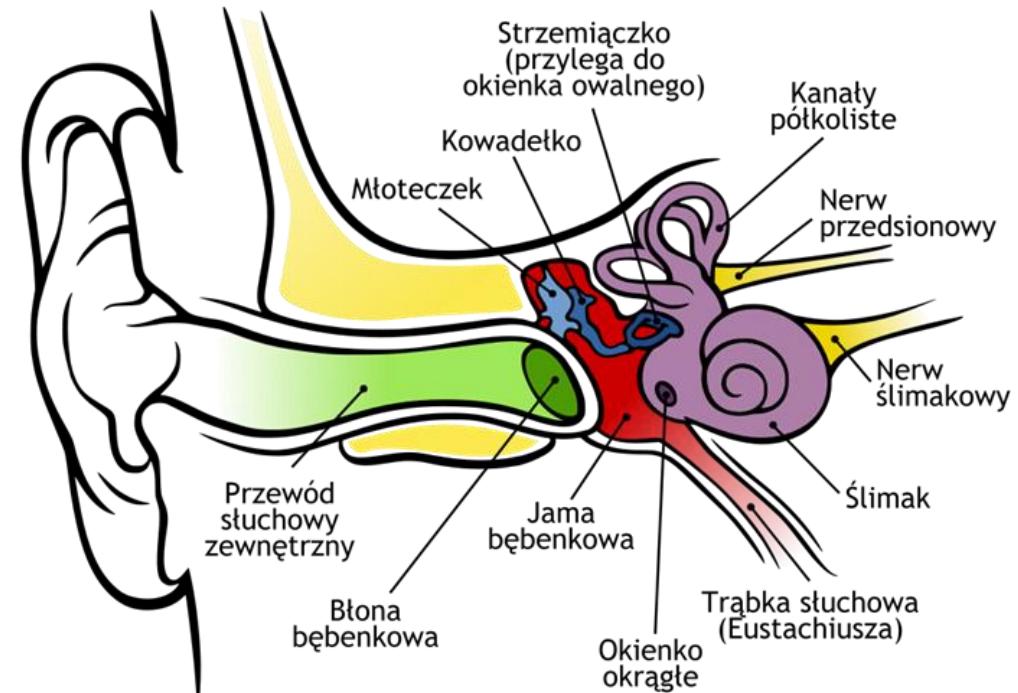
# Anatomia słuchu

- Ucho ludzkie składa się z ucha zewnętrznego (1), środkowego (2) i wewnętrznego (3).
- Integralnym elementem narządu słuchu jest jego część ośrodkowa (ośrodki słuchowe w mózgu).



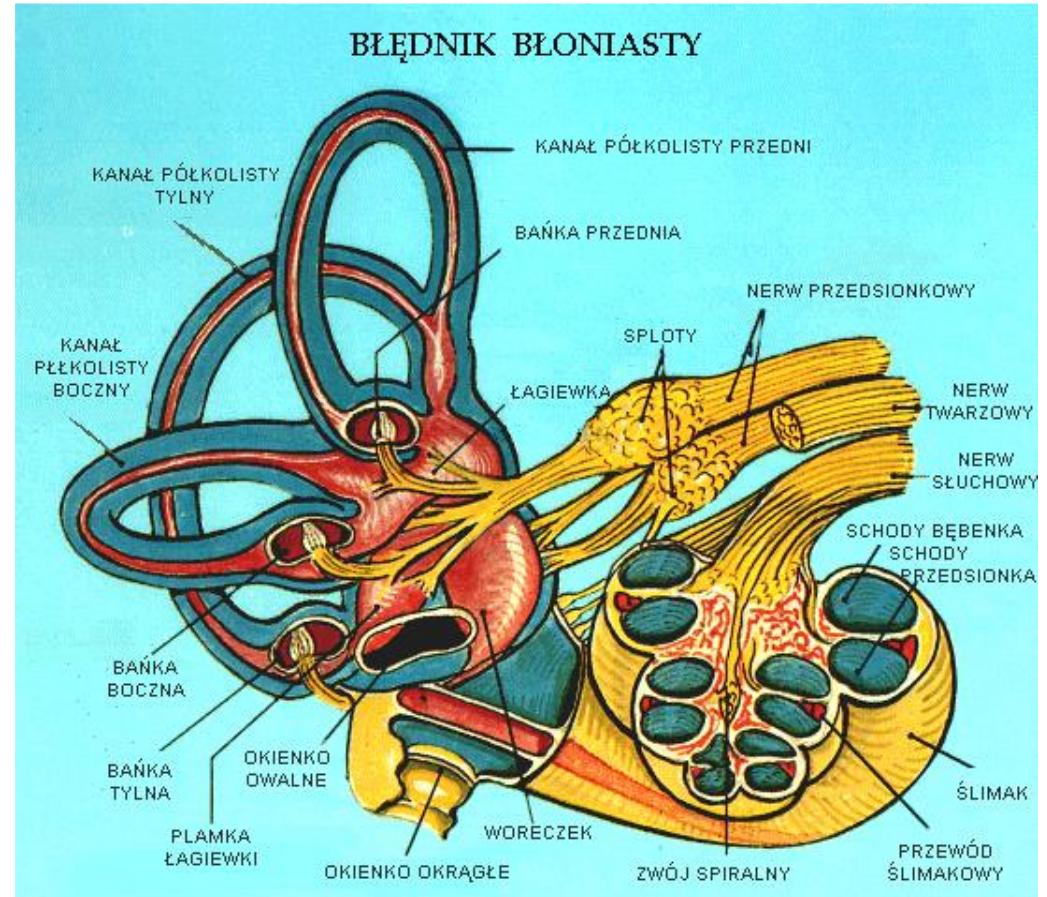
# Anatomia słuchu

- **Ucho zewnętrzne:** mażowina uszna, przewód słuchowy zewnętrzny.
- **Ucho środkowe:** błona bębenkowa, jama bębenkowa, kosteczki słuchowe: młoteczek, kowadełko i strzemiączko oraz trąbka słuchowa (Eustachiusza).
- **Ucho wewnętrzne (błędnik)** składa się z szeregu połączonych ze sobą przestrzeni, w których wyróżnia się błędniak błoniasty i błędniak kostny.
- Błędniak kostny zawiera przedsionek, ślimak i trzy kanały półkoliste. Jest on wypełniony płynem, zwanym przychlonką (perylimfą).



# Anatomia słuchu

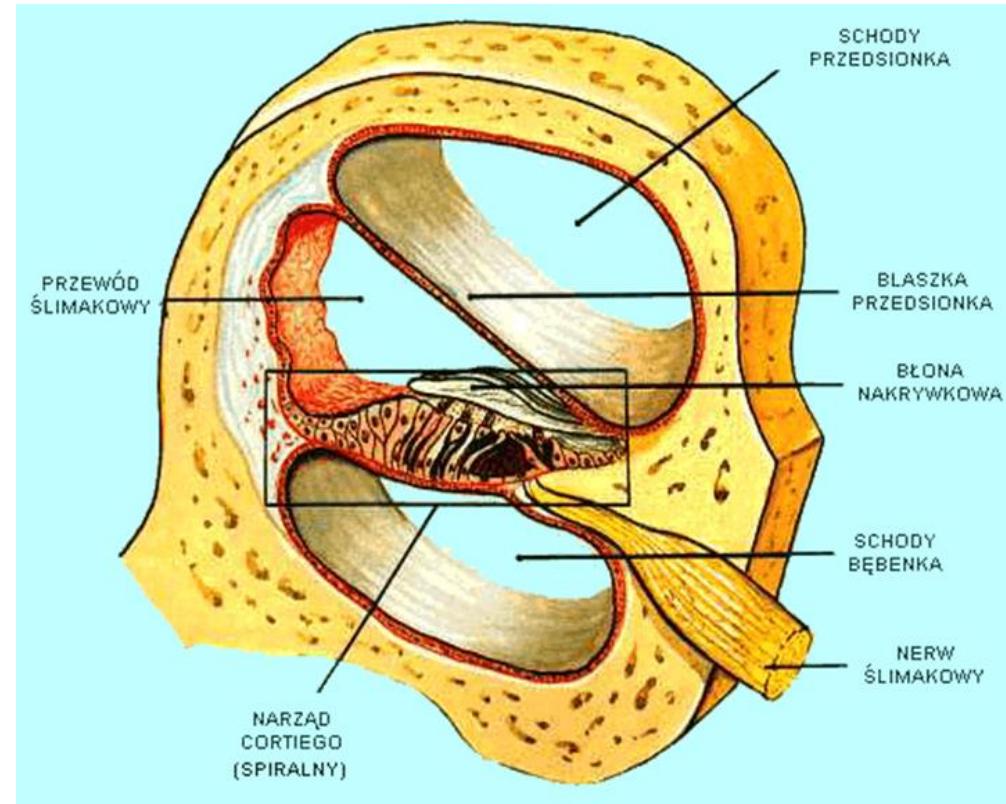
- Wewnątrz błednika kostnego mieści się błednik błoniasty zawieszony na licznych pasemkach łącznotkankowych, umocowanych na ścianach kostnych. Błednik błoniasty wypełniony jest śródchlönką (endolimfą).
- Błednik łączy się z jamą bębenkową dwoma otworami: okienkiem przedzionka (ówalnym), o które opiera się płytka strzemiączka i okienkiem ślimaka (okrągłym), zamkniętym błoną bębenkową wtórną.
- Ze względu na funkcje błednik dzieli się na:
  - część statyczną (narząd równowagi), do której należą przedzionek i kanały półkoliste
  - część słuchową, czyli ślimak.
- Zadaniem części statycznej jest sygnalizacja zmian położenia głowy i utrzymywanie równowagi.



# Anatomia słuchu

Za słuch odpowiadają:

- **okienko owalne** (przedzionka) - błona stykająca się bezpośrednio ze strzemiączkiem,
- **okienko okrągłe** (ślimaka) - błona nie stykająca się z zewnątrz z żadną z kostek; stanowi wylot stykający się ze schodami ślimaka (środkowa komora ślimaka).
- **ślimak** - długi, z wyglądu przypominający muszlę ślimaka, zwężający się kanał kostny, zwinięty spiralnie i wypełniony w całości płynem, w którym zawieszone są kryształki CaCO<sub>3</sub>; w środku przedzielony jest dwoma błonami - błoną podstawową i błoną przedzionkową (Reisnera), które dzielą ślimaka na trzy komory zwane schodami przedzionka, ślimaka i bębenka.
- **narząd Cortiego** – narząd spiralny znajdujący się wewnętrz przewodu ślimakowego (schodów ślimaka).



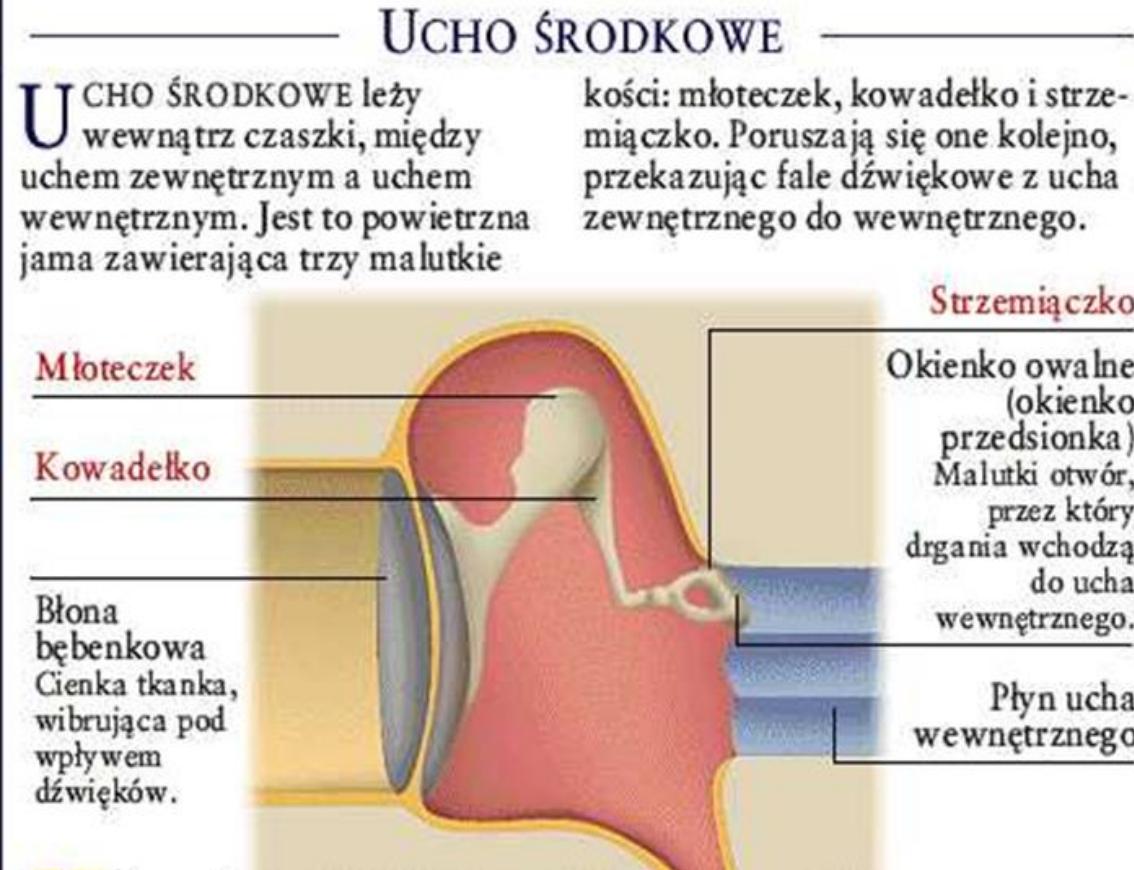
# Fizjologia słuchu

- W procesie percepcji dźwięku **ucho zewnętrzne** wychwytuje fale dźwiękowe, wzmacnia je i kieruje na błonę bębenkową
- Mażowina uszna swym kształtem jest przystosowana do zbierania fal dźwiękowych i doprowadzania ich do przewodu słuchowego. Wpływa też na polepszenie lokalizacji źródeł dźwięku, ponieważ stanowi zasłonę dla fal dźwiękowych przychodzących z tyłu i z boku. Dzięki temu dźwięki dobiegające z przodu są lepiej słyszalne.
- Przewód słuchowy zewnętrzny to kanał o długości ok. 26 – 30 mm i średnicy ok. 7 mm, który doprowadza fale dźwiękowe do błony bębenkowej (do jej powierzchni zewnętrznej).
- W przewodzie słuchowym występuje fala stojąca, która powstaje w wyniku odbicia się fali dźwiękowej prostopadle od powierzchni błony bębenkowej.
- Drgania słupa powietrza zawartego w przewodzie słuchowym powodują wystąpienie rezonansu dla częstotliwości ok. 3000 Hz i w konsekwencji wzmacnienie dźwięku dla tej częstotliwości o ok. 10 dB.
- Ponadto przewód słuchowy chroni błonę bębenkową przed uszkodzeniami mechanicznymi i zapewnia jej odpowiednią (stałą) temperaturę oraz wilgotność powietrza, co w znacznym stopniu wpływa na sprężystość błony bębenkowej.



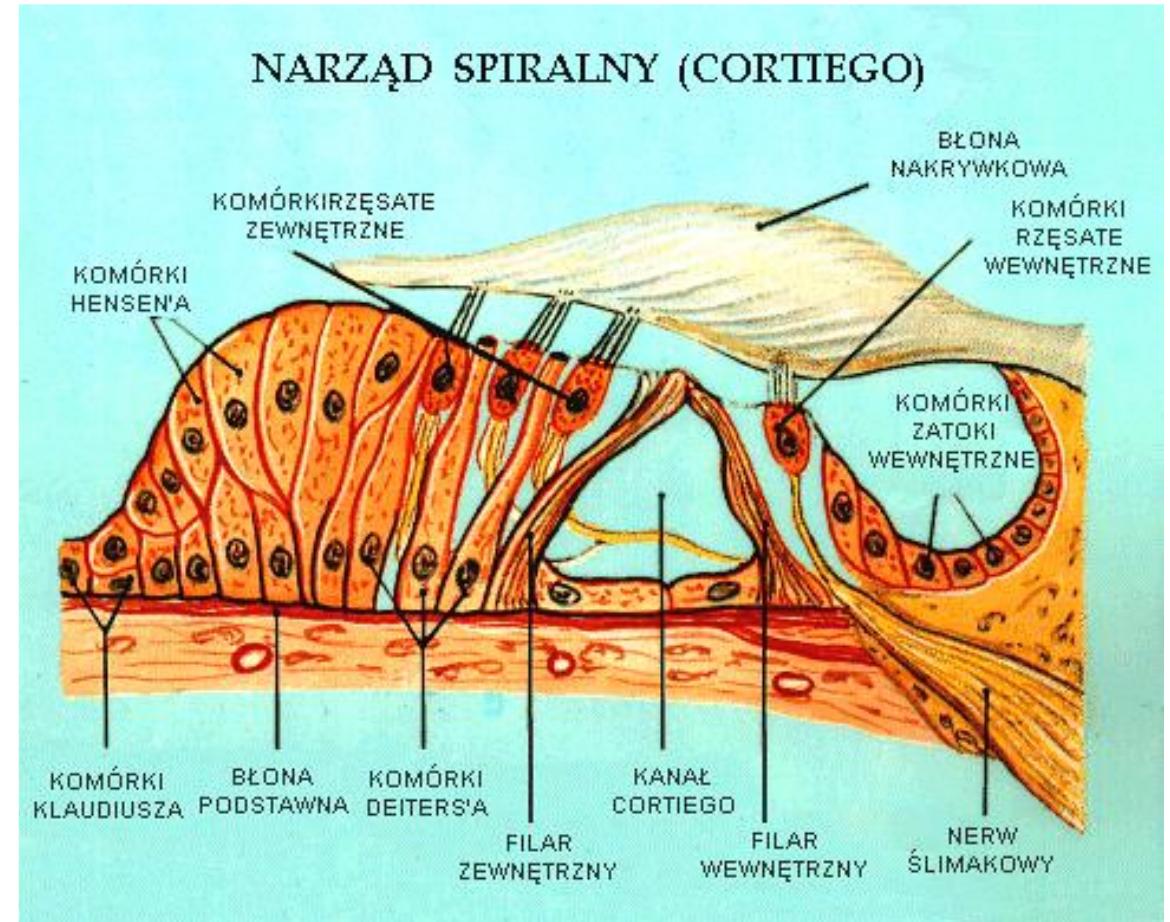
# Fizjologia słuchu

- Fala akustyczna, po przejściu przez przewód słuchowy zewnętrzny, uderza w błonę bębenkową i wprawia ją w drgania.
- W uchu środkowym, drgania błony bębenkowej wprawiają w ruch łańcuch kosteczek słuchowych: młoteczko, kowadełko i strzemiączko.
- W wyniku działania układu dźwigni kosteczek drgania błony bębenkowej zostają wzmacnione dwukrotnie przed przeniesieniem na podstawę strzemiączka.
- W procesie słyszenia istotną rolę odgrywają dwa mięśnie - napinacz błony bębenkowej, który rozluźniając się osłabia drgania zbyt mocnych dźwięków oraz mięsień strzemiączkowy pełniący analogiczną rolę.
- Strzemiączko, ma bezpośredni kontakt z błoną okienka owalnego, dzięki czemu przekazuje drgania do ucha wewnętrznego.



# Fizjologia słuchu

- Wzmocnione przez kosteczki słuchowe fale dźwiękowe przechodzą poprzez okienko ovalne do schodów przedśionka ślimaka.
- Organia mogą przechodzić również przez okienko okrągłe ślimaka bezpośrednio do schodów ślimaka. Wówczas jednak nie są one wzmocnione.
- Wewnątrz schodów ślimaka znajduje się narząd spiralny Cortiego, który zamienia pobudzenia znajdujących się na nim rzęsek w impulsy nerwowe.
- W skład narządu Cortiego wchodzą dwa rodzaje komórek: komórki zmysłowe włoskowate, czyli właściwe komórki słuchowe oraz komórki tworzące zrąb narządu i utrzymujące komórki włoskowate w ich położeniu (m.in. komórki filarowe, falangowe (Deitersa), graniczne (Hensena), podporowe (Claudiusa)).
- W narządzie Cortiego znajduje się też nabłonek zmysłowy narządu słuchu i w nim zaczynają się włókna części ślimakowej nerwu przedśionkowo-ślimakowego.
- Nerw ślimakowy przekazuje impulsy nerwowe powstałe pod wpływem fali akustycznej do ośrodków słuchu w mózgu wywołując w ten sposób wrażenia słuchowe.



# Teorie słyszenia

- Teorie słyszenia są przedmiotem dociekań fizjologów i akustyków.
- Zasadniczym ich celem jest wyjaśnienie w jaki sposób drgania bębenka, przeniesione za pośrednictwem układu kostek ucha środkowego do ucha wewnętrznego, powodują pobudzenie określonych zakończeń nerwowych w narządzie Cortiego.
- Kluczowym zagadnieniem jest sposób przeniesienia drgań strzemiączka na błonę podstawową, z którą jest związany narząd Cortiego.
- Dotychczas powstały cztery koncepcje wyjaśniające zjawisko słyszenia, które odpowiadają różnym ujęciom teoretycznym:
  - Teoria rezonansowa
  - Teoria fali stojącej
  - Teoria fali biegącej
  - Teoria rezonansu komórkowego



# Teorie słyszenia

## Teoria rezonansowa

- Zakłada się, że ze względu na małe rozmiary ślimaka można uważać, że zaburzenia wytworzone wskutek drgań strzemiączka rozchodzą się natychmiast w całej objętości ślimaka.
- Gdy strzemiączko drga z bardzo małą częstotliwością, wytworzone przezeń zmiany ciśnienia powodują przepływ perylimfy w obu kierunkach przez dziurkę osklepka ślimaka, co wymusza drgania błonki zamykającej okienko ślimaka.
- Gdy częstotliwość drgań jest znaczna, perylimfa nie zdąży przepływać przez dziurkę osklepka i aby wyrównać ciśnienia między schodami przedsionka i bębenka, przewód ślimakowy a wraz z nim błona podstawna musi choć w jednym miejscu drgać, przesuwając nadmiar cieczy.
- Ponieważ podatność błony podstawnej zmienia się nieznacznie wzduż przewodu ślimakowego, to przy większych częstotliwościach muszą drgać części błony podstawnej leżące bliżej przedsionka.
- Przeciwko teorii rezonansowej przemawiają obserwacje wskazujące na zdolności ucha do rozróżniania tonów o nieznacznie różniących się częstotliwościami i szybko następujących po sobie.



# Teorie słyszenia

## Teoria fali biegnącej

- Głoszona przez znakomitego uczonego, noblistę w tej dziedzinie, György Békésy.
- Stwierdził on, że pod wpływem drgań strzemiączka powstają w błonie podstawnej fale, rozchodzące się od przedsionka ku dziurce osklepka.
- Amplitudy tych fal w miarę oddalania się od przedsionka wolno wzrastają, osiągając w pewnym miejscu błony podstawowej wartość największą, a następnie szybko zanikają.
- Miejsce występowania największej amplitudy drgań zależy od częstotliwości i dla największych słyszalnych częstotliwości występuje w części błony najbliższej przedsionka, dla najmniejszych zaś - w części najbliższej dziurki osklepka.
- Dodatkowe zaostrzenie właściwości selektywnych jest związane z właściwościami układu nerwowego.



# Teorie słyszenia

## Teoria fali stojącej

- Wyjaśnia, że wyodrębnienie częstotliwości następuje w wyniku właściwego dla każdej częstotliwości rozkładu fal stojących wzdłuż błony podstawnej, a więc i przestrzennego rozkładu pobudzania tej błony.
- Wynikałoby stąd, że przeprowadzaniem analizy dźwięków obarczony wyłącznie mózg.
- Ze względu na to, że rozkład fal stojących, wywołanych dźwiękiem złożonym, zależy od wzajemnych faz tonów składowych, ucho powinno odczuwać różnice faz, czego doświadczalnie nie stwierdzono.

## Teoria rezonansu komórkowego

- Najnowsza i nadal rozwijana teoria, zgodnie z którą każdy organ ludzkiego ciała (a więc i narząd słuchu) jest zbudowany z komórek, z których każda posiada sobie właściwą częstotliwość rezonansową.
- Wzmocnienie tej częstotliwości przez bodźce zewnętrzne (falę akustyczną) powoduje określone reakcje ludzkiego ciała (wrażenie słyszenia).



# Perceptualny model słuchu

- Psychoakustyczny (perceptualny) model słuchu jest modelem matematycznym opisującym sposób postrzegania dźwięku przez człowieka.
- Model uwzględnia informacje o tym, które dźwięki są rozpoznawane przez ludzki narząd słuchu, a które nie są.
- Wytyczne do modelowania pochodzą z pomiarów psychoakustycznych (odsłuchowych), w których słuchacze oceniają wrażenia wywołane różnymi dźwiękowymi sygnałami testowymi prezentowanymi w określonym kontekście (np. czy słyszą ton sinusoidalny prezentowany na tle szumu).
- Model opisuje taki sposób przetwarzania sygnału dźwiękowego przez system słuchowy, aby wyjście systemu stanowiło predykcję subiektywnych ocen słuchaczy.



# Perceptualny model słuchu

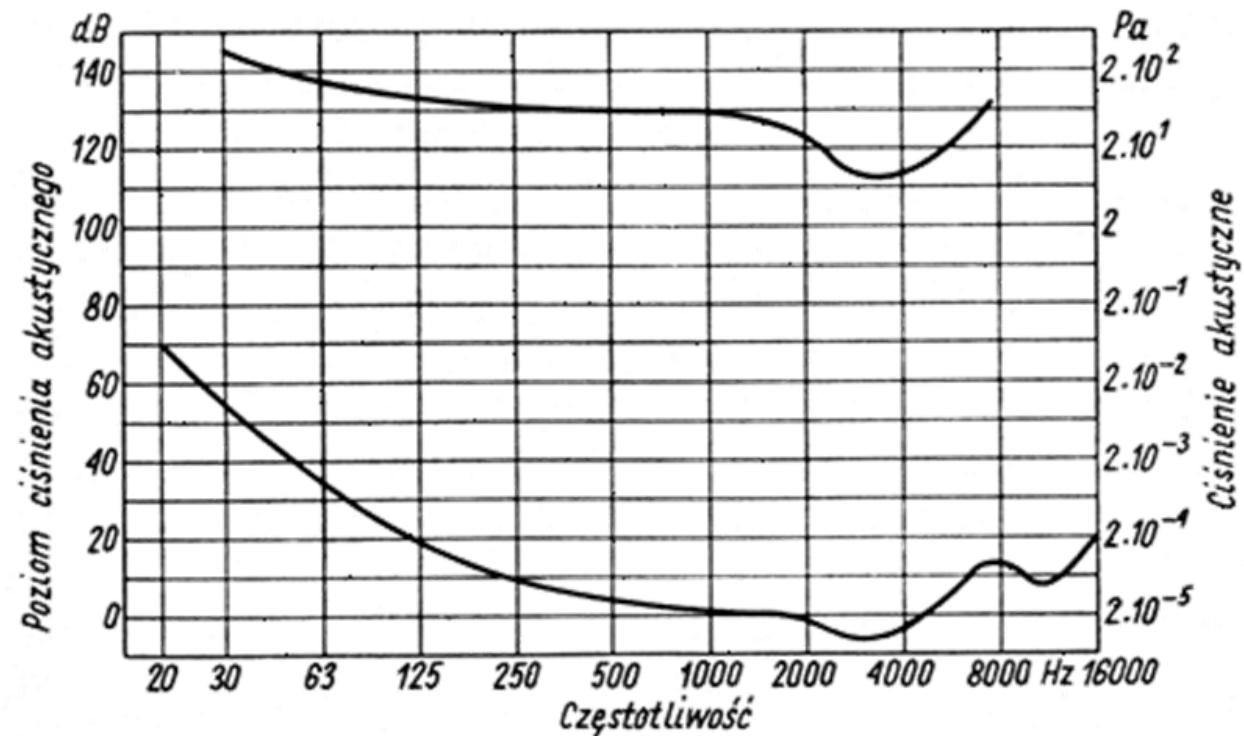
- Zjawiskami słuchowymi najczęściej wykorzystywanyymi przy opracowywaniu modeli psychoakustycznych są:
  - dolny i górny próg słyszalności,
  - percepceja wysokości dźwięków (logarytmiczna charakterystyka słuchu),
  - pasma krytyczne i maskowanie dźwięków,
  - dyskryminacja modulacji amplitudy.
- Modele psychoakustyczne przewidują zwykle zakres słyszalności od 20 Hz-20 kHz (dla tego właśnie większość współczesnych odtwarzaczy muzyki zapisanej cyfrowo ma takie pasmo przenoszenia) i maksymalną czułość w zakresie od 2 kHz do 4 kHz.



# Perceptualny model słuchu

## Progi słyszalności

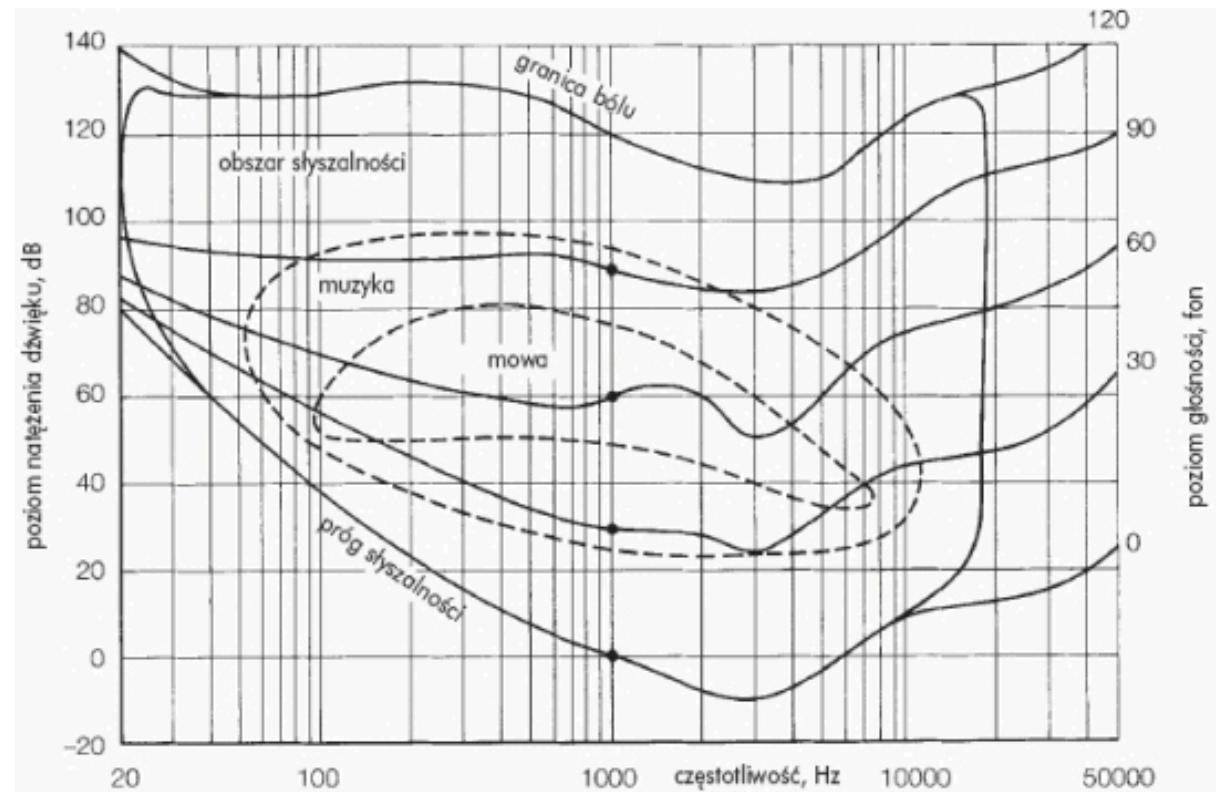
- Zakres częstotliwości słyszalnych przez człowieka jest ograniczony budową anatomiczną ucha.
- Najmniejsza częstotliwość słyszalna wynosi przeciętnie 16Hz, największa - 20kHz. Częstotliwość ta obniża się z wiekiem do 16 kHz, a nawet do 10 kHz.
- Podobne granice związane są z natężeniem dźwięku.
- Wartości ciśnień akustycznych dźwięków, przy których ucho zaczyna odbierać wrażenia dźwiękowe i ciśnień wywołujących wrażenie bólu zostały zmierzone przez wielu badaczy. Średnie wyniki badań przedstawia wykres na rysunku obok.



# Perceptualny model słuchu

## Progi słyszalności

- Krzywa, odpowiadająca najmniejszym akustycznym ciśnieniom słyszalnym, nazywa się dolną granicą (progiem) słyszalności.
- Krzywa, odpowiadająca ciśnieniom, przy których zaczyna występować uczucie bólu w uszach, nazywa się granicą słyszenia bolesnego, granicą bólu lub ogólnie - górną granicą słyszalności.
- Powierzchnia (obszar) słyszalności wyznaczona jest przez dolną i górną granicę słyszalności przecinające się w zakresie małych i dużych częstotliwości.
- Wewnątrz tego obszaru można wyróżnić mniejsze powierzchnie odpowiadające: dźwiękom mowy (najmniejszy obszar), dźwiękom muzycznym (nieco większy obszar zawierający dźwięki mowy) oraz pozostałą powierzchnię obejmującą wszystkie dźwięki słyszalne przez człowieka



# Perceptualny model słuchu

## Progi słyszalności

- Dolny próg słyszenia określony jest przez najmniejszą wartość poziomu ciśnienia akustycznego, która wywołuje ledwie postrzegalne wrażenie słuchowe.
- Próg ten pokazuje, że czułość słuchu jest największa (poziom ciśnienia akustycznego ma wartość najmniejszą) w zakresie częstotliwości średnich (1000-4000 Hz) i wyraźnie maleje w paśmie częstotliwości niskich i wysokich.
- Spadek czułości poniżej 1000 Hz wynosi ok. 6 dB/oktawę, natomiast powyżej 4000Hz – ok. 24 dB/oktawę.
- Wysoka czułość słuchu w zakresie średnich częstotliwości wynika z rezonansowego charakteru działania przewodu słuchowego oraz szczególnie efektywnego działania ucha środkowego w tym zakresie częstotliwości.
- W modelach psychoakustycznych dolny próg słyszenia przybliża się za pomocą równań matematycznych opisujących krzywe, których przebieg jest zbliżony do przebiegu krzywej empirycznej.
- Najpopularniejszy, wykorzystywany w wielu standardach kodowania, jest model Terharda:

$$LT_q = 3,64 * f^{-0,8} - 6,5 * \exp [-0,6 * (f - 3,3)^2] + 10^{-3} * f^4$$

$LT_q$  – poziom progu słyszenia w dB,

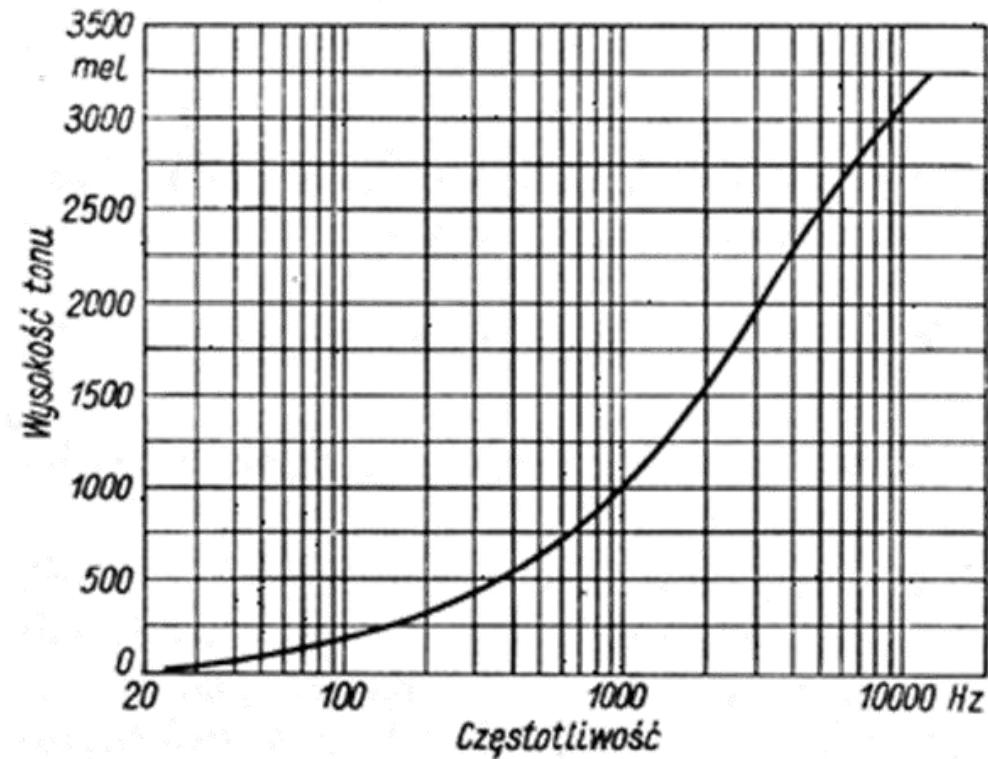
f – częstotliwość w kHz.



# Perceptualny model słuchu

## Percepcja wysokości dźwięków

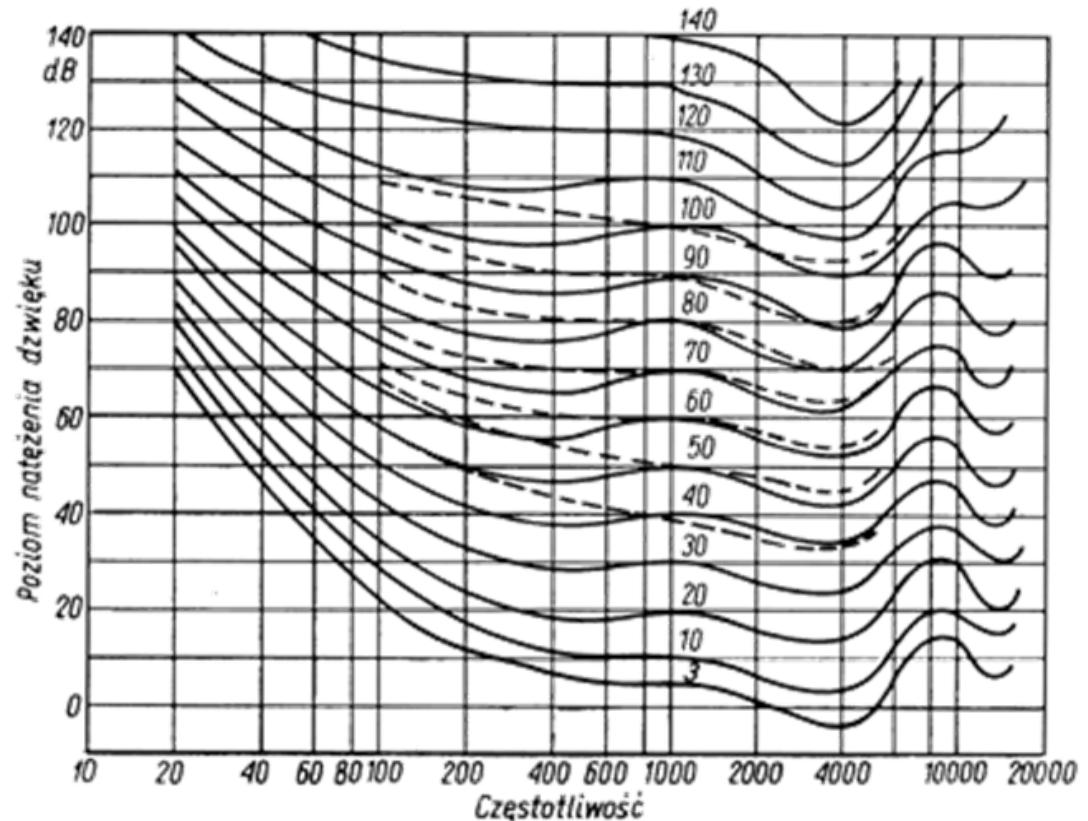
- Wysokość tonu jest wrażeniem słuchowym, umożliwiającym określenie położenia tonu na skali częstotliwości.
- Jednostką wysokości tonu jest mel.
- Wrażenie wysokości tonu zależy od natężenia dźwięku.
- Na podstawie pomiarów ustalono związek między subiektywną skalą wysokości tonu a obiektywną skalą częstotliwości.
- Zakłada się przy tym, że wysokość tonu o częstotliwości 1000Hz i poziomie ciśnienia akustycznego równym 40 dB wynosi 1000 melów.



# Perceptualny model słuchu

## Percepcja wysokości dźwięków

- Wrażeniem subiektywnym związanym z natężeniem dźwięku jest głośność dźwięku (wyrażona w dB), która zależy również od częstotliwości.
- Jednostką poziomu głośności jest fon
- 1 fon odpowiada poziomowi głośności dźwięku o częstotliwości równej 1 kHz i o poziomie ciśnienia akustycznego 1 dB (1 fon = 1 dB, przy 1 kHz).
- Na podstawie pomiarów wyznaczono krzywe jednakowej głośności
- Wartości otrzymane przez Robinsona i Dadsona przyjęte zostały za standardowe.
- Liczby nad krzywymi na wykresie obok wskazują poziom głośności w fonach



# Perceptualny model słuchu

## Percepcja wysokości dźwięków

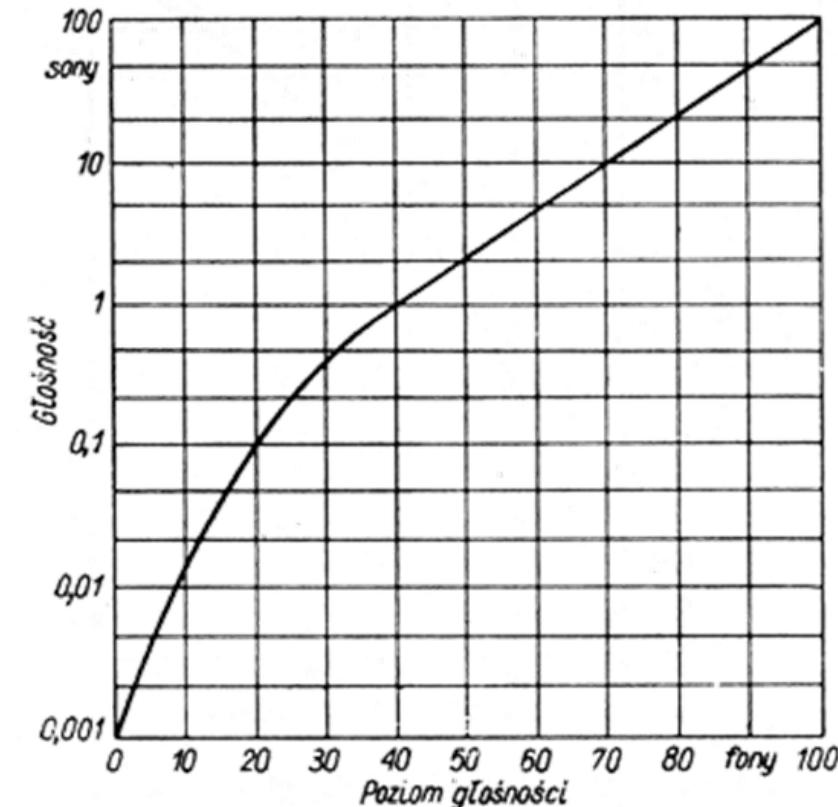
- Jednostką głośności jest son
- 1 son odpowiada głośności tonu o poziomie głośności 40 fonów.
- Zależność między głośnością tonów wyrażoną w sonach a poziomem głośności w fonach przedstawia wykres obok.
- Dla celów standaryzacyjnych przyjęto krzywą uzyskaną na podstawie zależności:

$$S = 2^{\frac{(L_s - 40)}{10}}$$

Gdzie:

S - głośność w sonach;

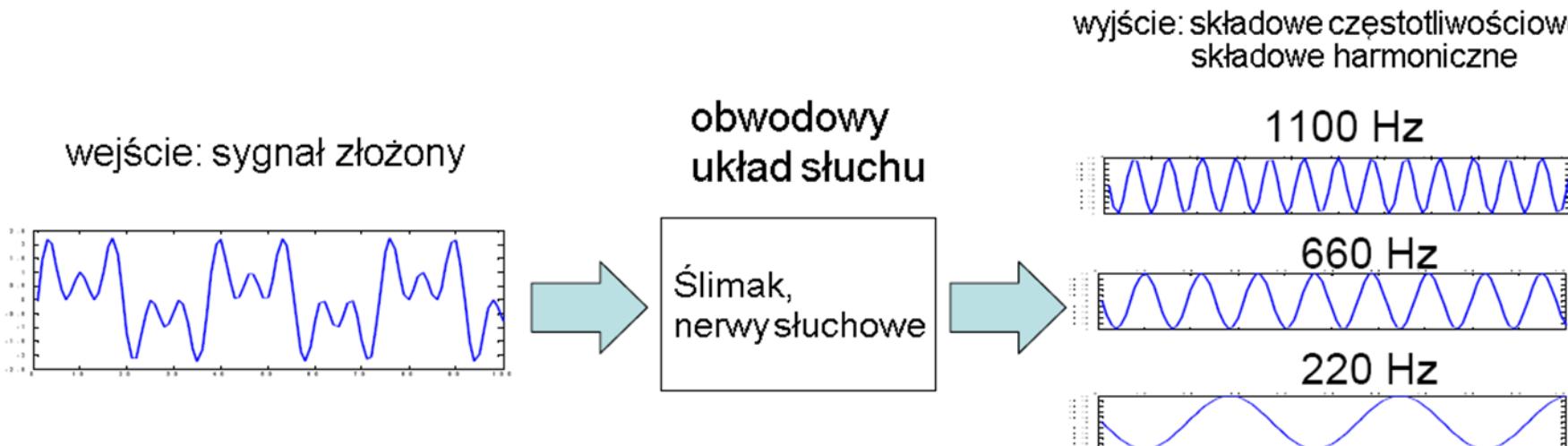
Ls - poziom głośności w fonach



# Perceptualny model słuchu

## Teoria pasm krytycznych

- Zgodnie z teorią słyszenia Békésy'ego (teoria fali biegnącej) system słuchowy człowieka przetwarza dźwięk w pewnych podpasmach, tzw. pasmach krytycznych.
- W każdym podpaśmie sygnał analizowany jest niezależnie od sygnałów, które występują w pozostałych podpasmach.



# Perceptualny model słuchu

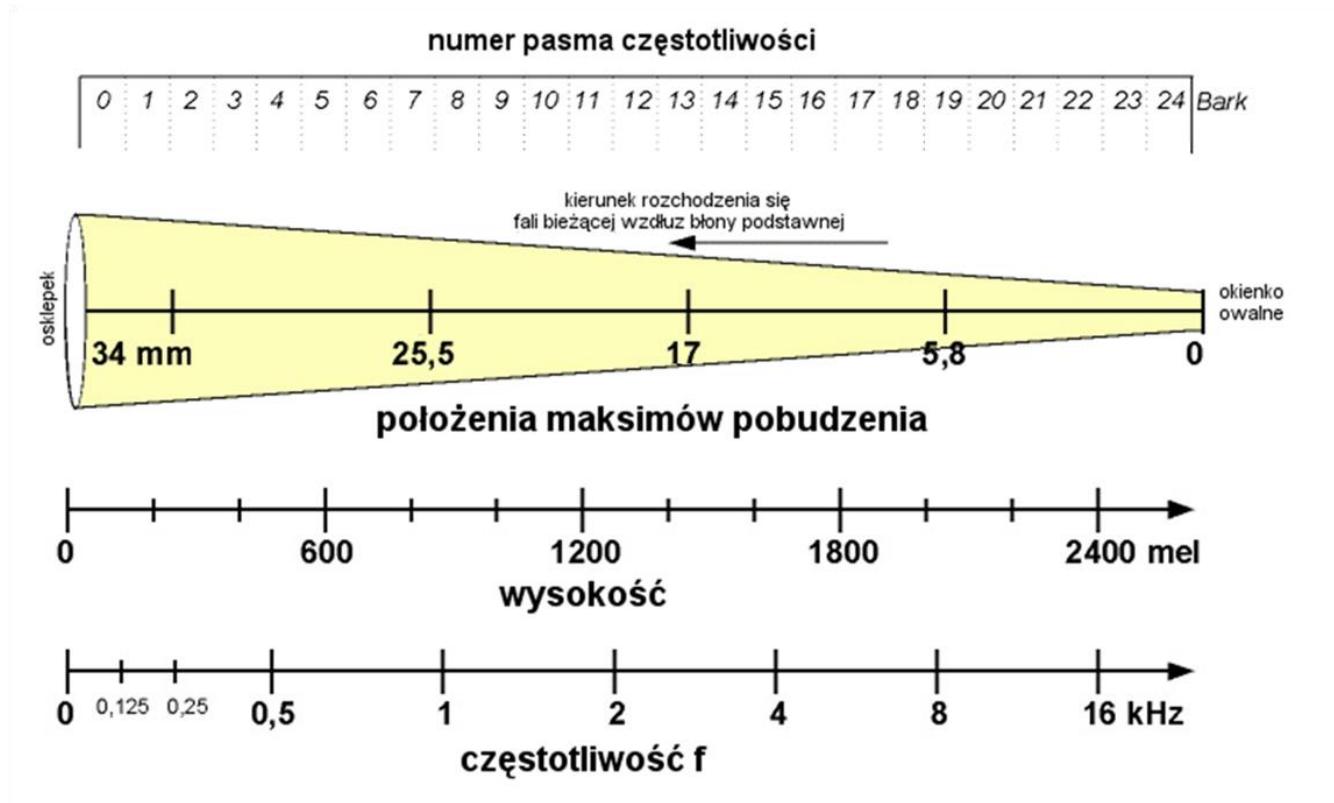
## Teoria pasm krytycznych

- Analiza sygnału dokonuje się w ślimaku.
- Fala rozchodząca się wzdłuż błony podstawnej pobudza określone jej miejsca do drgań.
- Percepcja częstotliwości sygnału odbywa się poprzez pasma krytyczne, określające rozdzielcość częstotliwościową narządu słuchu.
- Można wyznaczyć około 24 pasm krytycznych rozmieszczonych na błonie podstawnej ślimaka.
- Każdemu pasmowi krytycznemu odpowiada odcinek około 1,3 mm długości na błonie podstawnej (ok. 1300 neuronów).
- Istnieje zależność między miejscem położenia maksymalnej odpowiedzi na błonie podstawnej ślimaka a częstotliwością odbieranego dźwięku.
- Im wyższa częstotliwość, tym miejsce to znajduje się bliżej okienka owalnego, przy podstawie ślimaka.
- Zakres częstotliwości 20 - 5000 Hz rozkłada się na ponad 2/3 długości błony podstawnej (od 12 do 35 mm od okienka owalnego)
- Wyższy zakres częstotliwości (5 000-20 000 Hz) przypada na pozostałą część błony podstawnej, co stanowi mniej niż 1/3 jej długości.

# Perceptualny model słuchu

## Teoria pasm krytycznych

Wykres na rysunku obok ilustruje zależność między wysokością dźwięku a miejscem pobudzenia błony podstawnej ślimaka.



# Perceptualny model słuchu

## Teoria pasm krytycznych

- Definicja pasma krytycznego według *Fletchera*:

Pasmo krytyczne jest elementarnym pasmem częstotliwości o szerokości równej  $\Delta f$ , wydzielonym z ciągłego widma mocy szumów i zawierającym w sobie moc akustyczną równą mocy akustycznej tonu prostego o częstotliwości  $f$  położonej w środku tego pasma, przy czym rozpatrywany ton prosty ma taką intensywność, że zagłuszany przez nieograniczone widmo szumów ciągłych, znajduje się dokładnie na granicy słyszalności.

- Definicja pasma krytycznego według *Zwickera*:

Przy zwiększaniu szerokości pasma szumu, głośność pozostaje na tym samym poziomie, dopóki nie zostanie przekroczona szerokość pasma krytycznego. Wówczas wrażenie głośności ulega zmianie.



# Perceptualny model słuchu

## Teoria pasm krytycznych

- W kontekście teorii pasm krytycznych i licznych badań w tym zakresie okazało się, że system słuchowy człowieka może być modelowany jako zestaw filtrów pasmowo-przepustowych, dla których szerokość pasma jest równa szerokości odpowiedniego pasma krytycznego.
- Bezwzględne szerokości poszczególnych pasm krytycznych nie są jednakowe.
- Poniżej  $f = 500$  Hz szerokość pasma jest stała i wynosi oko 100 Hz.
- Powyżej  $f = 500$  Hz, szerokość każdego następnego pasma jest o 20% większa od szerokości poprzedniego pasma krytycznego.
- Szacowanie szerokości pasma krytycznego odbywa się na podstawie wzoru empirycznego:

$$\Delta f = 25 + 75(1 + 1,4 f^2)^{0,69}$$

gdzie  $f$  wyrażona jest w kHz, zaś szerokość pasma  $\Delta f$  w Hz.



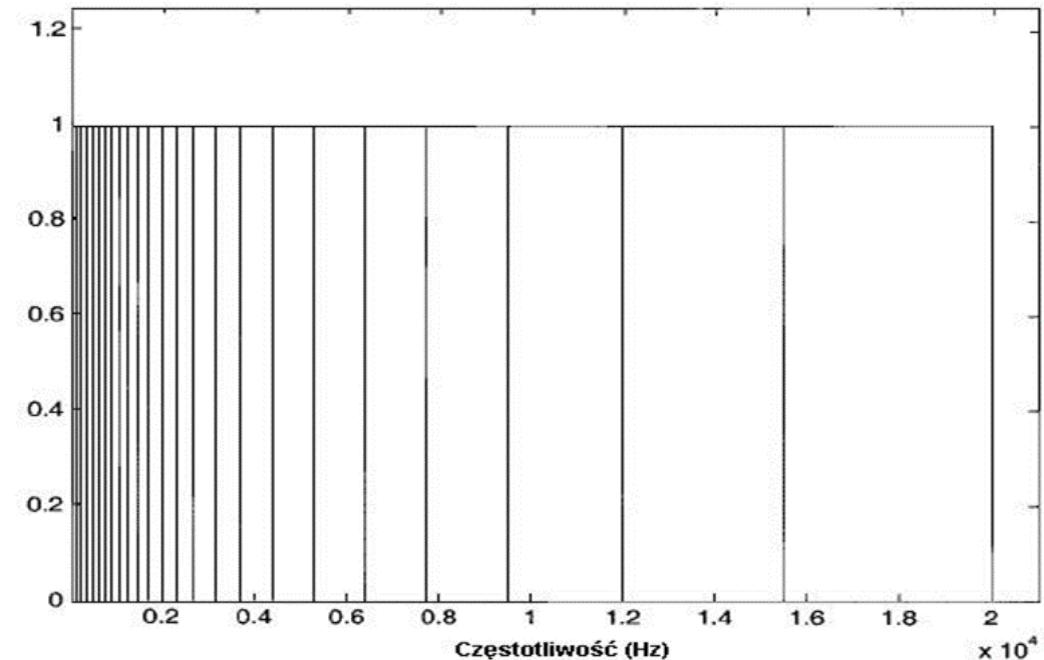
# Perceptualny model słuchu

## Teoria pasm krytycznych

- Ze względu na szerokie zastosowanie teorii pasm krytycznych zdefiniowano perceptualną jednostkę dźwięku – Bark, ku pamięci niemieckiego naukowca Barkhausena, który wprowadził jednostkę głośności fon.
- Jeden Bark odpowiada szerokości pojedynczego pasma krytycznego.
- Na podstawie obserwacji zaproponowano szereg zależności wyrażających związek między częstotliwością w Hz a numerem pasma krytycznego w barkach.
- Najczęściej stosowane jest odwzorowanie określone wzorem:

$$b = 13 \operatorname{arctg} (0,76f) + 3,5 \operatorname{arctg} \left( \left( \frac{f}{7,5} \right)^2 \right)$$

Gdzie:  $f$  częstotliwość wyrażona w kHz



# Perceptualny model słuchu

## Maskowanie

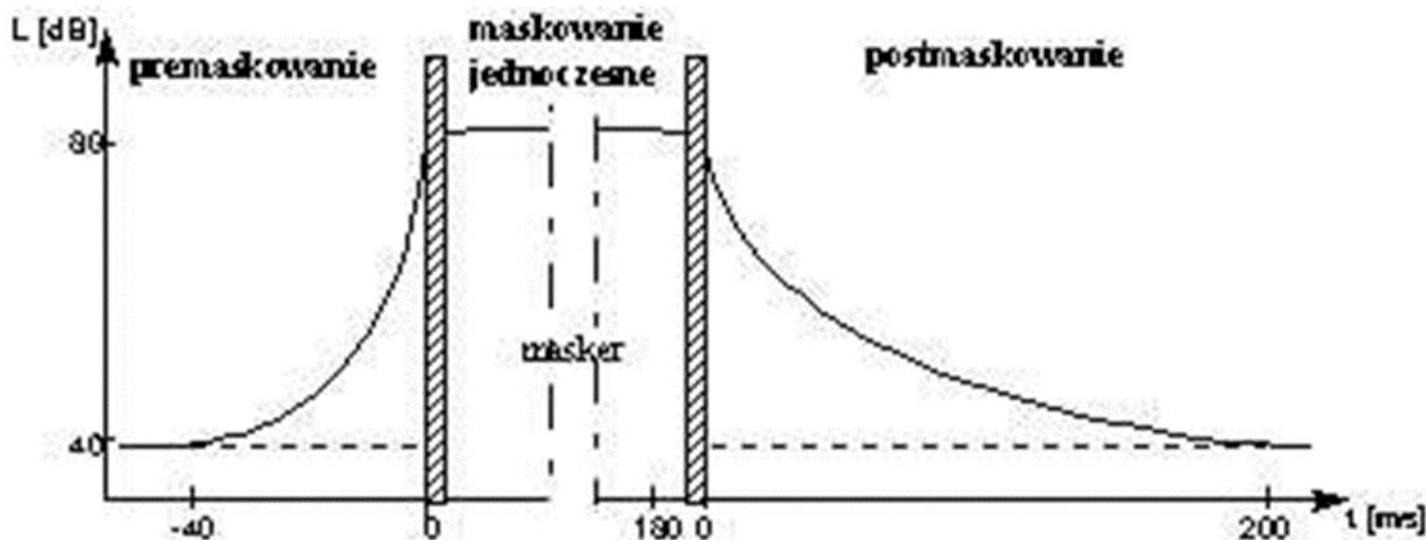
- W algorytmach kompresji stratnej wykorzystuje się zjawiska maskowania (zagłuszania) dźwięku.
- Interpretacja anatomiczna tych zjawisk wskazuje, że są one spowodowane wychylaniem się błony podstawnej ślimaka pod wpływem bodźców akustycznych.
- Rzęski komórek słuchowych są drażnione przez pobudzaną na drgań błonę podstawną.
- Pobudzone silnym bodźcem słuchowym zakończenia nerwowe nie mogą odbierać innych bodźców o częstotliwościach i poziomach głośności powodujących mniejsze wychylenie błony podstawnej byłoby niż aktualnie występujące.
- W praktyce zjawisko maskowania powoduje podwyższenie progu detekcji sygnału maskowanego w wyniku obecności innego sygnału (maskera). Masker utrudnia (lub uniemożliwia) odbiór sygnału maskowanego.



# Perceptualny model słuchu

## Maskowanie

- Wyróżnia się dwa rodzaje maskowania:
  - nierównoczesne (modelowane w dziedzinie czasu),
  - równoczesne (maskowanie w dziedzinie częstotliwości).
- Maskowanie nierównoczesne może występować jako:
  - maskowanie wstępne (premaskowanie, maskowanie wsteczne),
  - postmaskowanie (maskowanie pobodźcowe).



# Perceptualny model słuchu

## Maskowanie

- Maskowanie wsteczne (premaskowanie) występuje wówczas, gdy pojawiający się dźwięk o dużym natężeniu jest w stanie zagłuszyć tony, które wystąpiły do kilkunastu milisekund wcześniej.
- Występowanie maskowania wstępnego wyjaśnia się tym, że dźwięki głośniejsze są przetwarzane przez system słuchowy szybciej niż ciche.
- Maskowanie pobodzowe (postmaskowanie) pojawia się wtedy, gdy słabszy dźwięk następuje w odstępie 0 . . 200 ms po dźwięku silniejszym.
- Czas utrzymywania się maskowania zależy w tym przypadku od natężenia tonu maskującego i czasu jego trwania.
- Zjawisko to spowodowane jest stosunkowo długim czasem relaksacji neuronów.

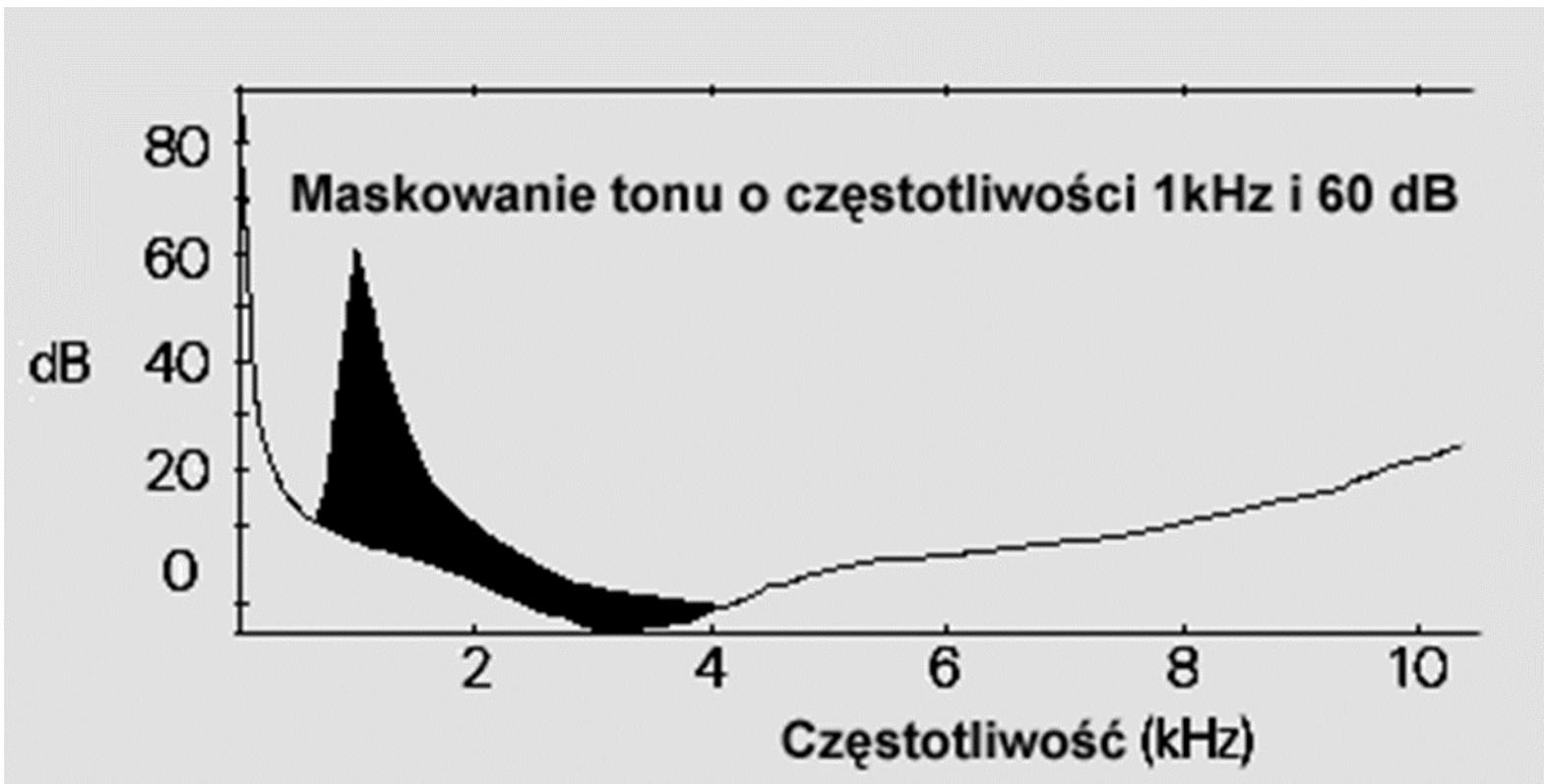
# Perceptualny model słuchu

## Maskowanie

- Maskowanie równoczesne charakteryzuje się tym, że pewne tony stają się niesłyszalne w obecności innych tonów, maskerów.
- Efekt maskowania równoczesnego zależy od wzajemnego natężenia i częstotliwości tonu maskującego i maskowanego.
- Zależność tę opisują tzw. krzywe maskowania dla maskerów o danym natężeniu i częstotliwości.
- Dźwięki, których wartości natężenia leżą poniżej krzywej maskowania stają się niesłyszalne.
- W efekcie dolny próg słyszenia podwyższa się do poziomu wyznaczonego przez dolną granicę słyszalności i próg maskowania.
- Ogólnie:
  - Tony o niższej częstotliwości maskują tony o wyższych częstotliwościach.
  - Im ton maskujący jest głośniejszy, tym efekt wyraźniejszy.

# Perceptualny model słuchu

## Maskowanie



Fundusze  
Europejskie  
Wiedza Edukacja Rozwój



Politechnika Wrocławska



Unia Europejska  
Europejski Fundusz Społeczny

# Perceptualny model słuchu

## Maskowanie

Dźwięki o niskiej częstotliwości maskują skutecznie te o wyższej częstotliwości

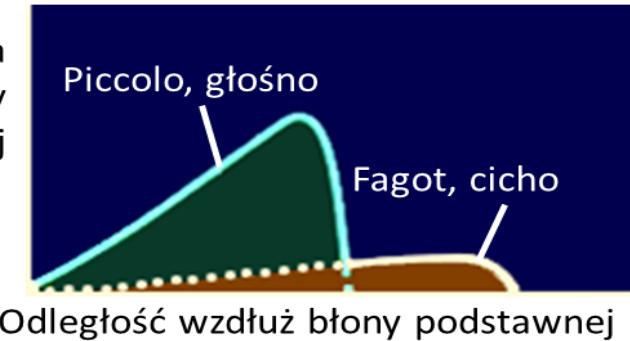
Jednak w przypadku „relatywnie głośnych” dźwięków o wysokiej częstotliwości zachodzi sytuacja odwrotna

Amplituda  
drgań błony  
podstawnej



Maskowanie fletu piccolo przez fagot

Amplituda  
drgań błony  
podstawnej



Maskowanie fagotu przez flet piccolo

# Perceptualny model słuchu

## Próg maskowania

- Próg maskowania jest ściśle związaną z pobudzeniem (wychyleniem z położenia równowagi) błony podstawnej ślimaka.
- Wyznacza minimalny poziom energii, poniżej której sygnał dźwiękowy jest niesłyszalny.
- Jedną z metod wyznaczania progu maskowania jest wykorzystanie tzw. funkcji wagowej, zależnej od numeru pasma krytycznego.
- Próg maskowania określony jest wówczas zależnością:

$$LT(i) = w(i) * LE(i),$$

gdzie:

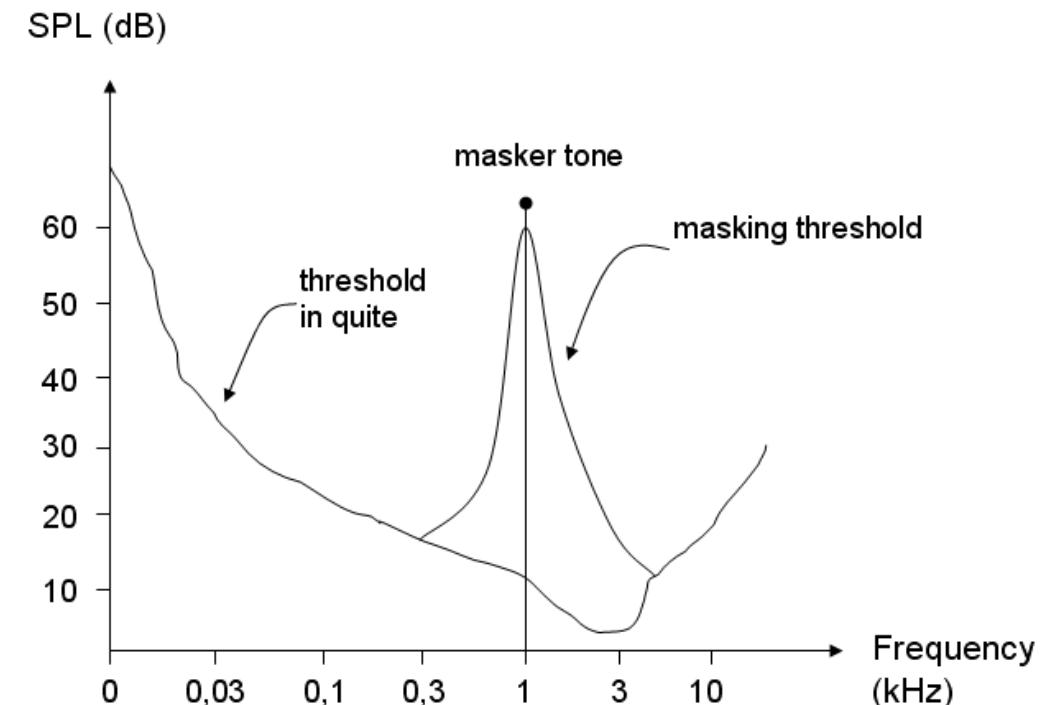
$$10 \log_{10} (w_i) \cong - (15,5 + i),$$

$i$  – numer pasma krytycznego,

$w_i$  – waga  $i$ -tego pasma.

$LT(i)$  – próg maskowania dla  $i$ -tego pasma krytycznego w dB,

$LE(i)$  – pobudzenie w  $i$ -tym paśmie krytycznym.



# Standard MPEG

## Historia

- Komitet standaryzacyjny MPEG (ang. Moving Picture Export Group) powstał w roku 1988 w celu opracowanie standardu kodowania sygnału audiowizualnego o niskiej przepływności binarnej.
- I Etap prac – standard MPEG-1 - listopad 1992 r.:
  - Kodek foniczny (ISO/IEC 11172-3) został zaprojektowany dla sygnałów jedno- i dwukanałowych przy częstotliwości próbkowania 32; 44,1 oraz 48 kHz.
  - Kodek zapewnia wysoką jakość dźwięku przy przepływności binarnej równej 128 kbit/s.
- Cel dalszych prac grupy MPEG - opracowanie wielokanałowego rozszerzenia standardu fonicznego MPEG-1 oraz zdefiniowanie nowej metody kodowania sygnałów o niższych niż w standardzie MPEG-1 częstotliwościach próbkowania, tj. 16; 22,05 i 24 kHz.
- II Etap – standard MPEG-2 BC i MPEG-2 LSF (ISO/IEC 13818-3) – listopad 1994 r.:
  - MPEG-2 BC (Backward Compatible) zapewnia dobrą jakość dźwięku w zakresie przepływności binarnej 640 – 896 kbps dla pięciu kanałów,
  - MPEG-2 LSF (Low Sampling Frequency) – metoda kodowania sygnałów o niskich częstotliwościach



# Standard MPEG

## Historia

- III Etap – kwiecień 1997 r.:
  - MPEG-2 NBC (Non Backward Compatible) wersja standardu niekompatybilna z wcześniejszymi, przemianowana następnie na MPEG-2 AAC (Advanced Audio Coding) i zdefiniowana normą ISO/IEC 13818-7
  - MPEG-2 AAC definiuje sposób kompresji wielokanałowego sygnału fonicznego o jakości określonej przez Międzynarodową Unię Telekomunikacyjną (ITU-R) dla sygnałów pięciokanałowych i przepływności binarnej 384 kbps lub mniejszej.
- Standard MPEG -3 przewidywał włączenie kompresji obrazu przeznaczonej dla HDTV (High Definition Television), jednak ostatecznie stwierdzono, iż korzystniejsze będzie jego dołączenie do standardu MPEG-2



# Standard MPEG

## Historia

- Standard MPEG-4 wprowadzono pod koniec roku 1998:
  - Zbiór kodeków kompresji dla kodowania percepcyjnego sygnałów wraz z pewnymi wariantami kodowania Advanced Audio Coding (AAC) oraz innymi narzędziami kodowania audio/mowy.
  - W ramach standardu opracowano technologię HE-AAC (ang. High Efficiency-AAC) będącą kombinacją trzech technologii kodowania: Advanced Audio Coding (AAC), Spectral Band Replication (SBR) oraz Parametric Stereo (PS).
- Technologia AAC+ jest stosowana między innymi w 3G i WiFi audio oraz audio-video usługach, w strumieniowaniu i pobieraniu danych audio w Internecie, w radiofonii cyfrowej, w cyfrowej telewizji satelitarnej i kablowej oraz w przenośnych odtwarzaczach audio.



# Standard MPEG

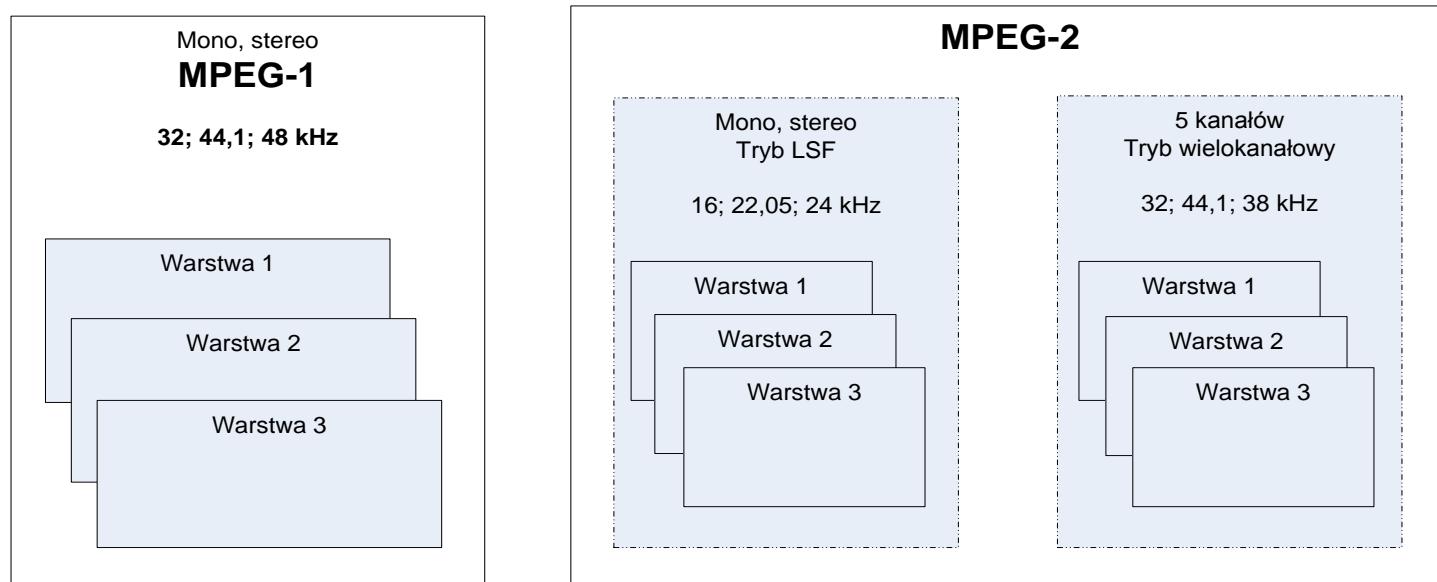
## Historia

- MPEG-7 jest standardem opisu treści multimedialnych i w odróżnieniu do MPEG-1, MPEG-2 czy MPEG-4 nie odnosi się do kodowania ruchomego obrazu czy dźwięku
- MPEG-21 (ISO/IEC 21000) jest najmłodszym rozszerzeniem rodziny standardów MPEG.
- Jego celem było opracowanie otwartego środowiska dla aplikacji multimedialnych, to jest zdefiniowanie technologii potrzebnej do wspierania użytkowników w zakresie wymiany, dostępu, konsumpcji, handlu lub manipulowania przedmiotami cyfrowymi w wydajny i przejrzysty sposób
- MPEG-21 opiera się na dwóch podstawowych koncepcjach:
  - definicja przedmiotu cyfrowego (podstawowa jednostka dystrybucji i transakcji)
  - użytkownicy wchodzący w interakcje z elementami cyfrowymi
- Elementy cyfrowe mogą być uważane za jądro struktury multimediarów, a użytkowników można uważać za tych, którzy wchodzą z nimi w interakcję w ramach.
- Na najbardziej podstawowym poziomie, MPEG-21 zapewnia strukturę, w której jeden użytkownik wchodzi w interakcję z innym, a przedmiotem tej interakcji jest element cyfrowy.



# Standard MPEG

- W ramach standardów MPEG-1 i MPEG-2 istnieje możliwość wyboru jednego z trzech trybów pracy - warstw (ang. layers).
- Numer warstwy decyduje o uzyskiwanej jakości dźwięku: im wyższy numer, tym wyższa jakość dźwięku, ale też tym wyższa złożoność kodera.



# Standard MPEG

Porównanie przepływności binarnej dla MPEG-1 warstwa III.

Oznaczenia w tabeli: CD – płyta kompaktowa, FM – modulacja częstotliwościowa, AM – modulacja amplitudowa, SW – krótkie fale

Jakość subiektywna	Zakres pasma [kHz]	Rodzaj sygnału	Przepływność binarna [kbps]	Współczynnik kompresji
„telefoniczna”	2,5	monofoniczny	8	1 : 96
lepsza niż SW	4,5	monofoniczny	16	1 : 48
lepsza niż AM	7,5	monofoniczny	32	1 : 24
podobna do FM	11	stereofoniczny	56 ... 64	1 : 26 ... 1 : 24
bliska CD	15	stereofoniczny	96	1 : 16
CD	> 15	monofoniczny	112 ... 128	1 : 14 ... 1 : 12



# Standard MPEG

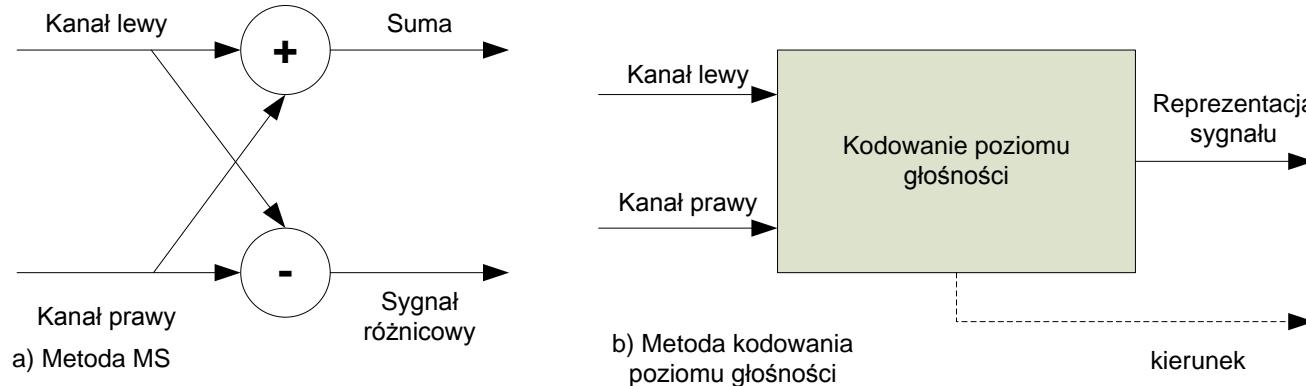
## Kodowanie

- Standard MPEG-1 został zaprojektowany do kodowania sygnałów fonicznych:
  - o częstotliwościach próbkowania 32; 44,1 oraz 48 kHz
  - co najwyżej dwukanałowych
- Z założenia tego wynikają następujące tryby kodowania:
  - Monofoniczny,
  - Stereofoniczny,
  - Dwukanałowy,
  - Stereofoniczny łączny.
- Tryby stereofoniczny i dwukanałowy są sobie równoważne. Oznacza to, że kodowanie sygnału w każdym kanale zachodzi niezależnie od drugiego.
- W przypadku kanałów stereofonicznych sygnały w obu kanałach muszą być skorelowane. Wykorzystując ten fakt, można uzyskać wyższy współczynnik kompresji niż przy oddzielnym kodowaniu poszczególnych kanałów. Zależność ta wykorzystywana jest w trybie łącznym stereofonicznym.

# Standard MPEG

## Kodowanie

- Wyróżnia się dwie metody kodowania stereofonicznego łącznego:
  - kodowanie Mid/Side,
  - kodowanie głośności.
- W kodowaniu MS (Mid/Side) przed kwantyzacją wyznacza się znormalizowaną sumę z kanałów lewego i prawego. Dzięki temu unika się zjawiska zanikania maskowania, występującego w przypadku sygnałów stereofonicznych. Dodatkowo dla silnie skorelowanych sygnałów w obu kanałach nie trzeba przesyłać sygnału różnicowego.



# Standard MPEG

## Kodowanie

- W kodowaniu głośności zachowywana jest tylko informacja o intensywności sygnału w każdym podpaśmie, dla którego wyznaczane są czynniki skalujące.
- Informacja o kierunkowości sygnału jest przesyłana poprzez zakodowanie niezależnych wartości czynników skalujących dla lewego i prawego kanału. W rezultacie transmitowana jest tylko obwiednia energii dla obu kanałów.
- Jest to efektywna metoda kodowania sygnałów przy niskich przepływnościach binarnych, ponieważ jak zauważono wprowadzone w tym procesie zniekształcenia są łatwiejsze do zaakceptowania niż zniekształcenia spowodowane zbyt małą rozdzielczością. Dla pewnych sygnałów, zwłaszcza krótkotrwałych, możliwa jest jednak utrata informacji przestrzennej.



# Standard MPEG

## Kodowanie

- Standard MPEG umożliwia pracę z zadanym trybem przepływności typu:
  - FBR (ang. Fixe Bit Rate): praca ze stałą wartością przepływności,
  - VBR (ang. Variable Bit Rate): prace ze zmienną wartością przepływności.

Tryb FBR wykorzystywany był w pierwszych wersjach standardu. Opcja VBR jest dostępna jedynie w warstwie III i obecnie jest najczęściej używanym trybem.

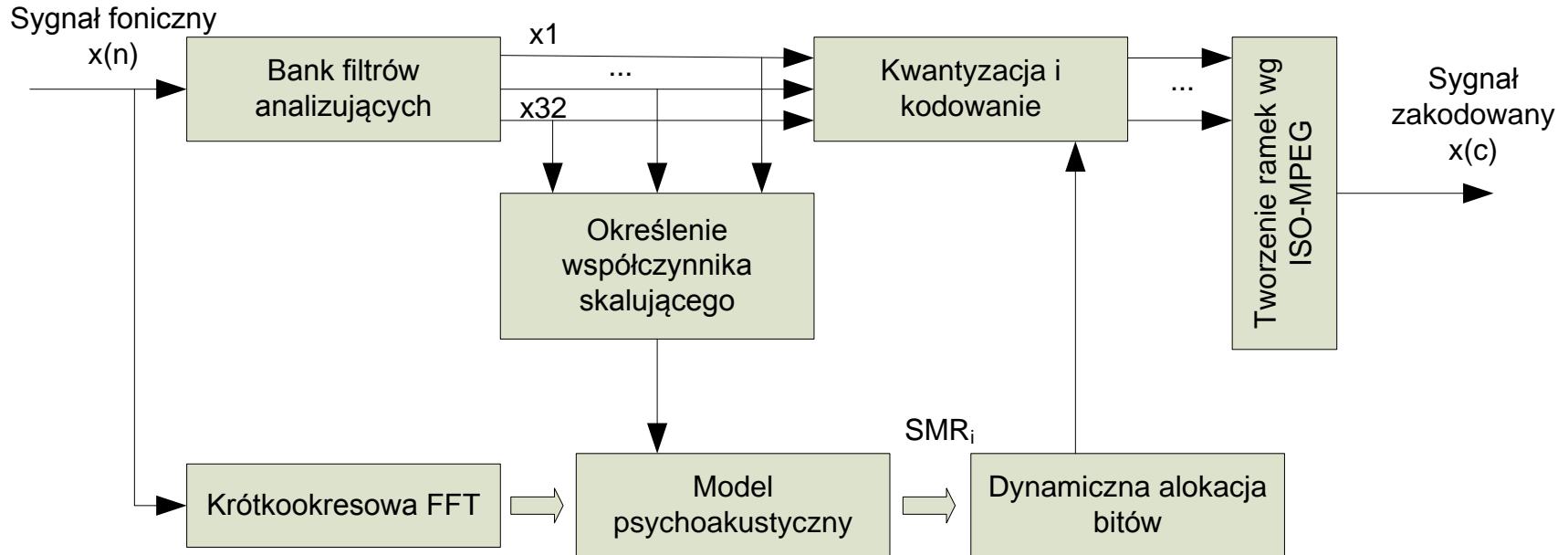
- Podczas kodowania dwa procesy przebiegają równolegle:
  - Analiza sygnału za pomocą banku filtrów analizujących
  - Wyznaczanie krótkookresowej transformaty Fouriera.
- Dzięki drugiemu procesowi uzyskuje się widmo sygnału, które wykorzystywane jest do wyznaczenia progu maskowania, a w konsekwencji do określenia stosunku sygnał-próg maskowania ( $SMR_i$ ), istotnego do utrzymywania szumu kwantyzacji poniżej progu słyszenia.
- Cechą charakterystyczną kodeków MPEG jest to, że modelowanie zjawisk psychoakustycznych odbywa się wyłącznie po stronie kodera. Dzięki temu dekodery są znacznie prostsze, zaś przetwarzanie sygnału odbywa się przy niższym koszcie obliczeniowym. Poza tym, kolejność wykonywanych operacji jest w dekoderze odwrotna niż w koderze.



# Standard MPEG

## Kodowanie

### Schemat kodera MPEG



# Standard MPEG

## Kodowanie

- Bank filtrów analizujących rozdziela sygnał wejściowy na 32 podpasma (analogicznie jak błona podstawnia ślimaka w uchu).
- W standardzie MPEG bank filtrów analizujących oparto na filtrach polifazowych, które umożliwiają redukcję zjawiska nakładania się widm w sąsiednich podpasmach. Ponadto, dla standardu opracowano implementację szybkiej wersji procedur filtracji: analizy i syntezy.
- Dla warstwy I bank filtrów jest typu SOI (ang. FIR) o 512 węzłach. Dzieli on sygnał wejściowy (384 próbki) na 32 podpasma o stałej szerokości po 12 próbek w każdym podpaśmie.
- W warstwie II bank filtrów obejmuje 1024 węzły, przy czym na jego wejście podawany jest sygnał PCM (w blokach po 1152 próbek), a następnie dokonywany jest podział całego zakresu częstotliwości na 32 podpasma po 36 próbek w każdym.



# Standard MPEG

## Kodowanie

- W celu pełniejszego wykorzystania możliwości kwantyzera, próbki są skalowane według współczynnika skalującego (SCF – scale factor). Dzięki temu ich wartości mogą być znormalizowane w zakresie [-1;1].
- W standardzie ISO-MPEG wartości współczynnika skalującego zapisane są na 6 bitach i są stabelaryzowane.
- Największa predefiniowana wartość współczynnika skalowania wynosi 2, zaś stosunek dwóch kolejnych wartości wyraża się zależnością:

$$\frac{scf_{i+1}}{scf_i} = \sqrt[3]{2} \cong 2dB$$

- Dzięki temu pokryty jest cały zakres dynamiki równy 126 dB.
- Po wybraniu największej wartości próbki  $s_{\max}(i)$  w każdym podpaśmie można wyznaczyć czynnik skalujący  $SCF_i$  dla każdego podpasma osobno:

$$\forall_{i=1,\dots,32} \forall_{k=1,\dots,62} scf_k < s_{\max}(i) \leq scf_{k+1} \Rightarrow SCF(i) = scf_{k+1}$$

Gdzie  $scf_k$  jest predefiniowaną wartością czynnika skalującego o indeksie k.



# Standard MPEG

## Kodowanie

- W zależności od trybu pracy kodera MPEG, rozmiar bloku próbek użytych do wyznaczenia DFT (Dyskretnej Transformaty Fouriera) waha się od N=512 (warstwa I) do N=1024 (warstwa II).
- Przed dokonaniem transformacji sygnał wejściowy jest okienkowany przy pomocy funkcji okienkowej Hanninga:

$$w(i) = \gamma \cdot \frac{1}{2} \cdot [1 - \cos\left(\frac{2 \cdot \pi \cdot i}{N-1}\right)]$$

Gdzie: N – liczba próbek w ramce do analizy DFT, i=0, ..., N-1 - indeks

$$\gamma = \begin{cases} \sqrt{\frac{8}{3}}, & \text{dla warstwy I} \\ 1, & \text{dla warstwy II} \end{cases}$$

# Standard MPEG

## Kodowanie

- Model psychoakustyczny wykorzystywany w warstwie I opiera się głównie na badaniach Terhardta, Stolla i Seemanna.
- Modelowanie składa się z następujących etapów:
  - transformacja z liniowej skali częstotliwości (Hz) na perceptualną (Barki) dla kolejnych prążków widma,
  - wyznaczenie progu słyszenia,
  - wyznaczenie składowych tonowych i szumowych,
  - wyznaczenie globalnego progu maskowania,
  - wyznaczenie minimalnego poziomu maskowania dla każdego podpasma,
  - wyznaczenie stosunku sygnału do poziomu maskowania
- Transformacja Hz → Bark dokonywana jest w oparciu o wzór empiryczny Zwickera, zaś wartości progu słyszenia są obliczane na podstawie zależności Terhardta (obie te wartości są tablicowane).



# Standard MPEG

## Kodowanie

- Wyznaczenie progu słyszenia dla każdego prążka widma odbywa się niezależnie, zgodnie z krzywą słyszenia (model Terhardta).
- W kolejnym etapie kodowania wyznacza się składowe tonowe i nietonowe zgodnie z algorytmem:
  - Dla wszystkich prążków widma (poza pierwszym i ostatnim) wyznaczane są maksima lokalne, tj. jeśli wartość prążka przewyższa prążki sąsiednie, to prążek ten dołączany jest do listy kandydatów na składowe tonowe.
  - W tak utworzonej liście wyszukuje się te prążki-kandydatów, które przewyższają sąsiednie o 7 dB.
  - Dla każdego znalezionego prążka usuwa się z listy prążki sąsiednie, a prążek uznaje się za składową tonową. Liczba sąsiednich prążków jest zmienna i zależy od częstotliwości kandydata: im jest ona wyższa, tym więcej prążków sąsiednich.
  - W następnym kroku sumuje się energie składowej tonowej z energią dwóch sąsiednich prążków i wyznaczana jest ich moc sumaryczna w dB. Wartość ta odpowiada mocy wybranej składowej tonowej. Na koniec z widma usuwane są prążki leżące w sąsiedztwie składowej tonowej.
  - W kolejnym kroku w każdym paśmie krytycznym sumowana jest energia prążków nietonowych i jednocześnie prążki te usuwane są z widma. Następnie wyznaczany jest indeks prążka reprezentującego usunięte składowe nietonowe. Na ogół wykorzystuje się tutaj metodę środka ciężkości.
  - Wartość prążka o tak wyznaczonym indeksie i obliczana jest jako suma prążków nietonowych w i-tym paśmie wybranej w dB. Procedura kończy się dołączeniem prążka do listy prążków reprezentujących składowe nietonowe.
- Tak utworzona lista prążków tonowych i nietonowych wykorzystywana jest do wyznaczenia prążków leżących poniżej progu słyszenia (są one usuwane z listy).
- Ponadto, jeśli odstęp między dwoma składowymi tonowymi jest mniejszy od 0,5 Barka, to usuwany jest prążek o niższym poziomie.

# Standard MPEG

## Kodowanie

Aby wyznaczyć globalny próg maskowania, w pierwszym kroku wyznacza się próg maskowania dla kolejnych prążków. Poziom maskowania [dB] dla i-tego prążka obliczany jest następująco:

- Dla j-tego prążka maskującego  $X_{tm}(j)$  o charakterze tonowym

$$LT_{tm} = (j, i) = X_{tm}(j) + a_{vtm}(j) + v_f(j, i)$$

- Dla j-tego prążka maskującego  $X_{nm}(j)$  o charakterze szumowym

$$LT_{nm}(j, i) = X_{nm}(j) + a_{vnm}(j) + v_f(j, i)$$

gdzie  $a_{vtm}$  i  $a_{vnm}$  są indeksami maskowania dla maskera tonowego i szumowego określonymi następująco:

$$a_{vtm} = -1,525 - 0,275 \cdot b(j) - 4,5$$

$$a_{vnm} = -1,525 - 0,175 \cdot b(j) - 0,5$$

$b(k)$  – częstotliwość k-tego prążka wyrażona w Barkach

$v_f(j, i)$  – funkcja maskowania.



# Standard MPEG

## Kodowanie

- Funkcja maskowania opisuje rodzinę krzywych maskowania. Jej konkretna postać wynika ze spostrzeżenia, że w praktyce zakresy maskowania są mniejsze od 3 i 8 barków w przypadku maskowania składowych o odpowiednio niższych oraz wyższych częstotliwościach.
- Na podstawie znajomości dolnego progu słyszenia  $LT_q(i)$  dla częstotliwości odpowiadającej i-temu prążkowi oraz wpływu na maskowanie przez składowe tonowe  $LT_{Tm}(j,i)$  oraz szumowe  $LT_{nm}(j,i)$  można wyznaczyć globalny próg maskowania dla i-tego prążka:

$$LT_G(i) = 10 \log_{10} \left[ 10^{\frac{LTq(i)}{10}} + \sum_{j=1}^{Tm} 10^{\frac{LTtm(j,i)}{10}} + \sum_{j=1}^{Rm} 10^{\frac{LTnm(j,i)}{10}} \right]$$

Gdzie:

$LT_q$  – absolutny próg słyszenia w dB

Tm i RM – liczba składowych tonowych i szumowych



# Standard MPEG

## Kodowanie

Wyznaczenie stosunku sygnału do poziomu maskowania odbywa się następująco:

- Na podstawie modelu psychoakustycznego wyznaczany jest minimalny poziom maskowania w każdym paśmie krytycznym, a następnie, w każdym podpaśmie, wyszukuje się prążka o największej wartości  $X_{\max}(i)$  i oblicza się jego moc korzystając z zależności:

$$L_{\max}(i) = \max[X_{\max}(i), 20 \cdot \log_{10}(scf_{\max}(i) \cdot 32768) - 10]$$

- Na tej podstawie można wyznaczyć stosunek sygnału do poziomu maskowania w każdym podpaśmie zgodnie ze wzorem:

$$SMR(i) = L_{\max}(i) - LT_{G \min}(i),$$

dla  $i = 1, 2, \dots, 32$



# Standard MPEG

## Kodowanie

- Współczynnik kompresji w metodzie MPEG (i odpowiadająca mu wartość przepływności binarnej) są zdeterminowane przez rozmiar kodowanej ramki. Dlatego liczba bitów kodujących sygnał nie może przekroczyć rozmiaru ramki.
- Zadaniem modułu dynamicznej alokacji bitów jest taki ich przydział, aby w warunkach ograniczonego rozmiaru ramki zachować jak najlepszą jakość sygnału fonicznego.
- Jakość sygnału fonicznego wyraża się stosunkiem poziomu maskowania do poziomu szumu wprowadzonego podczas kwantyzacji – MNR (Masking to Noise Ratio):

$$MNR_i = SNR_i - SMR_i,$$

gdzie:

$i$  – numer podpasma,

$SMR_i$  – stosunek sygnału do poziomu maskowania, wyznaczony na podstawie modelu psychoakustycznego,

$SNR_i$  – predefiniowane estymaty stosunku sygnału do poziomu szumu. Wartości te są związane z liczbą poziomów kwantyzera. Dopuszcza się również inne metody wyznaczania estymaty  $SNR$ .



# Standard MPEG

## Kodowanie

- Dla zachowania jak najlepszej jakości sygnału optymalizuje się współczynnik MNR dla każdego pasma.
- Optymalizacja rozpoczyna się od pasma, w którym wartość MNR jest najmniejsza.
- Proces trwa dopóki nie zostaną zoptymalizowane wszystkie współczynniki MNR albo nie zostaną wyczerpane wszystkie dostępne bity użyte w procesie kodowania.
- Aby utrzymać szum kwantowania poniżej progu słyszalności, w koderze MPEG stosuje się kwantowanie nielinowe.
- Tak przygotowane dane łączy się, dodając dane sterujące. Umożliwiają one dekoderowi odkodowanie dźwięku i mówią, które z pasm z jaką dokładnością zostały zakodowane.
- Dalej dane audio i dane sterujące kompresuje się za pomocą algorytmu Huffmmana (bezstratna metoda kompresji).

# Standard MPEG

## Kodowanie

- Ten model kodowania jest możliwy do osiągnięcia. Jednak często okazuje się, że użytkownik określając maksymalną przepływność nie zapewnił możliwości tak dokładnego zakodowania dźwięku.
- Jeśli po kompresji okaże się, że rozmiar danych przekracza założenia, zmniejsza się liczbę bitów, w których koduje się pasma.
- Dlatego najlepsze efekty jakościowe dają kompresory działające ze zmienną przepływnością VBR, „oszczędzając” miejsce na dane w momentach kiedy się da, a zwiększając przepływność wtedy, gdy to jest konieczne.
- Subiektywna jakość dźwięku zależy od modelu psychoakustycznego zaimplementowanego w koderze oraz od słuchacza.
- Jeśli jego słuch będzie bardzo odbiegał od modelu użytego w kodeku, oceniona jakość dźwięku może być odebrana jako znacznie różna od oryginału.



# Standard MPEG

## Zaawansowane kodowanie audio

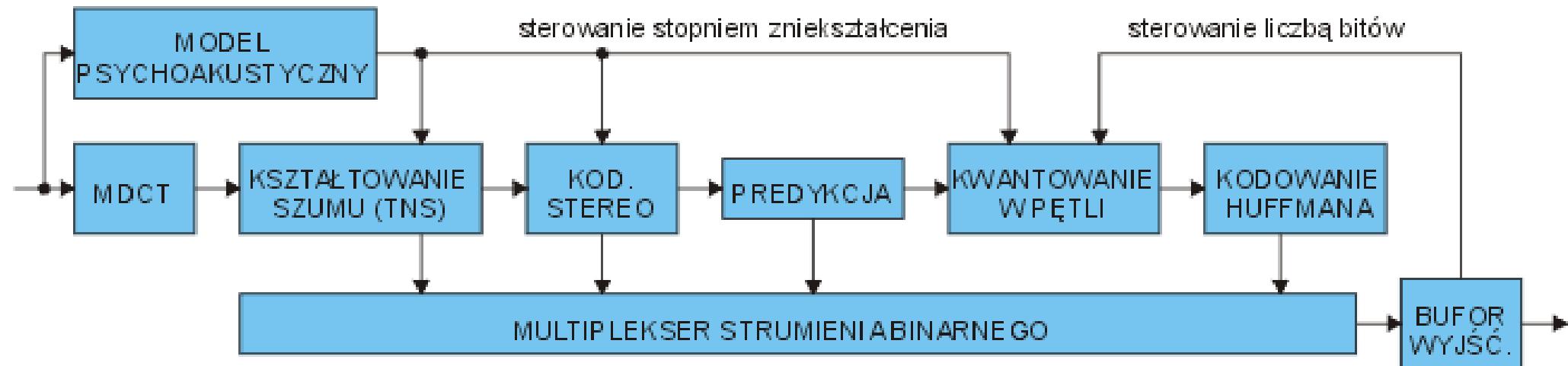
- Algorytm AAC (ang. Advanced Audio Coding) został zaprojektowany jako następca MP3, oferujący lepszą jakość dźwięku przy podobnym rozmiarze danych.
- AAC został zdefiniowany przez normę ISO/IEC 13818-7 w roku 1997 jako część specyfikacji MPEG-2 część 7 (MPEG-2 NBC – nie kompatybilny z wcześniejszymi rozwiązaniami).
- W roku 1999 MPEG-2 część 7 został zmodyfikowany i dołączony do standardu MPEG-4 część 3 (MPEG-4 Audio or ISO/IEC 14496-3).
- Częścią standardu MPEG-Audio są algorytmy AAC oraz HE-AAC (AAC +), zaadoptowane do cyfrowych standardów radiowych DAB + i Digital Radio Mondiale oraz standardów telewizji mobilnej DVB-H i ATSC-M / H .
- Standard MPEG-4 Audio jest złożonym narzędziem do przetwarzania dźwięku od kodowania mowy o niskiej przepływności binarnej do wysokiej jakości kodowania audio oraz syntezy muzyki.



# Standard MPEG

## Zaawansowane kodowanie audio

### Uproszczony schemat blokowy kodera MPEG-4 AAC



# Standard MPEG

## Zaawansowane kodowanie audio

Najważniejsze różnice między MPEG-1 a MPEG-4:

1. W AAC sygnał wejściowy dzielony jest adaptacyjnie na nachodzące na siebie bloki złożone z 2048 lub 512 próbek w zależności od chwilowych właściwości sygnału.
2. Każdy blok sygnału reprezentowany jest przez współczynniki zmodyfikowanej transformaty kosinusowej (MDCT)
3. Celem kształtowania szumu (TNS) w dziedzinie czasu jest minimalizacja tzw. Efektu echa wyprzedzającego. Narzędzie TNS przekształca wstępnie współczynniki MDCT tak, że ich kwantyzacja odbywa się w sposób różnicowy. Dzięki czemu obwiednia szumu kwantyzacji przypomina sygnał pierwotny i przez to jest skutecznie maskowana.
4. Do kodowania dźwięku wielokanałowego używany jest tryb stereofoniczny łączny



# Standard MPEG

## Zaawansowane kodowanie audio

5. Kwantyzacja współczynników transformaty MDCT jest sterowana w koderze przez model psychoakustyczny, który decyduje o tym, ile zniekształceń kwantowania można wprowadzić do poszczególnych podzakresów częstotliwości bez powodowania znaczących i słyszalnych artefaktów. Model psychoakustyczny oblicza bieżący kształt krzywej maskowania.
6. Aby utrzymać poziom szumu kwantowania poniżej progu maskowania, współczynnikom MDCT w poszczególnych przedziałach częstotliwości przydziela się odpowiednią liczbę bitów (różna rozdzielcość kwantyzacji dla różnych przedziałów).
7. Predykcja w MPEG AAC wykorzystuje podobieństwo odpowiadających sobie współczynników transformaty, które można zaobserwować pomiędzy kolejnymi blokami, gdy sygnał jest stacjonarny.
8. Do przewidywania wartości współczynników w następnym bloku na podstawie współczynników z bloków poprzednich wykorzystywany jest zbiór predyktorów adaptacyjnych (każdy współczynnik ma niezależny predyktor), który działa identycznie w koderze i dekoderze. Przesłanie błędu predykcji (zazwyczaj o mniejszej wartości) zamiast bezwzględnej wartości współczynnika, statystycznie wymaga znacznie mniejszej liczby bitów przy zachowaniu tej samej dokładności reprezentacji, a dzięki temu oferuje lepszą efektywność.

# Standard MPEG

## Zaawansowane kodowanie audio

8. W koderze AAC wykorzystano kwantyzator o stałej charakterystyce z nierównomierną szerokością przedziału kwantyzacji: wartości współczynników bliskie zeru są kwantowane z większą dokładnością, duże wartości są traktowane bardziej zgrubnie.
9. Stopień kwantyzacji jest sterowany poprzez odpowiednie skalowanie współczynników MDCT. Aby odtworzyć właściwe amplitudy współczynników transformatowych, czynniki skalujące muszą być również przesłane do dekodera. Ze względu na ilość dodatkowo transmitowanych danych, skalowaniu podlegają łącznie grupy współczynników w ramach tzw. pasma skalowania, dzięki czemu pojedynczy czynnik skalujący opisuje pewną liczbę współczynników transformaty.
10. Kwantyzacja i następujące po niej kodowanie współczynników jest źródłem zazwyczaj nieprzewidywalnej liczby bitów gromadzonych w buforze wyjściowych (co wynika z kodowania entropijnego słowami kodowymi o różnej długości). Dlatego w celu osiągnięcia docelowej prędkości transmisji, proces kwantowania jest iteracyjnie poprawiany z sukcesywnie dobieranym globalnym czynnikiem skalującym kwantyzatora.
11. Ta iteracyjna procedura jest najbardziej złożoną obliczeniowo częścią kodera. Można ją uprościć, ale skutkiem jest pogorszenie jakości zakodowanego dźwięku.



# Techniki przetwarzania mediów cyfrowych

Wykład 4: Synteza dźwięku (4 h)



Fundusze  
Europejskie  
Wiedza Edukacja Rozwój



Politechnika Wrocławskiego



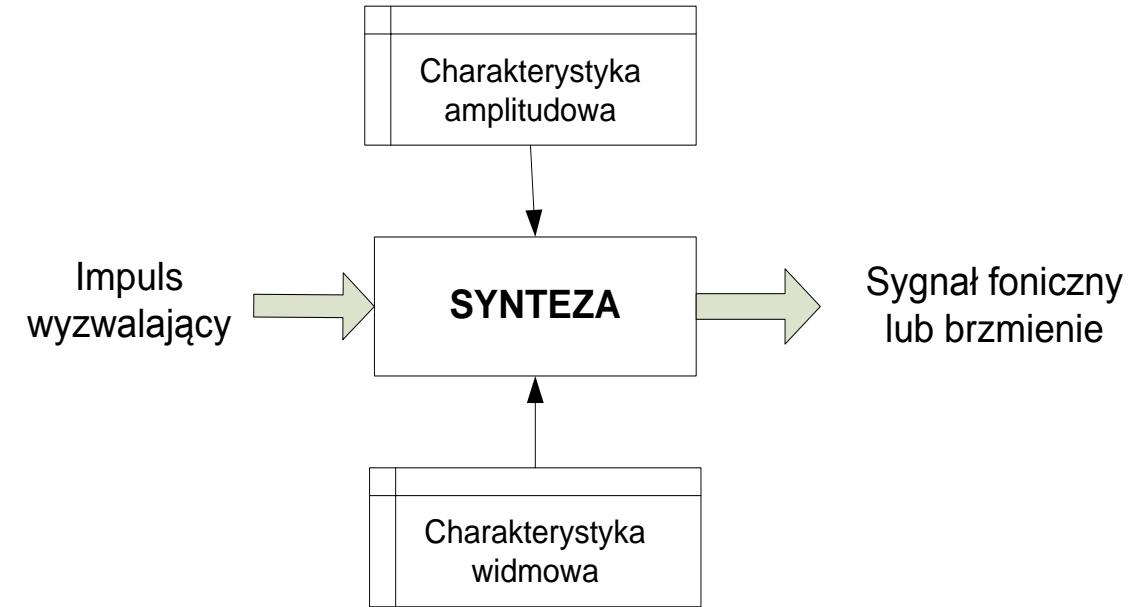
Unia Europejska  
Europejski Fundusz Społeczny

# Agenda

- Wprowadzenie
- Elektroniczne instrumenty muzyczne
- Klasyfikacja metod generowania dźwięku ze względu na liczbę brzmień, polifonię, metody i algorytmy syntezy
- Klasyfikacja algorytmów generowania dźwięku:
  - Przetwarzanie zapisu: metoda tablicowa i samplingowa
  - Metody widmowe: synteza addytywna, synteza substratywna
  - Algorytmy abstrakcyjne: synteza FM, przekształcanie fali, metody oparte o chaos deterministyczny
  - Modelowanie fizyczne: modelowanie matematyczne, falowodowe, metoda komórkowa.

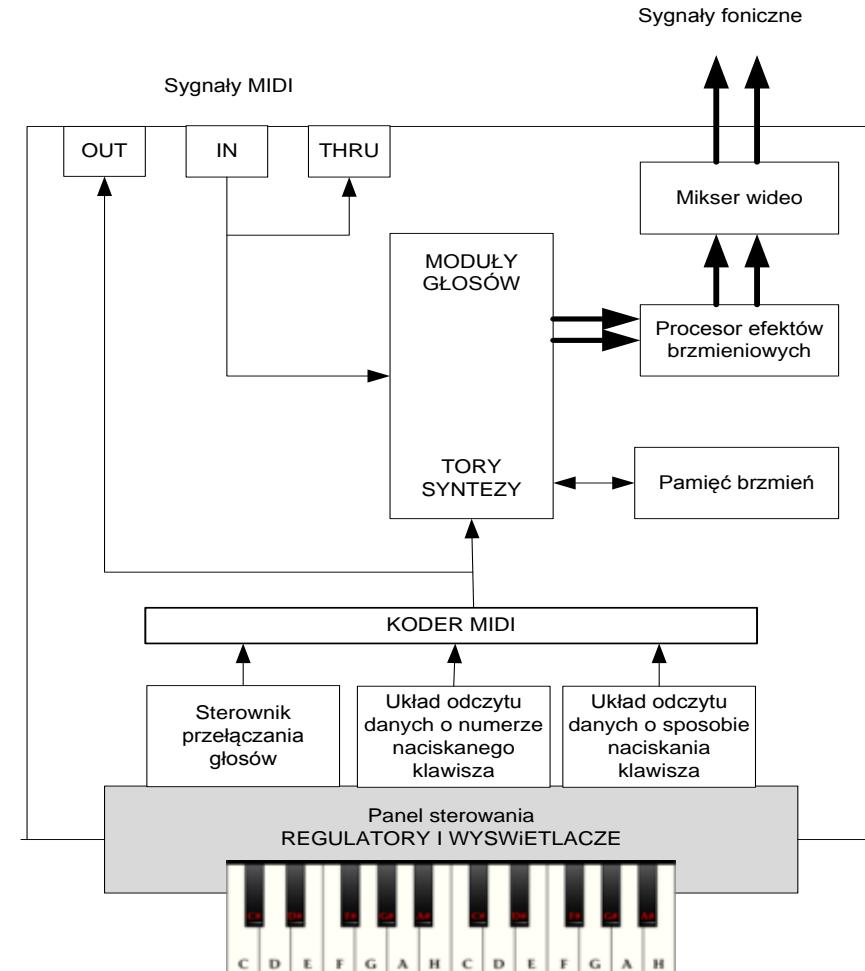
# Wprowadzenie

- Synteza dźwięku jest procesem umożliwiającym wytwarzanie sygnału fonicznego i kształtowanie brzmień na podstawie zbioru parametrów.
- Parametry te opisują w sposób bezpośredni lub pośredni charakterystykę amplitudową wytwarzanego dźwięku oraz jego charakterystykę widmową.
- Charakterystyka amplitudowa to inaczej obwiednia dźwięku.
- Charakterystyka widmowa mówi o tym, jakie są proporcje między poszczególnymi składnikami widma.



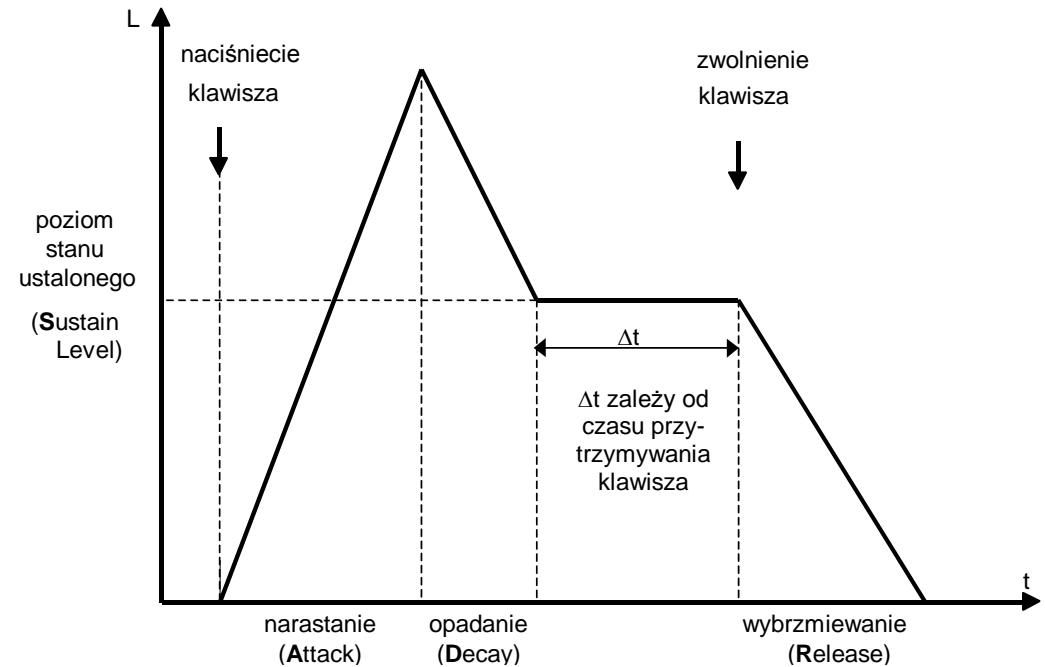
# Elektroniczne instrumenty muzyczne

- Instrumenty elektroniczne (syntezatory) mogą być urządzeniami analogowymi lub cyfrowymi.
- Niezależnie od technologii wykonania, ich ogólna architektura oraz funkcje poszczególnych podzespołów są podobne.
- Najważniejszymi podzespołami syntezatorów dźwięku są moduły głosów (tory syntezy).
- Tory syntezy to bloki, w których wytwarzany jest dźwięk zgodnie z przyjętym algorytmem syntezy. Najczęściej modułami tymi są specjalizowane układy scalone. Od ich jakości oraz efektywności zastosowanych algorytmów zależy ostateczna ocena jakości brzmienia instrumentu oraz jego możliwości (polifonia, dynamika, możliwości artykulacyjne).
- Większość syntezatorów (niezależnie od zaimplementowanego w nich algorytmu syntezy) w swych torach generacji zawiera wiele identycznych lub bardzo podobnych bloków.



# Elektroniczne instrumenty muzyczne

- Jednym z takich bloków wspólnych jest generator obwiedni (ang. Envelope Generator – EG).
- EG pozwala na simulację zmian wybranych parametrów syntezy w czasie (np. amplitudy, głębokości vibrato, częstotliwości odcięcia filtru).
- Działanie generatora obwiedni polega na tym, że w wyniku pojawienia się impulsu wyzwalającego, pochodzącego z klawiatury muzycznej, układ generatora wytwarza pojedynczy przebieg o kształcie standardowej obwiedni ADSR.
- Użytkownik może ustawić wartości faz A, D i R podając czasy ich trwania.
- Faza S zależy od czasu przytrzymywania klawisza instrumentu przez muzyka.



# Elektroniczne instrumenty muzyczne

- Ostatecznie sygnał dźwiękowy przetwarzany jest w procesorze efektów muzycznych, gdzie może być wzbogacony o echo, pogłos czy efekt chóralny.
- Innym, ważnym układem modyfikującym parametry dźwięku jest generator wolnych przebiegów (ang. Low Frequency Oscillator – LFO).
- LFO tworzy przebieg sinusoidalny o małej, rzędu pojedynczych Hz, częstotliwości.
- Wykorzystuje się go do tworzenia takich efektów jak vibrato, tremolo lub efektu cyklicznych zmian barwy w czasie.
- W zależności od tego, który z parametrów dźwięku jest modulowany przez LFO, uzyskuje się odpowiedni efekt:
- vibrato – LFO moduluje częstotliwość tworzonego dźwięku,
- tremolo – LFO moduluje amplitudę,
- cykliczne zmiany barwy – LFO moduluje częstotliwość odcięcia filtra.

# Klasyfikacja metod generowania dźwięku

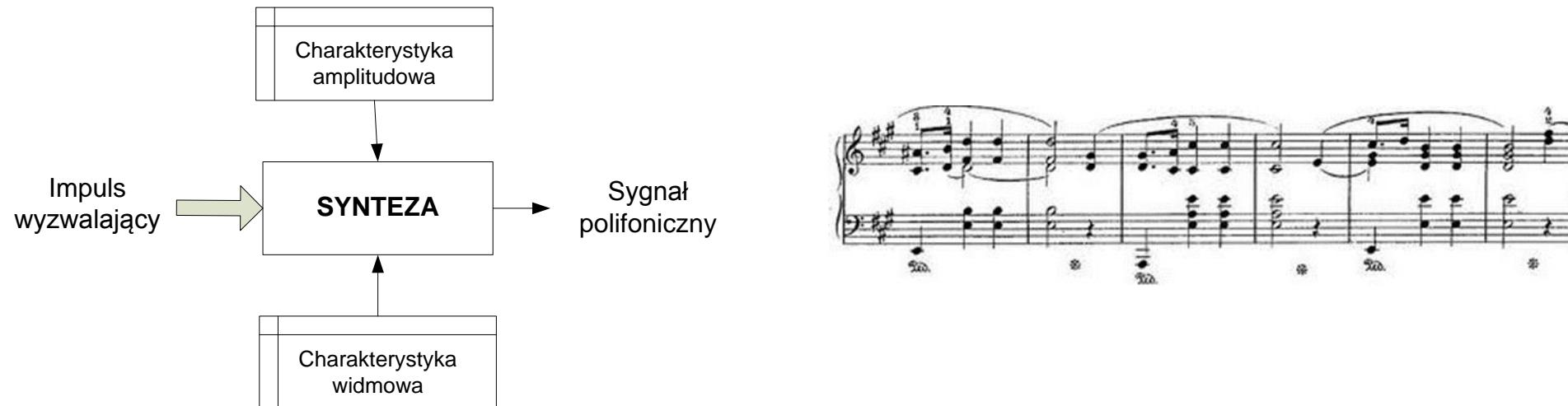
- Ze względu na liczbę dźwięków generowanych w jednostce czasu metody syntezy dzieli się na monofoniczne (homofoniczne) i polifoniczne.
- Synteza monofoniczna polega na tym, że w systemie wytwarzany jest w danym momencie czasu tylko jeden dźwięk - tak, jak w przypadku niektórych instrumentów naturalnych, np. trąbki czy saksofonu.
- Przykładami syntezatorów monofonicznych mogą być najstarsze instrumenty analogowe.



Minimoog – jeden z najpopularniejszych syntezatorów analogowych, [Wikipedia]

# Klasyfikacja metod generowania dźwięku

- Synteza polifoniczna pozwala na generowanie kilku różnych dźwięków jednocześnie.
- W rzeczywistych instrumentach taki sposób grania (akordami) jest typowy dla fortepianu, akordeonu czy organów.
- Współczesne syntezatory nie są w pełni polifoniczne. Generują równocześnie tyle dźwięków, ile mają głosów (tj.



# Klasyfikacja metod syntezy

- Ze względu na liczbę generowanych brzmień metody syntezy dźwięku dzieli się na jedno- i wielobrzemieniowe (ang. Multi-timbral).
- Synteza jednobrzemieniowa umożliwia wytwarzanie dźwięków tylko o jednym brzmieniu, np. fortepianu.
- Synteza wielobrzemieniowa pozwala na jednoczesne wytwarzanie dźwięków o różnych brzmieniach, np. fortepianu równocześnie z sekcją instrumentów smyczkowych.
- Syntezatory wielobrzemieniowe mogą być wykorzystywane do generowania dźwięków tworzonych przez rozbudowany zespół instrumentalny.

Synteza jednobrzemieniowa



Synteza wielobrzemieniowa



Fundusze  
Europejskie  
Wiedza Edukacja Rozwój



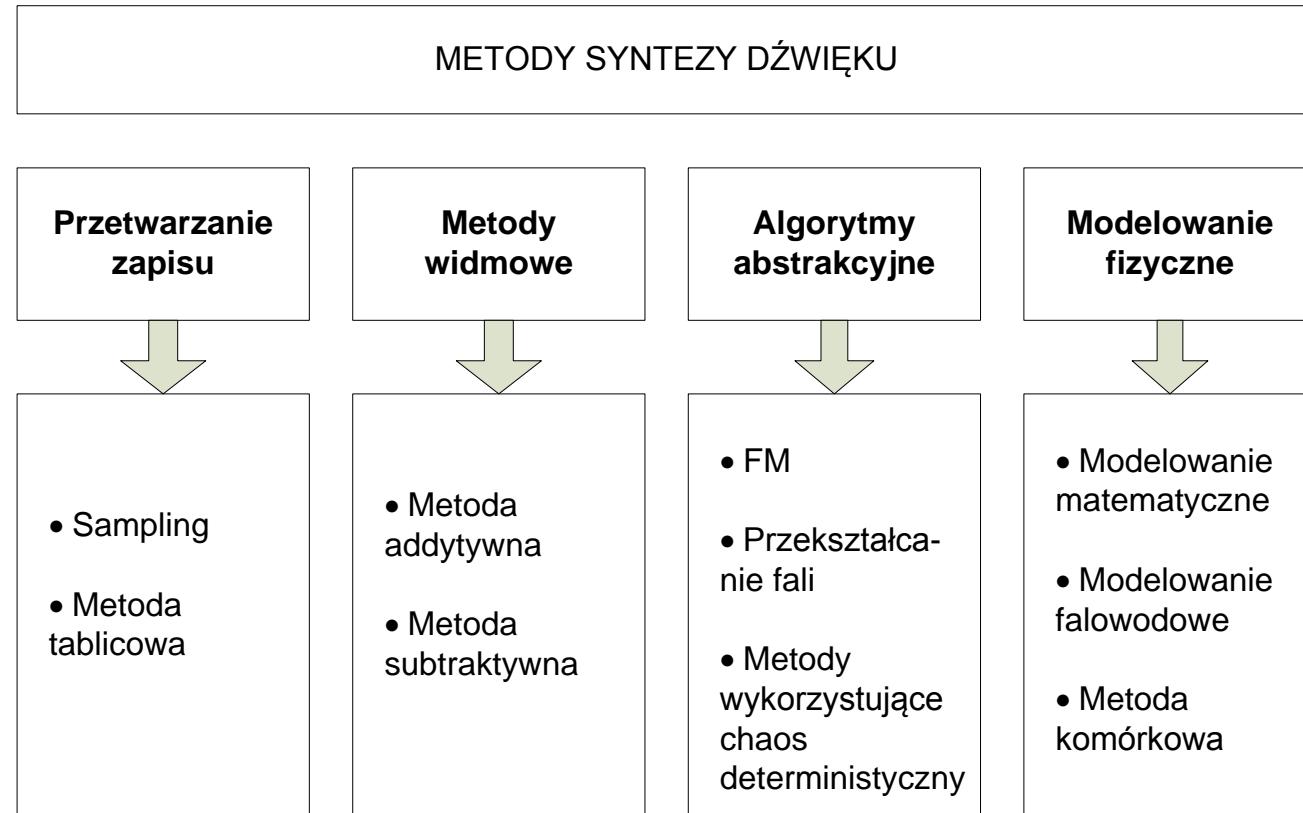
Politechnika Wrocławska

Unia Europejska  
Europejski Fundusz Społeczny



# Klasyfikacja metod generowania dźwięku

Klasyfikacja metod generowania dźwięku ze względu na algorytm syntezy



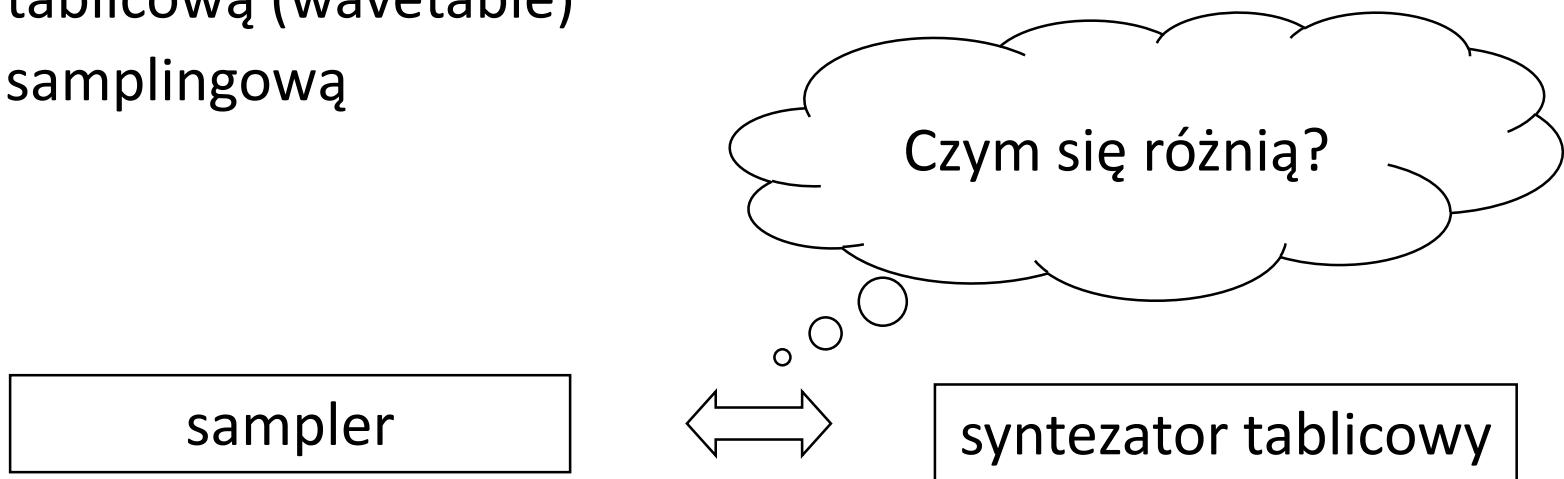
# Metody generowania dźwięku na podstawie przetwarzania zapisu

- Metody generowania dźwięku na podstawie przetwarzania zapisu wykorzystują tzw. banki brzmień.
- Banki są tablicami, w których przechowuje się próbki rzeczywistych instrumentów muzycznych.
- Próbka (ang. sample) w kontekście metod przetwarzania zapisu ma dwojakie znaczenie:
  - Zakodowane cyfrowo brzmienie nuty o określonej wysokości wytwarzanej na konkretnym instrumentem muzycznym (określenie odnosi się do próbek w syntezie wavetable)
  - Cyfrowy fragment wcześniejszej dokonanego nagrania muzycznego, najczęściej stanowią je pojedyncze dźwięki instrumentów naturalnych (np. sample perkusyjne) lub odpowiednio wycięte części innych utworów. Sample uzyskuje się w sposób naturalny lub sztuczny.
- Proces syntezy dźwięku sprowadza się do odtwarzania próbek zgodnie z określonym algorytmem, a następnie ich konwersji na postać analogową.

# Metody generowania dźwięku na podstawie przetwarzania zapisu

W grupie metod generowania dźwięku na podstawie zapisu wyróżnia się:

- Metodę tablicową (wavetable)
- Metodę samplingową



syntetyzer tablicowy = sampler bez możliwości nagrywania dźwięków i ich edycji

# Metody generowania dźwięku na podstawie przetwarzania zapisu

- W instrumentach działających na zasadzie syntezy wavetable cały proces związany z próbkowaniem dźwięków instrumentów akustycznych oraz zapisywaniem ich w pamięci wykonuje producent banku brzmień.
- Użytkownik nie ma wpływu na sposób rejestracji i syntezy zapisanych dźwięków.
- Inaczej jest w przypadku dźwięku generowanego metodą samplingową.
- Samo „samplowanie” (ang. sampling) polega na wykorzystaniu fragmentu wcześniej dokonanego nagrania muzycznego, zwanego samplem (próbką), jako elementu nowo tworzonego utworu, przy wykorzystaniu specjalnego instrumentu zwanego samplerem.

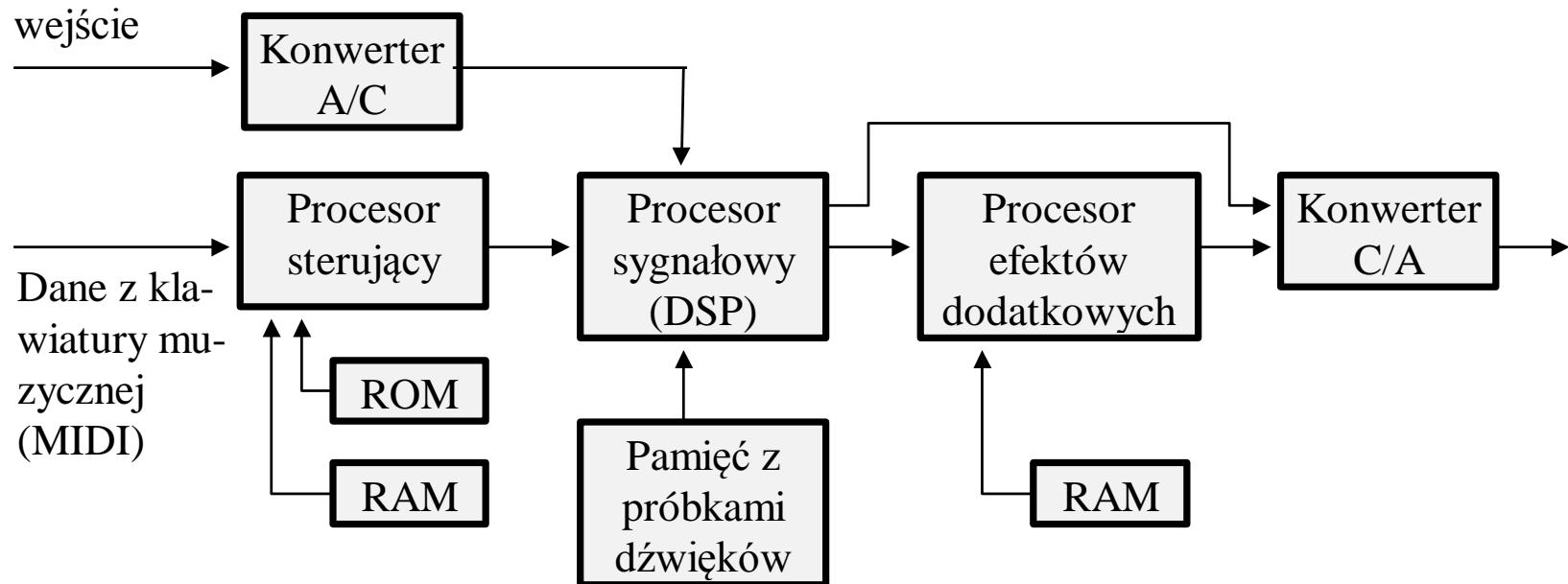
# Metody generowania dźwięku na podstawie przetwarzania zapisu

- Sampler to elektroniczny lub cyfrowy instrument muzyczny, który wykorzystuje w procesie tworzenia pojedyncze oraz wielowarstwowe nagrania dźwiękowe, tak zwane sample czyli próbki i fragmenty rzeczywistych dźwięków (np. może to być pojedynczy dźwięk gitary basowej, skrzypiec lub trąbki ale też rozbudowana melodia czy fragment nagrania muzycznego).
- W procesie tworzenia muzyk producent rejestruje inspirujące go sample – próbki, by następnie załadować je na swój sampler.
- Kolejnym procesem jest już praca na samym samplerze, gdzie za pomocą odpowiedniego oprogramowania, sekwencera lub innego urządzenia wyzwalającego oraz klawiatury MIDI dźwięki zostają zagrane.
- Ponadto istnieje możliwość modyfikowania szeregu parametrów związanych ze sposobem odtwarzania uprzednio nagranych dźwięków.
- Dzięki temu, że sample przechowywane są w pamięci cyfrowej można szybko uzyskać do nich dostęp, co wypływa istotnie na komfort procesu komponowania.



# Metody generowania dźwięku na podstawie przetwarzania zapisu

## Architektura typowego samplera



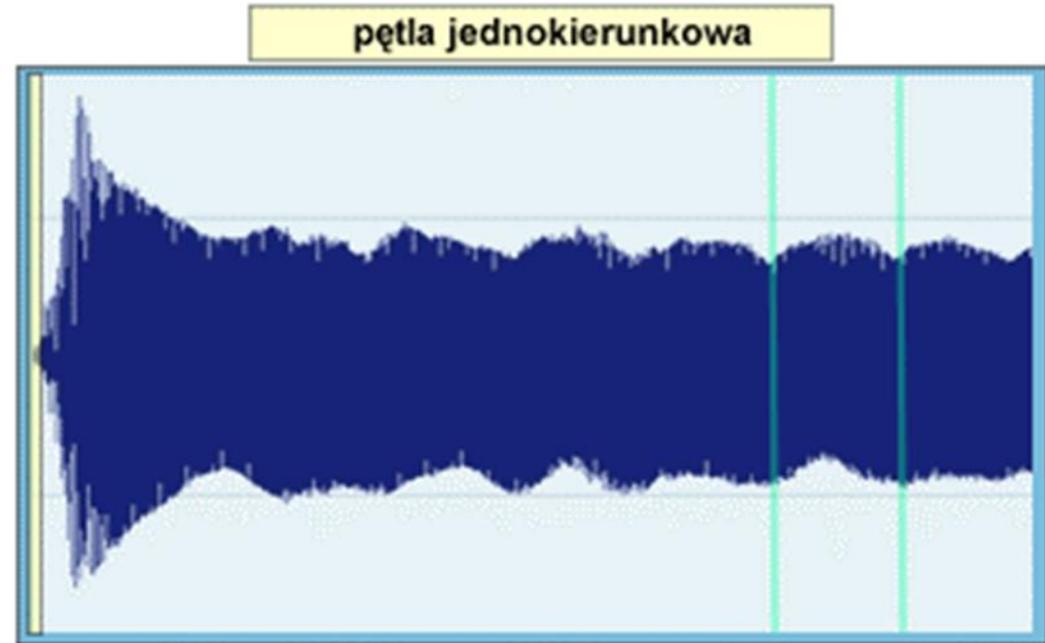
# Metody generowania dźwięku na podstawie przetwarzania zapisu

- Ze względu na duże wymagania dotyczące pojemności pamięci próbek dźwiękowych w obu metodach: wavetable i samplingowej stosuje się pewne rozwiązania kompromisowe.
- Pierwsze z nich polega na tym, że w pamięci próbek zapisuje się jedynie wybrane (reprezentatywne) dźwięki danego instrumentu z poszczególnych zakresów skali muzycznej (np. co kwintę).
- „Brakujące” dźwięki uzupełnia się przez transpozycję dźwięków uprzednio spróbkowanych.
- Proces ten dokonywany jest przy wykorzystaniu procesora sygnałowego DSP na podstawie algorytmów interpolacyjno - decymacyjnych.
- Drugi kompromis związany z ograniczeniami pamięci polega na tym, że w pamięci instrumentu umieszcza się stosunkowo krótkie fragmenty dźwięków, co pozwala na oszczędne wykorzystanie pamięci.
- Sztuczne wydłużenie spróbkowanych dźwięków uzyskuje się poprzez ich odpowiednie zapętlenie.
- W zależności od typu dźwięku (a raczej jego obwiedni) oraz odległości między zapętlonymi punktami, stosowane są alternatywnie trzy rodzaje pętli: jednokierunkowa, dwukierunkowa oraz pętla z przenikaniem.



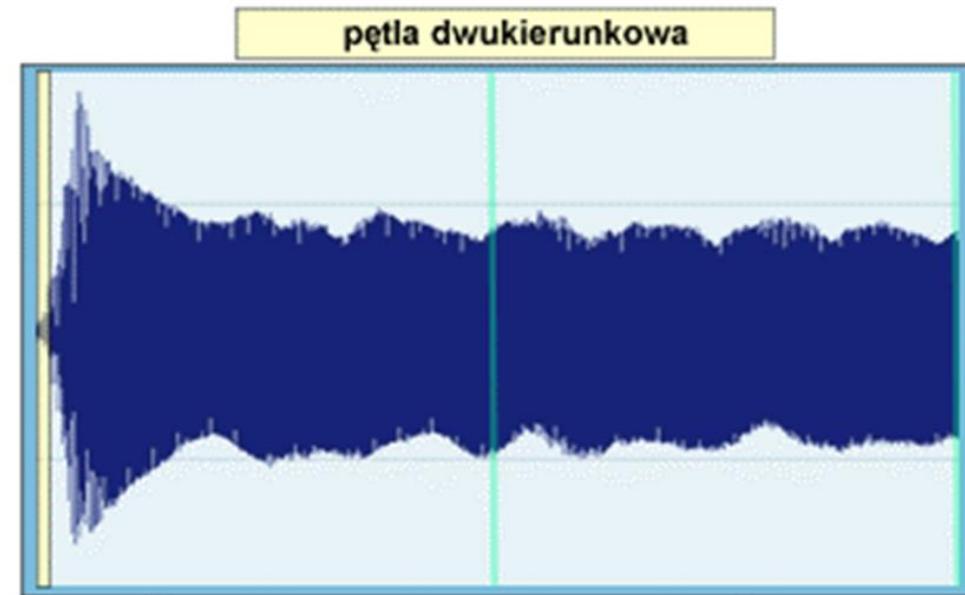
# Metody generowania dźwięku na podstawie przetwarzania zapisu

- W pętli jednokierunkowej następuje przeskok od końca pętli do jej początku.
- Pętla jednokierunkowa może wprowadzać zniekształcenia brzmienia dźwięku (skoki amplitudy). Dobrze nadaje się dźwięków okresowych.



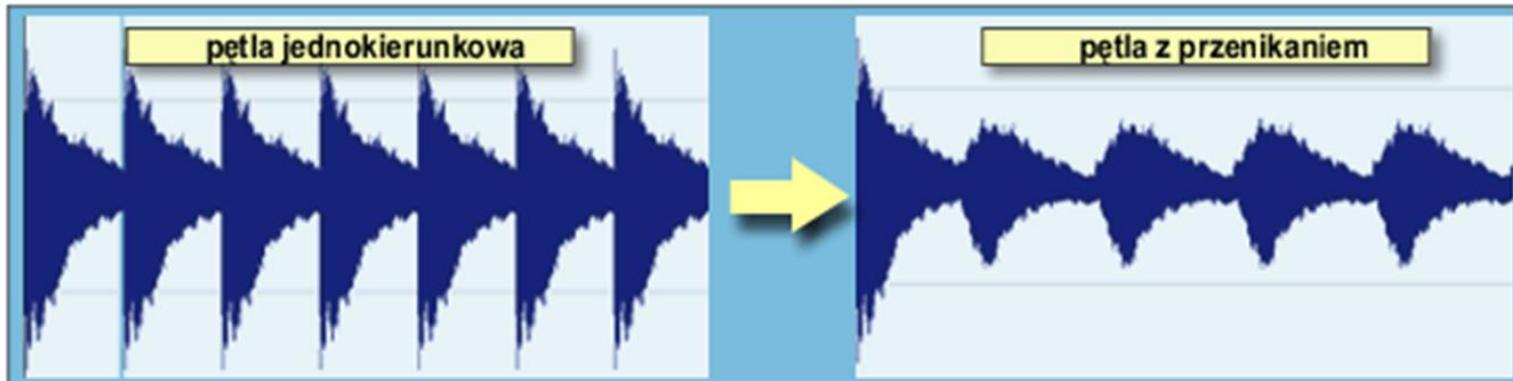
# Metody generowania dźwięku na podstawie przetwarzania zapisu

- W pętli dwukierunkowej po dojściu do końca pętli, zapętlony fragment jest odtwarzany od tyłu.
- Ten typ pętli może powodować, że dźwięk brzmi nienaturalnie. Mogą wystąpić zniekształcenia fazowe. Nadaje się do krótkich fragmentów.



# Metody generowania dźwięku na podstawie przetwarzania zapisu

- Pętla z przenikaniem (*cross-fade*) jest podobna do pętli jednokierunkowej.
- Przy przejściu końcowy i początkowy odcinek pętli nakładane są na siebie. Dzięki temu uzyskuje się bardziej płynne przejście dźwięku.



# Metody generowania dźwięku na podstawie przetwarzania zapisu

- Wszystkie algorytmy wykorzystujące zapętleńcia działają podobnie.
- Po naciśnięciu klawisza na klawiaturze muzycznej odtwarzany jest początkowy fragment dźwięku (faza attack – narastania), po czym dźwięk odtwarzany jest cyklicznie pomiędzy dwoma punktami granicznymi.
- Proces powtarzania trwa dopóty, dopóki przytrzymywany jest klawisz. Po jego zwolnieniu odtwarzany jest pozostały fragment dźwięku.
- Jakość syntezy zależy od doboru punktów, pomiędzy którymi odbywa się cykl.
- Wybór odpowiedniego rodzaju pętli oraz związanych z nią optymalnych punktów granicznych nie jest rzeczą prostą.
- W wielu przypadkach dźwięków instrumentów naturalnych w fazie quasi ustalonej (sustain) charakteryzuje się określonymi fluktuacjami zarówno częstotliwości jak i amplitudy, co utrudnia dobór optymalnych warunków granicznych pętli.



# Metody generowania dźwięku na podstawie przetwarzania zapisu

## Zalety:

Ze względu na to, że metody wavetable i sampling oparte są na odtwarzaniu i przetwarzaniu nagrań, do ich zalet należy wysoka wierność brzmienia

## Wady:

Trudności implementacyjne związane z artykulacją, szczególnie w przypadku instrumentów dętych.



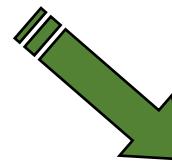
# Widmowe metody generowania dźwięku

- W metodach widmowych proces generowania dźwięku polega na rekonstrukcji sygnału na podstawie parametrów widma sygnału rzeczywistego, uzyskanych na drodze analizy widmowej.
- Wyróżnia się dwa rodzaje metod widmowych:
  - Synteza addytywna
  - Synteza substraktywna
- Synteza addytywna polega na „budowaniu” widma dźwięku z pojedynczych sygnałów harmonicznych, z których każdy może być modulowany amplitudowo i fazowo.
- Wyróżnia się dwie odmiany syntezy addytywnej:
  - PV (ang. Phase Vocoder)
  - MQ (od nazwisk jej twórców McAulay'a i Quatieri)

# Widmowe metody generowania dźwięku

Dwie odmiany metody addytywnej

metoda addytywna



PV

- ang. Phase Vocoder
- $f_n = nf_1$

MQ

- McAulay i Quatieri
- $f_n$  nie musi być  $= nf_1$



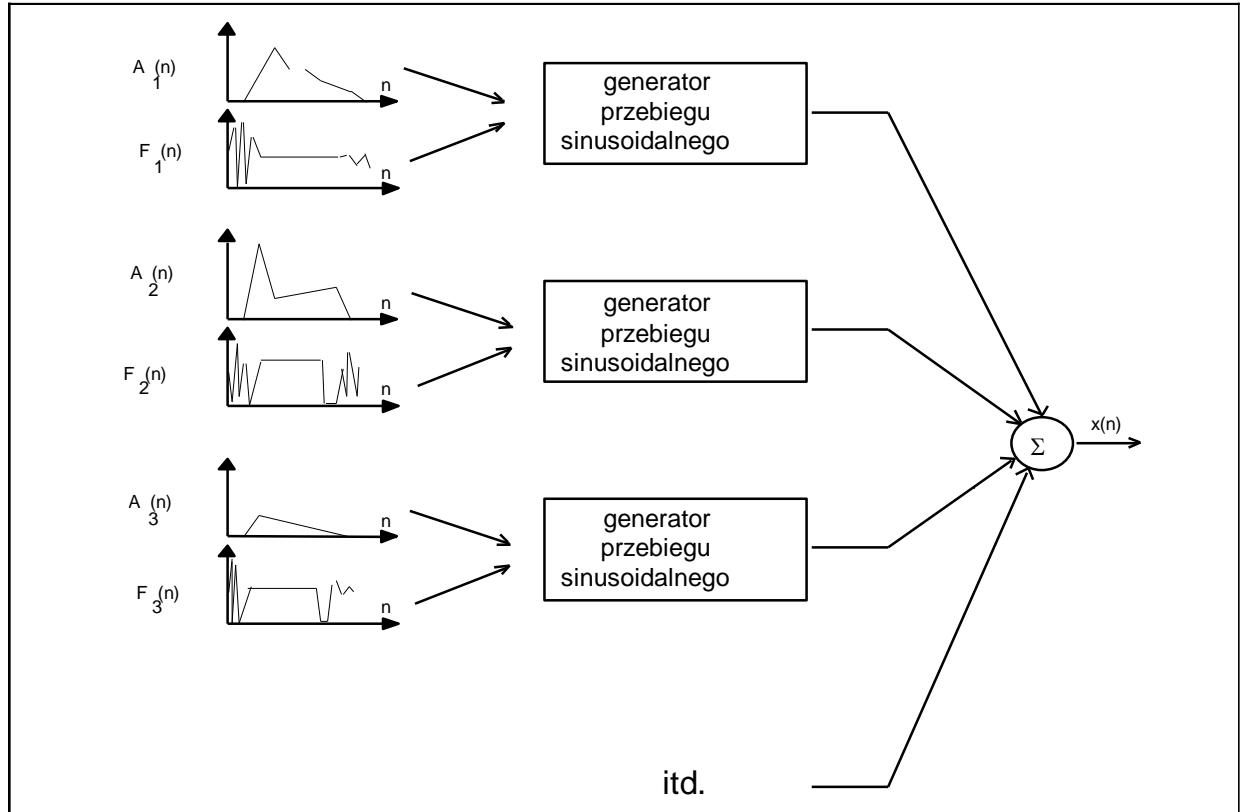
# Widmowe metody generowania dźwięku

- W wokoderze fazowym PV zakłada się, że generowany sygnał jest okresowy.
- W procesie analizy dokonuje się ekstrakcji składowych widma oraz oblicza się częstotliwości poszczególnych składowych harmonicznych.
- Na podstawie tych parametrów dokonuje się resyntezy dźwięku.
- Wadą tej metody jest konieczność znajomości częstotliwości podstawowej  $f_1$ .
- Nie można też za jej pomocą wygenerować dźwięków zawierających składowe nieharmoniczne.
- Odmiana MQ syntezy addytywnej wykorzystuje transformatę Fouriera do wyznaczenia ciągu widm chwilowych.
- Na ich podstawie dokonuje się wyszukiwania ekstremów lokalnych.
- Częstotliwości odpowiadające tym ekstremom nie muszą być harmonicznymi.
- Uzyskane w wyniku tej analizy dane stanowią podstawę procesu resyntezy.

# Widmowe metody generowania dźwięku

Niezależnie od zaimplementowanej odmiany metody addytywnej:

- Syntezator składa się z zestawu generatorów sygnałów sinusoidalnych
- Syntezowany dźwięk jest wynikiem sumowania przebiegów składowych.



# Widmowe metody generowania dźwięku

## Zalety:

- Duża wierność uzyskiwanych dźwięków syntetycznych w porównaniu z ich oryginalnymi odpowiednikami oraz bogate możliwości modyfikacji uzyskiwanego brzmienia.
- Poprzez odpowiednią modyfikację parametrów uzyskanych w procesie analizy można w prosty sposób uzyskać takie efekty jak:
  - kompresję lub ekspansję czasową bez zmiany wysokości dźwięku,
  - zmianę wysokości dźwięku bez zmiany czasu jego trwania,
  - redukcję szumu.

## Wady:

- Wadą metody jest duża złożoność obliczeniowa wynikająca z konieczności stosowania dużej liczby generatorów.
- Na przykład, aby skonstruować syntezyzator 32-głosowy, w którym każdy dźwięk składałby się z 64 składowych, należy zastosować

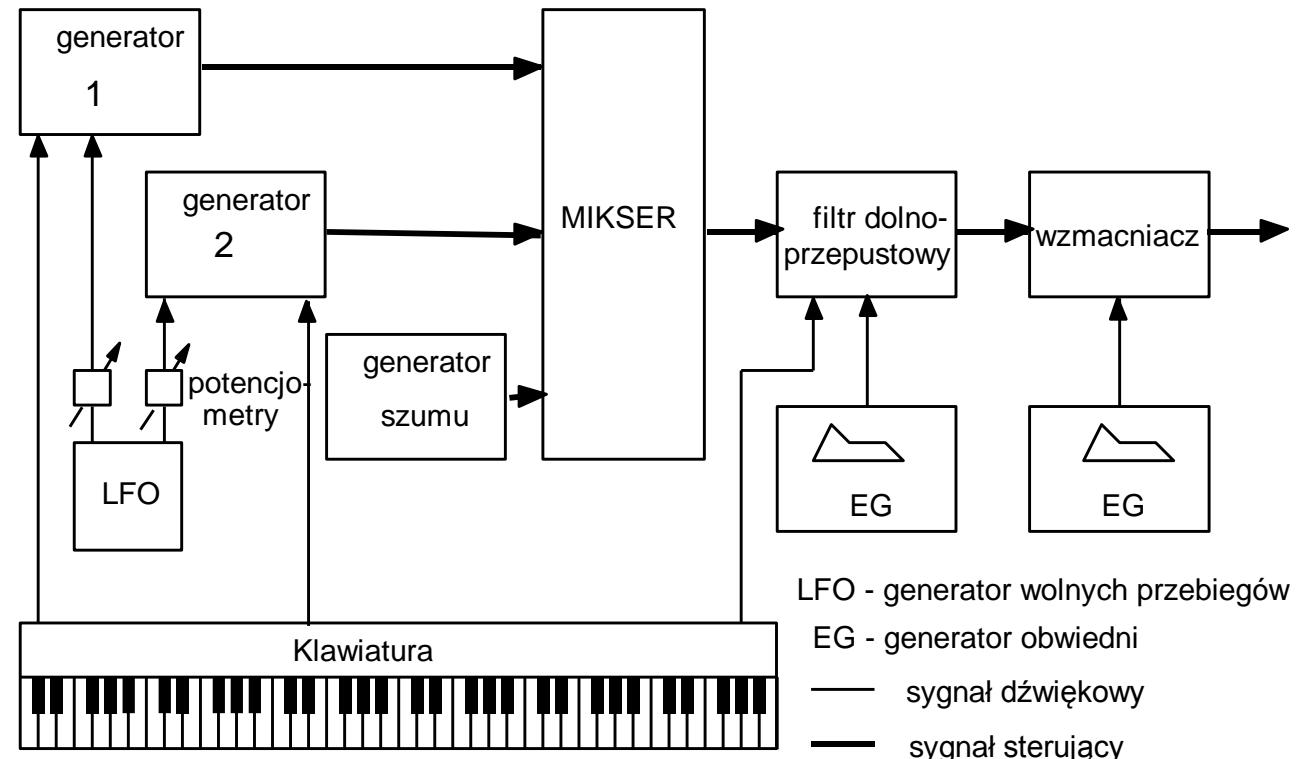
$$32 * 64 = 2048$$

generatorów przebiegów sinusoidalnych.



# Widmowe metody generowania dźwięku

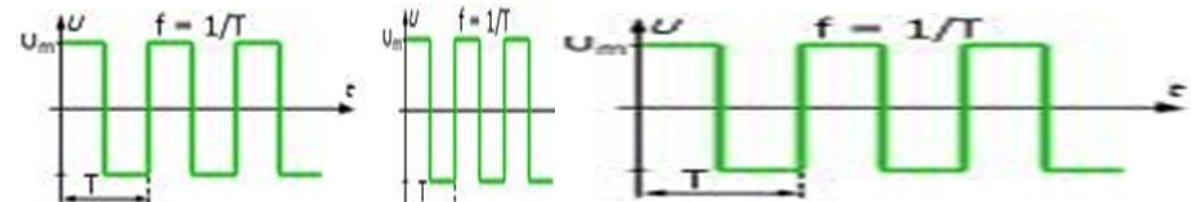
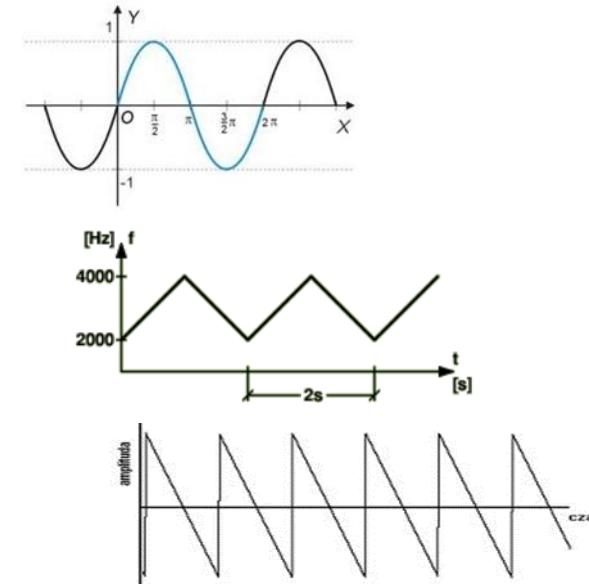
- Metoda subtratywna jest powszechnie stosowana zarówno w klasycznych syntezatorach analogowych, ale również w syntezatorach cyfrowych.
- Polega ona na odejmowaniu określonych składowych widma z szumu lub sygnału szerokopasmowego, w celu uzyskania pożądanej barwy dźwięku.
- Dokonuje się tego w układzie filtracyjnym, do którego kierowany jest sygnał o bogatym widmie.



# Widmowe metody generowania dźwięku

Większość syntezatorów subtraktywnych składa się z:

- generatora szumu (na ogólny biały),
- dwóch generatorów, które pozwalają na wytworzenie co najmniej czterech podstawowych typów przebiegów:
  1. Sinusoidalne
  2. Trójkątne
  3. Piłokształtne
  4. Prostokątne (z możliwością regulacji wypełnienia)



# Widmowe metody generowania dźwięku

- Sygnały z generatorów mieszane są w mikserze w odpowiednich proporcjach.
- W efekcie, na wyjściu miksera pojawia się złożony sygnał dźwiękowy.
- Sygnał ten kierowany jest na wejście filtra, np. dolnoprzepustowego, którego zadaniem jest wycięcie składowych nie występujących w widmie syntezowanego dźwięku.
- Ostatnim członem modyfikującym dźwięk jest wzmacniacz sterowany generatorem obwiedni (EG), pozwalający na nadanie sygnałowi odpowiedniej obwiedni amplitudowej.
- Wadą metody substraktywnej jest ograniczona wierność uzyskiwanych dźwięków syntetycznych w porównaniu z ich odpowiednikami naturalnymi.
- Mimo to, instrumenty działające na podstawie metody substraktywnej są powszechnie wykorzystywane przez muzyków, ponieważ pozwalają na uzyskanie interesujących brzmień, nie tylko tych występujących w naturze.

# Abstrakcyjne algorytmy generowania dźwięku

- Algorytmy abstrakcyjne są nazwą grupy metod syntezy, których wspólną cechą jest to , że - w odróżnieniu od wszystkich pozostałych – nie nawiązują bezpośrednio do:
  - Sposobu powstawania dźwięków w rzeczywistych instrumentach muzycznych,
  - Opisu widm dźwięków, których źródłem są rzeczywiste instrumenty muzyczne
- Do tej grupy metod należą:
  - Modulacja częstotliwości FM
  - Metoda przekształcania fali
  - Metody związane z wykorzystaniem chaosu deterministycznego
- Główną zaletą metod należących do tej grupy jest niska złożoność obliczeniowa.
- Zasadniczą ich wadą jest niewielkie podobieństwo brzmień dźwięków syntetycznych do tradycyjnych brzmień instrumentów akustycznych

# Abstrakcyjne algorytmy generowania dźwięku

- Synteza z wykorzystaniem modulacji FM jest najbardziej znaną i rozpowszechnioną metodą z tej grupy algorytmów.
- U jej podstaw leży ta sama technika, którą stosuje się do generowania radiofonicznych sygnałów FM.
- Polega ona miksowaniu uchwytnych dla ucha ludzkiego sygnałów dźwiękowych małej częstotliwości, składających się na mowę czy muzykę, z prostym sygnałem prądowym wysokiej częstotliwości.
- W efekcie powstaje sygnał prądowy zmodulowany częstotliwościowo.
- Przy spełnieniu określonych założeń dotyczących częstotliwości sygnałów, można w ten sposób uzyskać sygnał złożony o widmie harmonicznym.

# Abstrakcyjne metody generowania dźwięku

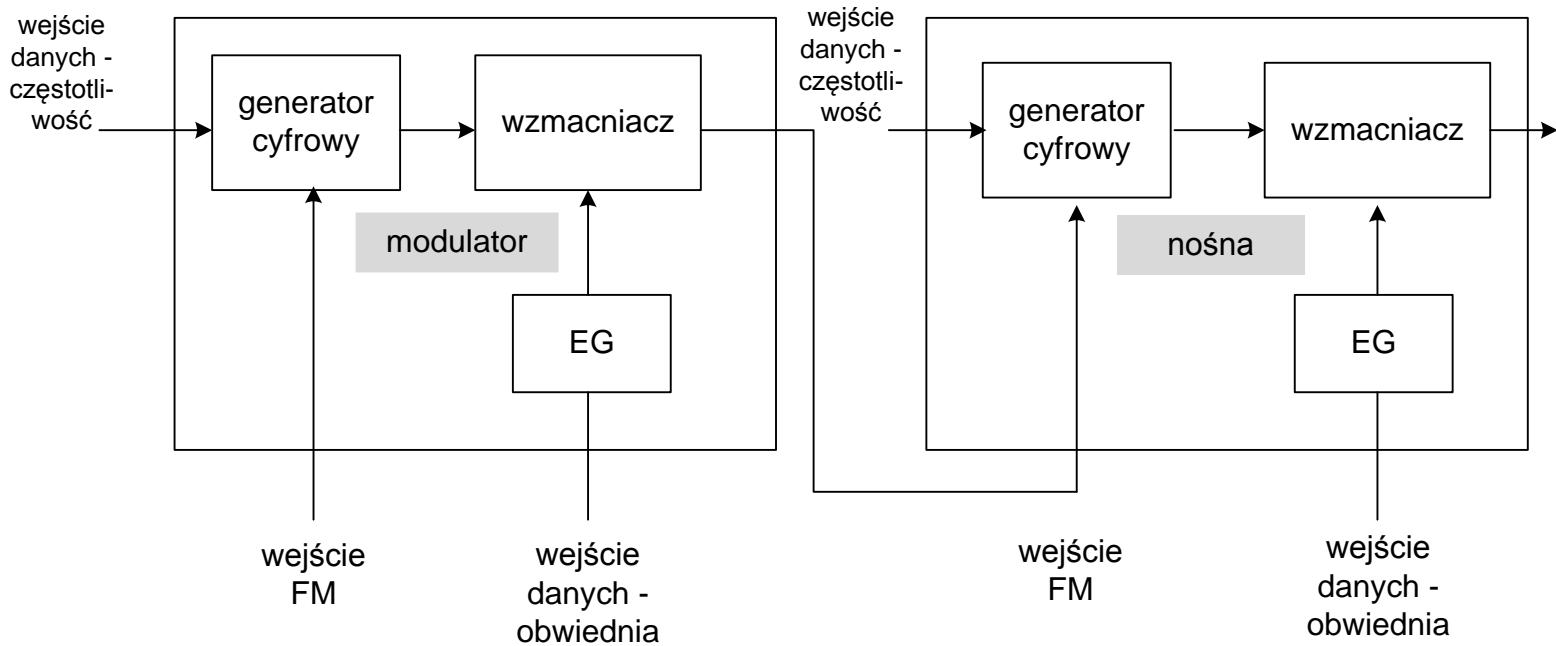
- W pierwszym etapie syntezy FM generowana jest fala nośna o częstotliwości  $f_c$ .
- Nośna jest modulowana przez inną falę (modulator) o częstotliwości  $f_m$ .
- Jeśli amplituda modulatora równa jest 0, wtedy częstotliwość nośnej wynosi  $f_c$ .
- Dodatnia wartość amplitudy modulatora powoduje wahanie częstotliwości sygnału modulowanego.
- Zjawisko to nosi nazwę odchyłki częstotliwości (ang. *frequency deviation*).
- Im "głośniejsza" jest fala modulacyjna, tym większe zmiany w fali modulowanej.
- Na przykład dla częstotliwości nośnej  $f_c = 1000 \text{ Hz}$ , jeśli modulator powoduje wahanie częstotliwości nośnej w zakresie  $900 \text{ Hz} - 1100 \text{ Hz}$ , to odchyłka  $d$  wynosi  $100 \text{ Hz}$ .
- Stosunek odchyłki do częstotliwości modulacji nazywa się indeksem modulacji i oznacza się:

$$I = \frac{d}{f_m}$$

- Barwa dźwięku syntezowanego metodą FM zależy od stosunku częstotliwości modulującej do częstotliwości nośnej oraz od indeksu modulacji

# Abstrakcyjne algorytmy generowania dźwięku

- Najważniejszym elementem syntezy FM jest operator (programowy lub sprzętowy), na który składają się: generator fal sinusoidalnych, wzmacniacz i generator obwiedni.



# Abstrakcyjne algorytmy generowania dźwięku

- Każdy z operatorów może działać jako generator częstotliwości modulującej albo jako generator nośnej, w zależności od konfiguracji i przyjętego sposobu połączenia operatorów.
- Jeżeli operator jest oznaczony jako operator nośnej, to jego sygnał wyjściowy jest bezpośrednio kierowany do konwertera C/A.
- W tym przypadku zmiana poziomu sygnału wyjściowego spowoduje zmianę głośności dźwięku.
- Jeśli dowolnie wybrany operator oznaczony jest jako operator sygnału modulującego, to jego wyjście łączone jest z wejściem danych modulujących innego operatora, wcześniej określonego jako operator nośnej.
- W tej sytuacji modulator powoduje swoim sygnałem wejściowym modulację nośnej.
- W każdym z operatorów sygnał z generatora kierowany jest o wzmacniacza, gdzie reguluje się jego amplitudę.

# Abstrakcyjne metody generowania dźwięku

- Połączenie operatorów tworzy tzw. algorytm syntezy FM.
- Najprostszy algorytm stanowią dwa operatory – taka konfiguracja nie pozwala na uzyskanie ciekawych brzmień
- W praktyce stosuje się więcej operatorów (często 6), które można łączyć na wiele sposobów w algorytmy.
- Te same operatory z tymi samymi ustawieniami, ale połączone w różny sposób, generują całkowicie inne brzmienia.
- Syntezatory FM zastosowano w kartach dźwiękowych dla komputerów PC (Sound Blaster 2/PRO/16).
- Synteza realizowana była przez układ OPL3 firmy Yamaha.
- Układ oferował bardzo uproszczoną syntezę: 2 algorytmy 2-op, 4 algorytmy 4-op.



# Abstrakcyjne metody generowania dźwięku

- Zaletą metody FM jest jej niska złożoność obliczeniowa.
- Wadą tej metody jest brak analitycznych metod pozwalających na dobór parametrów syntezy celem uzyskania pożądanego widma.
- Do rozwiązania tego problemu stosuje się bardziej zaawansowane algorytmy, w tym algorytmy genetyczne.

# Synteza dźwięku - modelowanie fizyczne

- Żadna z dotychczas prezentowanych metod generowania dźwięku nie pozwala muzykom na stosowanie artykulacji muzycznej w pełnym zakresie.
- Aby rozwiązać ten problem całkowicie zmieniono podejście do generowania dźwięku.
- Najnowsze, eksperymentalne syntezatory dźwięku działają w oparciu o modele fizyczne instrumentów akustycznych, symulując zjawiska fizyczne w nich zachodzące przy jednoczesnym uwzględnieniu modelowania procesów artykulacyjnych.
- Do tej grupy metod należą:
  - Metoda komórkowa
  - Modelowanie matematyczne
  - Modelowanie falowodowe

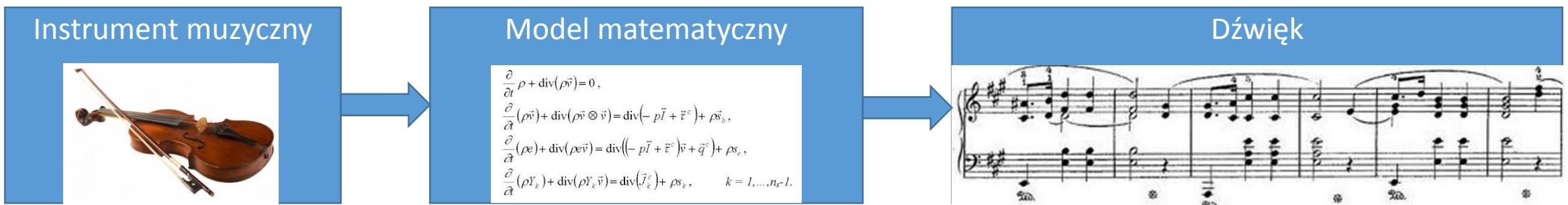


# Synteza dźwięku - modelowanie fizyczne

- Metoda komórkowa polega na tym, że drgający obiekt, będący źródłem dźwięku (np. strunę) modeluje się w postaci cząstek masy, tzw. komórek.
- Każda z komórek powiązana jest z sąsiednimi komórkami za pomocą wiązań sprężystych.
- Pobudzenie do drgań jednej z komórek wywołuje w zamodelowanym ośrodku falę, która w efekcie staje się źródłem dźwięku syntetycznego.
- Zaletą metody jest możliwość zamodelowania dowolnego instrumentu zbudowanego z dowolnie złożonego materiału.
- Wadą jest wysoka złożoność obliczeniowa, rosnąca w miarę rozbudowywania struktur i połączeń między nimi.
- Implementacja metody wymaga szybkich procesorów sygnałowych.

# Synteza dźwięku - modelowanie fizyczne

- Zjawiska fizyczne zachodzące w instrumentach muzycznych modeluje się również za pomocą równania falowego opisującego drgania w rozpatrywanym ośrodku.
- Rozwiązaniem tego równania jest funkcja, która stanowi przebieg czasowy dźwięku syntetycznego.
- Ze względu na wykorzystanie rozbudowanego aparatu matematycznego, metoda ta nazywa się modelowaniem matematycznym.



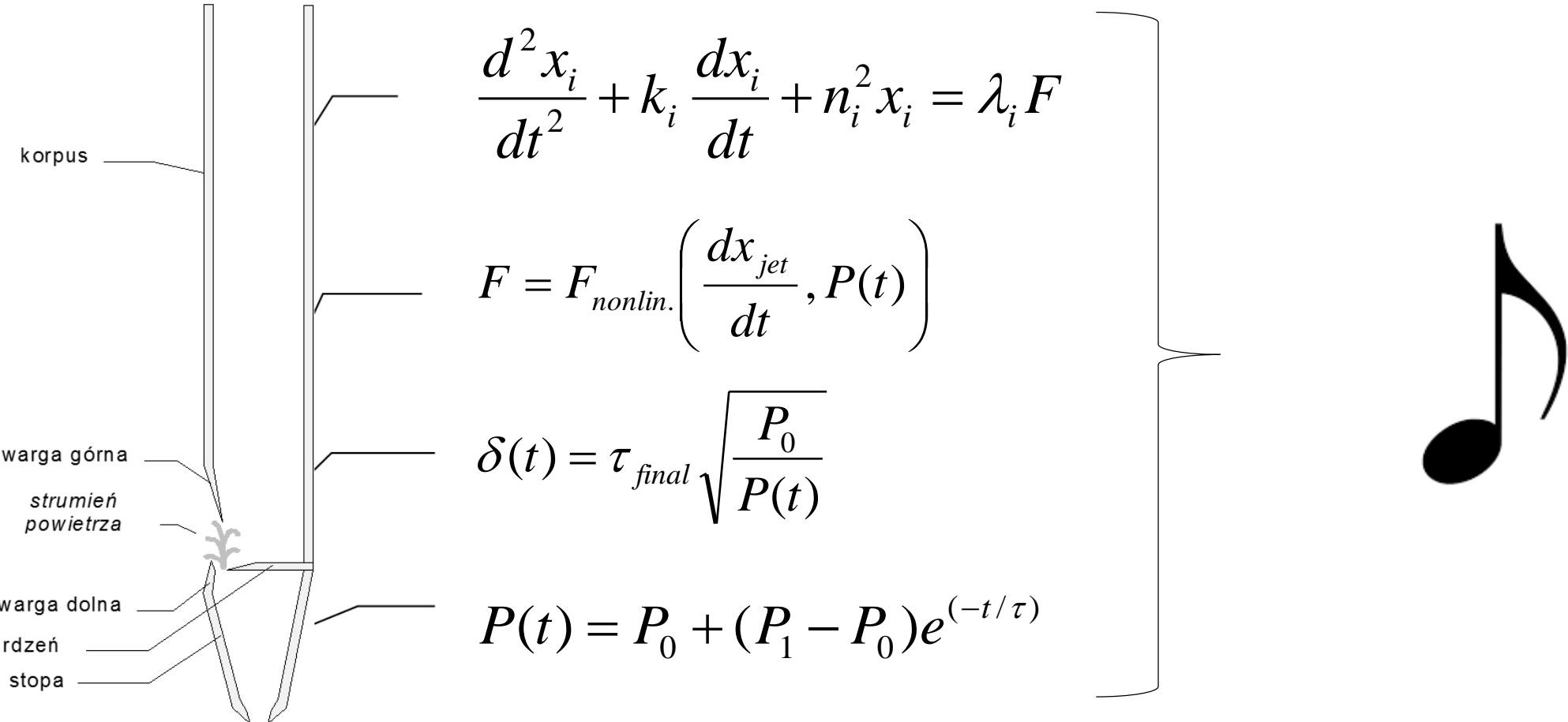
# Synteza dźwięku - modelowanie fizyczne

Metodologicznie rzecz ujmując, zagadnienie modelowania matematycznego można podzielić na cztery zasadnicze etapy.

1. W pierwszym etapie formułuje się opis, który odzwierciedla proces generowania dźwięku w rzeczywistym instrumencie muzycznym. Np. dla zamodelowania dźwięku instrumentu dętego wykorzystuje się układ równań różniczkowych drugiego stopnia.
2. Drugim etapem jest wyznaczenie wartości parametrów występujących w równaniach, np. wartości ciśnienia pobudzającego, początkowe wartości przemieszczenia strumienia powietrza itp. Wartości te mogą być wyznaczone na podstawie pomiarów bezpośrednich lub pośrednich, jeśli są to wielkości możliwe do zmierzenia.
3. W etapie trzecim przeprowadza się symulacje numeryczne. Wiąże się z nim szereg trudności wynikających m.in. z doboru odpowiedniej metody rozwiązywania równań różniczkowych.
4. Ostatnim etapem modelowania matematycznego jest zbadanie wrażliwości modelu oraz wybranych charakterystyk przebiegów wyjściowych na zmiany poszczególnych parametrów. Pozwala to na optymalizację zastosowanego modelu oraz sprawdzenie jego wiarygodności.



# Synteza dźwięku - modelowanie fizyczne



# Synteza dźwięku - modelowanie fizyczne

## Zalety:

- Możliwość rzetelnego sprawdzenia wiarygodności modeli opisujących rzeczywiste instrumenty (poprzez porównanie teoretycznego widma dźwięku, obliczonego na podstawie badanego modelu, z wartościami uzyskanymi na podstawie analiz dźwięków oryginalnych).
- Wgląd w złożoność procesów fizycznych zachodzących podczas generacji drgań oraz pozwala na sprawdzenie poprawności matematycznego opisu rzeczywistych instrumentów.
- Możliwość efektywnego symulowania artykulacji.

## Wady:

- Konieczność stosowania złożonych algorytmów numerycznych
- Duża złożoność obliczeniowa.
- Mała wierność brzmień.



Fundusze  
Europejskie  
Wiedza Edukacja Rozwój



Politechnika Wrocławskiego



Unia Europejska  
Europejski Fundusz Społeczny

# Synteza dźwięku - modelowanie falowodowe

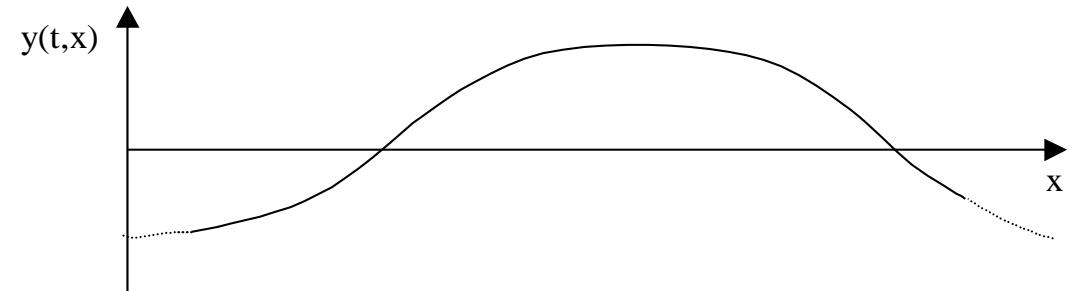
- Istotą metody wykorzystanie falowodu cyfrowego do modelowania propagacji fal bieżących składających się na falę stojącą w danym instrumencie.
- W ciągu ostatnich lat opracowano szereg modeli falowodowych dla różnych instrumentów, jak gitara, klarnet, fortepian i flet.
- Metodę zastosowano w instrumentach elektronicznych i w komputerowych kartach dźwiękowych.
- Eksperymentalnie zastosowano ją (z powodzeniem) do syntezy mowy oraz śpiewu.



# Synteza dźwięku - modelowanie falowodowe

- Ogólny zarys metody zostanie przedstawiony na przykładzie modelowania instrumentów strunowych.
- Rozważmy nieskończoność długą, idealną (idealnie giętką, sprężystą i bezstratną) strunę.
- Równanie falowe opisujące przemieszczenie  $y$  tego typu struny ma następującą postać:

$$K \frac{\partial^2 y}{\partial x^2} = \varepsilon \frac{\partial^2 y}{\partial t^2}$$



gdzie:

$y$  – wychylenie struny,

$K$  – naprężenie struny,

$\varepsilon$  - liniowa gęstość masy,

$x$  – współrzędna leżąca na osi równoległej do struny.



# Synteza dźwięku - modelowanie falowodowe

- Rozwiązanie ogólne tego równania można przedstawić jako superpozycję fal bieżących opisanych równaniem:

$$y(t, x) = y_p\left(t - \frac{x}{c}\right) + y_l\left(t + \frac{x}{c}\right)$$

- Funkcje  $y_p$  i  $y_l$  reprezentują fale bieżące przemieszczające się odpowiednio w lewo (w kierunku ujemnych wartości na osi x) oraz w prawo (w kierunku dodatnich wartości) z prędkością propagacji  $c$ .
- Przekształcając powyższe równanie i zastępując zmienne ciągłe zmiennymi dyskretnymi, uzyskujemy równanie:

$$y(t_n, x_m) = y^+(n-m) + y^-(n+m)$$

gdzie:  $y^+(n) = y_p(nT)$   $y^-(n) = y_l(nT)$   $T$  - okres próbkowania,

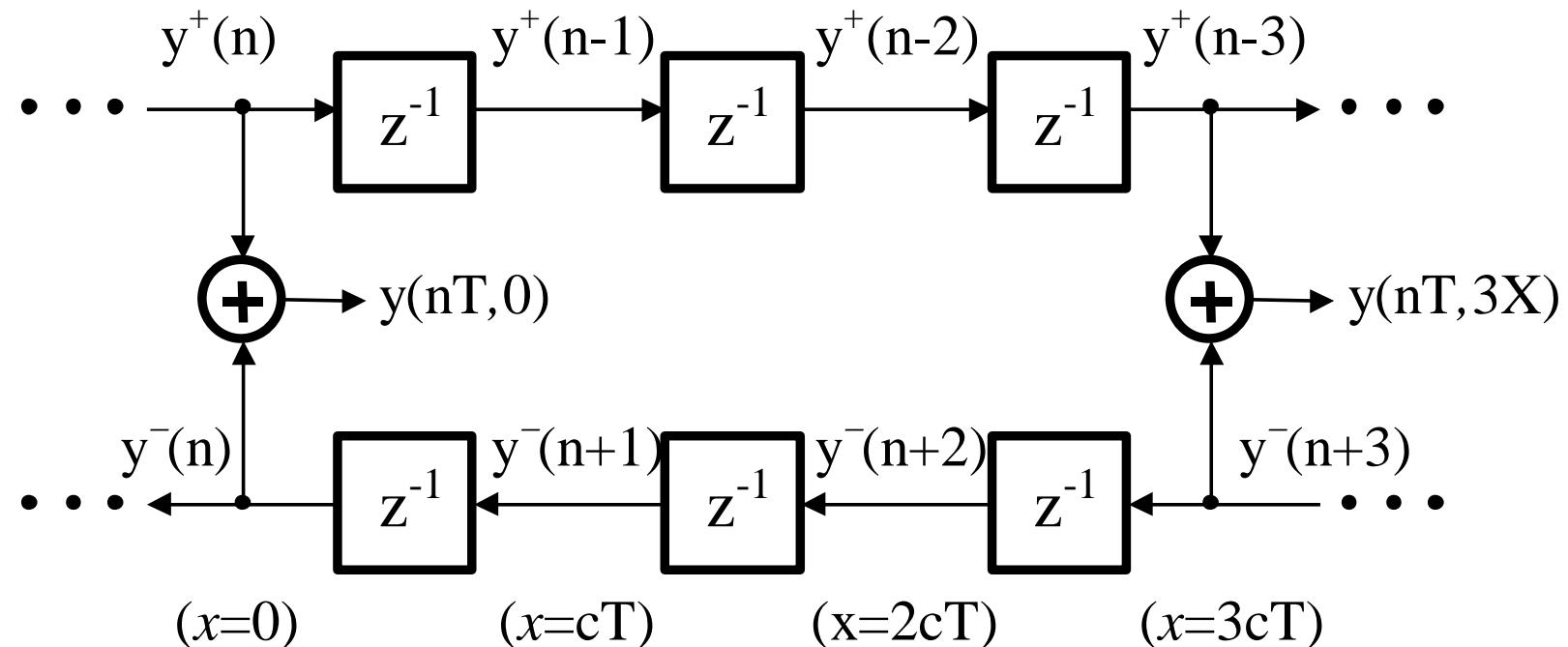


# Synteza dźwięku - modelowanie falowodowe

- Wyrażenia  $y^+(n-m)$  oraz  $y^-(n+m)$  mogą być interpretowane jako bieżąca wartość próbki na wyjściu linii opóźniającej o  $m$  próbek, na wejście której podano sygnał  $y^+(n)$  lub  $y^-(n)$ .
- Zgodnie z powyższym fale bieżące mogą być symulowane przy pomocy pary cyfrowych linii opóźniających zwanych falowodem cyfrowym.
- Wypadkowa wartość przemieszczenia struny w dowolnym punkcie  $x$  falowodu cyfrowego obliczana jest przez zsumowanie odpowiadających mu wartości poszczególnych fal bieżących.
- Zatem falowód cyfrowy można wykorzystać do symulacji propagacji fal bieżących przemieszczenia akustycznego w strunie.

# Synteza dźwięku - modelowanie falowodowe

Przykład cyfrowej symulacji bezstratnego falowodu z przykładowymi wyjściami w punktach  $x=0$  oraz  $x=3X=3cT$ . Symbol  $z^{-1}$  reprezentuje opóźnienie jednostkowe w dziedzinie  $z$ .



# Synteza dźwięku - modelowanie falowodowe

- Przedstawiona idea metody jest bardzo dużym uproszczeniem w stosunku do rzeczywistych modeli fizycznych, które uwzględniają dodatkowo takie czynniki, jak:
  - straty energii (wywołane tarciem wewnętrznym i oporem powietrza),
  - dyspersja związana ze sztywnością struny (a objawiająca się wzrostem prędkości propagacji fal wzdłuż struny towarzyszącym wzrostowi częstotliwości),
  - rzeczywiste rozmiary struny (w poprzednich rozważaniach przyjęliśmy, że struna jest nieskończonie dłuża),
  - warunki początkowe (określone przez trzy podstawowe możliwości pobudzenia struny do drgań: szarpnięcie, uderzenie młoteczkiem i pobudzenie smyczkiem).
- Powyższe czynniki wpływają na postać równania falowego i jego rozwiązanie.
- Przedstawiony model falowodowy dotyczy pojedynczej drgającej struny.
- W przypadku syntezy dowolnego instrumentu strunowego należy dodatkowo uwzględnić model pudła rezonansowego oraz drgania struny w różnych płaszczyznach.

# Techniki przetwarzania mediów cyfrowych

Wykład 5: Cyfrowy Interfejs dla Instrumentów Muzycznych  
MIDI (4 h)



Fundusze  
Europejskie  
Wiedza Edukacja Rozwój



Politechnika Wrocławska



Unia Europejska  
Europeani Fundusz Społeczny

# Wprowadzenie

- Standard MIDI (Musical Instruments Digital Interface) powstał w roku 1982 w wyniku prac przemysłowego konsorcjum MIDI Manufacturers Association.
- Standard definiuje szereg aspektów dotyczących:
  - Sterowania procesem tworzenia muzyki syntetycznej
  - Komunikacji między urządzeniami systemu MIDI
  - Fizycznego łączenia instrumentów elektronicznych
- W zakresie łączenia instrumentów oraz innych urządzeń kompatybilnych z MIDI standard definiuje interfejs fizyczny: gniazda, thru-boxy, port szeregowy specjalnego przeznaczenia – zapewniający transmisję szeregową asynchroniczną itp.
- Dzięki temu można łączyć różne urządzenia MIDI w jedną sieć



# Wprowadzenie

- Najbardziej złożoną częścią systemu MIDI jest protokół komunikacyjny określający sposób przesyłania danych.
- Protokół MIDI składa się z dwóch elementów:
  - cyfrowego protokołu prezentacji utworu,
  - sieciowego protokołu transmisji danych muzycznych między urządzeniami kompatybilnymi z MIDI.
- Urządzenia MIDI komunikują się między sobą za pomocą komunikatów.
- Standard precyzyjnie definiuje format każdego z nich.
- Każdy komunikat w systemie MIDI składa się z jednego bajtu sterującego i odpowiedniej dla tego komunikatu liczby bajtów danych.

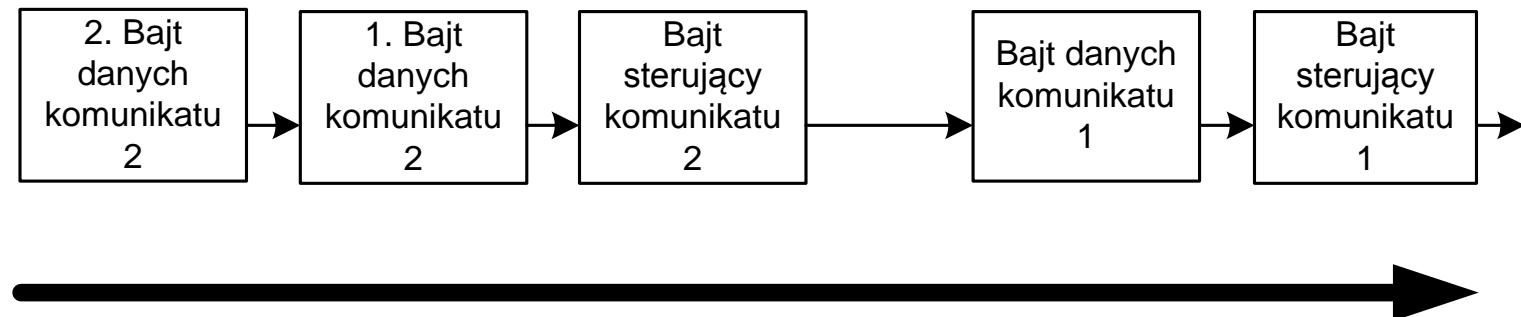


# Komunikacja w systemie MIDI

- Komunikaty dzieli się na dwie kategorie:
  - komunikaty kanałowe odnoszące się do konkretnego kanału MIDI
  - komunikaty systemowe dotyczące wszystkich kanałów
- Kanał jest podstawowym pojęciem dla systemu MIDI.
- W uproszczeniu można powiedzieć, że kanał jest ścisłe określonym miejscem w orkiestrze (kanałem głosowym), w którym skupione są instrumenty jednego rodzaju, np. skrzypce, altówki, oboje, wiolonczele itd.
- Muzycy numerują kanały od 1 do 16 (są to tzw. numery logiczne).
- W systemie MIDI stosuje się numerację fizyczną, a kanały numeruje się od 0 do 15 (jedynka logiczna, to fizyczne zero).

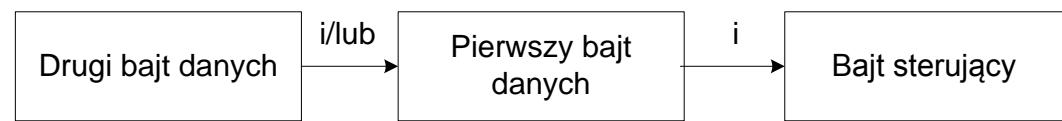
# Komunikaty kanałowe

- Rodzaj komunikatu (kanałowy lub systemowy) rozpoznaje się na podstawie bajtu sterującego.
- Dla komunikatów kanałowych w bajcie sterującym zakodowany jest również numer kanału.

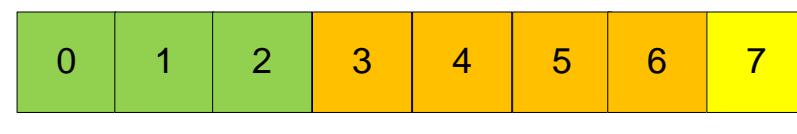


# Komunikaty kanałowe

- Komunikaty kanałowe dzielą się na dwie kategorie:
    - komunikaty głosowe
    - komunikaty trybu pracy
  - Komunikaty głosowe sterują pracą syntezatorów i odpowiadają za wytwarzanie muzyki
  - Komunikaty trybu pracy ustalają, w jaki sposób mają być interpretowane komunikaty głosowe.
  - Z punktu widzenia funkcji przez nie pełnionych, komunikaty trybu pracy różnią się od komunikatów głosowych, chociaż ze względu na składnię podlegają pod tę samą grupę komunikatów kanałowych służących do ustawiania kontrolerów instrumentu.



## Struktura bajtu sterującego w komunikatach kanałowych



### Numer kanału

## Kod komunikatu kanałowego

Dla wszystkich komunikatów sterujących ma wartość „1”

# Komunikaty kanałowe MIDI

Lp.	Bajt sterujący	Znaczenie	Liczba bajtów danych
1.	128 - 143	Note off - Wyłącz nutę	2
2.	144 - 159	Note on - Włącz nutę	2
3.	160 - 175	Polyphonic Key Pressure / After Touch - Docisk klawiszowy	2
4.	176 - 191	Control Change, Parameter - Ustawienie kontrolera	2
5.	192 - 207	Program Change - Wybór barwy	1
6.	208 - 223	Channel Pressure / After Touch - Docisk Kanałowy	1
7.	224 - 239	Pitch Wheel, Pitch Bender - Płynna zmiana wysokości dźwięku	2



# Komunikaty kanałowe

- Komunikaty Note off i Note on odpowiadają za „odegranie” nuty.
- W obu komunikatach pierwszy bajt danych, tzw. numer nuty (note number, pitch), określa wysokość dźwięku i może przyjmować wartości od 0 do 127.
- Kolejne wartości odpowiadają kolejnym dźwiękom w skali chromatycznej. Szczegółowe odwzorowanie przedstawia specyfikacja General MIDI. Np. C - wielkie ma numer 36, a c1 (razkreślne) - 60.
- Ponieważ w muzyce nie używa się dźwięków niższych niż C2 (C - subkontra - ok. 16 Hz), niektóre profesjonalne instrumenty mają pod numerami dźwięków 0 - 11 przyporządkowane dodatkowe brzmienia, często efekty specjalne.
- Drugi bajt danych - szybkość (velocity) w przypadku komunikatu Note on określa głośność brzmienia dźwięku. Nazwa szybkość bierze się stąd, iż najczęściej głośność brzmienia klawiatury odpowiedni układ elektroniczny określa na podstawie pomiaru odstępu czasu pomiędzy rozpoczęciem naciśnięcia klawisza a zakończeniem. Domyślnie przyjmuje się w tym miejscu głośność 64.
- Komunikat Polyphonic Key Pressure służy do zmiany parametrów dźwięku, który już brzmi (głośności, vibracji, wysokości). Podobne efekty osiąga się na klawiaturze poprzez zmianę siły nacisku klawisza.
  - Pierwszy bajt danych tego komunikatu określa numer dźwięku,
  - drugi zaś - głośność.

Komunikat ten powienien być wysłany po Note on i przed Note off.



# Komunikaty kanałowe

- Komunikat Program Change służy do zmiany barwy dźwięku. Nazwa komunikatu - Program Change - bierze się stąd, że zaprogramowane w syntezatorze składy barw nazywa się programami. W praktyce barwa dźwięku oznacza charakterystyczne brzmienie tradycyjnego instrumentu.
- Komunikat ma 1 bajt danych o wartościach od 0 do 127 i służy do wyboru instrumentu. Poszczególnym numerom przypisano na brzmienia instrumentów (wg specyfikacji General MIDI), np. 17 - organy Hammonda, 14 - Ksylafon, 26 - gitara akustyczna o strunach stalowych, 126 - helikopter (dźwięki specjalne).
- Komunikat Channel Pressure podobnie jak komunikat Polyphonic Key Pressure służy do zmiany parametrów dźwięku, z tym, że dotyczy wszystkich dźwięków odtwarzanych na danym kanale.
- Komunikat Pitch Bender umożliwia płynną zmianę wysokości dźwięku. Ponieważ komunikat ten ma dwa bajty danych, to zmiana głośności może odbywać się z rozdzielcością 14 bitową. W praktyce realizuje się to w ten sposób, że regulacji dokonuje się za pomocą pokrętła lub joysticka. Instrument zgodny ze standardem MIDI powinien zapewniać regulację co najmniej w granicach dwóch półtonów.



# Komunikaty kanałowe

- Komunikat Control Change ma dwojakie zastosowanie:
  - po pierwsze służy do ustawiania kontrolerów instrumentu,
  - po drugie do wyboru trybu pracy instrumentu.
- Kontroler jest „pokrętłem”, z pomocą którego można regulować dodatkowe efekty w trakcie brzmienia dźwięku, np. pedał pogłosu, głębokość tremolo lub vibrato, regulacja wszystkich faz trwania dźwięku, narastania, ustalenia oraz wybrzmiewania itp.
- Znaczenie komunikatu rozpoznaje się po wartościach ustawionych na dwóch bajtach danych.
- Jeśli pierwszy bajt danych zawiera wartości z przedziału od 0 do 101, oznacza to, że komunikat służy do ustawiania poziomu kontrolera.
- Komunikaty o pierwszym bajcie danych z przedziału od 120 do 127 służą do ustawiania trybu pracy instrumentu.



# Komunikaty kanałowe

Lp.	Pierwszy bajt danych	Znaczenie	Drugi bajt danych
1.	120	All Sound Off - Wyłączenie wszystkich dźwięków; komunikat powoduje całkowite wyciszenie instrumentu.	
2.	121	Reset All Controllers - Wyzerowanie wszystkich kontrolerów; komunikat powoduje ustawienie kontrolerów na ich wartości standardowe, zwykle na 0.	
3.	122	Local Control - Kontrola klawiatury lokalnej. W tym przypadku drugi bajt danych może mieć wartość 0 - klawiatura wyłączona lub 127 - klawiatura włączona.	
4.	123	All Notes Off - Wyłączenie wszystkich nut; komunikat powinien być używany jedynie w wyjątkowych sytuacjach, np. nagłe przerwanie odtwarzania.	
5.	124	Omni Mode Off - Wyłączenie trybu Omni.	Wyzerowany
6.	125	Omni Mode On - Włączenie trybu Omni.	Wyzerowany
7.	126	Monophonic Mode - Włączenie trybu monofonicznego.	Liczba kanałów monofonicznych
8.	127	Polyphonic Mode - Włączenie trybu polifonicznego.	



# Komunikaty kanałowe

- Tryb Omni informuje, czy komunikaty głosowe kierowane do konkretnego kanału (podstawowego) mają być odbierane wyłącznie na jednym kanale (Off) czy również na innych kanałach (On).
- Tryb Mono/Poly określa polifonię instrumentu:
  - Poly oznacza, iż dźwięki będą odtwarzane polifonicznie, tzn. na każdym kanale możliwe jest współbrzmienie dźwięków o różnych wysokościach.
  - Mono znaczy, że na każdym kanale możliwe jest brzmienie tylko jednego dźwięku.
- Właściwy tryb pracy instrumentu określa kombinacja Omni oraz Mono/Poly
- Na przykład tryb Omni OFF, Poly oznacza, że komunikaty głosowe odtwarzane są wyłącznie na kanale podstawowym, określonym w bajcie sterującym komunikatu kanałowego.
  - W tym trybie każdy kanał sterowany jest niezależnie.
  - Komunikat kierowany do konkretnego kanału jest dla pozostałych niewidoczny.

# Komunikaty systemowe

- Komunikaty systemowe MIDI dotyczą wszystkich kanałów i odnoszą się do ogólnych ustawień instrumentu.
- Wyróżnia się trzy typy komunikatów systemowych:
  - Komunikaty wspólne dla wszystkich kanałów (Common Messages),
  - Komunikaty „czasu rzeczywistego” (Real Time Messages),
  - Komunikaty typu System Exclusive - przeznaczone dla dowolnego wykorzystania przez producentów sprzętu muzycznego.



# Komunikaty systemowe

## Komunikaty Common Messages:

- **Quarter Frame**
  - Komunikat używany do synchronizacji w trybie SMPTE.
- **Song Position (pozycja melodii)**
  - Komunikat umożliwia zmianę aktualnej pozycji odtwarzanej melodii.
  - Przeznaczony dla automatów grających, sekvenserów i automatów perkusyjnych.
  - Posiada 2 bajty danych i określa pozycję melodii z 14 bitową rozdzielcością
  - Jednostką pozycji jest tzw. uderzenie (ang. beat) odpowiadające 6 komunikatom MIDI Clock. W momencie włączenia odtwarzania, pozycja melodii ustawiana jest na 0 i po czasie odpowiadającym sześciu komunikatom Clock zwiększa się o jeden.
- **Song Select (wybór melodii)**
  - Komunikat służy do wyboru jednej z wcześniej zaprogramowanych melodii.
  - Posiada jeden bajt danych określający numer melodii (może ich być 128).
  - Odnosi się do automatów grających, sekvenserów i automatów perkusyjnych (tj. podobnie jak Song Position).
- **Tune Request (strojenie)**
  - Przeznaczony jest dla syntezatorów analogowych, oznacza żądanie automatycznego dostrojenia oscylatorów
  - Nie posiada bajtów danych.
- **End of Exclusive (EOX)**
  - Koniec komunikatu System Exclusive.
  - Obowiązkowy, ponieważ w komunikatach SysEx liczba bajtów danych może być dowolna.

# Komunikaty systemowe

**Komunikaty Real Time Messages:** Komunikaty „czasu rzeczywistego” przeznaczone są do synchronizacji czasowej urządzeń MIDI. Wszystkie są jednobajtowe i mogą być wysyłane pomiędzy bajtami innego komunikatu. Do grupy tej należą:

- **Clock**
  - Podstawowy komunikat przeznaczony do synchronizacji czasowej instrumentów, w szczególności automatów perkusyjnych.
  - Komunikat ten pełni rolę metronomu, na jedną ćwierćnutek przypadają 24 komunikaty Clock.
  - Wszystkie instrumenty wchodzące w skład sieci MIDI powinny „słuchać” tego komunikatu i dostosować tempo odtwarzania do otrzymywanych komunikatów Clock (służy temu wewnętrzny zegar instrumentu).
- **Start** - powoduje rozpoczęcie odtwarzania od początku melodii.
- **Continue** - powoduje wznowienie odtwarzania po otrzymaniu najbliższego komunikatu Clock.
- **Stop** - oznacza przerwanie odtwarzania z wyzerowaniem licznika czasu.
- **Active Sensing**
  - Komunikat wysyłany dla potwierdzenia, iż połączenie między urządzeniami jest aktywne.
  - Jeżeli nadajnik nie wysyła innych komunikatów powinien wysłać Active Sensing.
  - Dopuszczalny maksymalny czas bez transmisji wynosi 300 ms, tzn. nadajnik ma prawo nie wysyłać niczego przez co najwyżej 300 ms, a odbiornik musi czekać co najmniej 300 ms na komunikat.
  - Jeśli czas ten zostanie przekroczony, odbiornik ma prawo przyjąć, iż nastąpiła przerwa połączenia; najczęściej wyłącza wszystkie aktualnie brzmiące nuty i ustawia parametry instrumentu w pozycji wyjściowej.
- **System Reset**
  - Komunikat ustawia wszystkie parametry instrumentu w pozycji wyjściowej.
  - Powinien być wysyłany wyłącznie na żądanie użytkownika.



# Komunikaty systemowe

## Komunikaty System Exclusive

- W obrębie grupy komunikatów System Exclusive wyróżnia się cztery ich podgrupy:
  - Komunikaty specyficzne dla producenta (**Manufacturer Specific System Exclusive**),
  - Komunikaty uniwersalne, niezależne od producenta (**Universal Noncommercial System Exclusive**),
  - Komunikaty uniwersalne czasu nierzeczywistego (**Universal Non Real Time System Exclusive**),
  - Komunikaty uniwersalne czasu rzeczywistego (**Universal Real Time System Exclusive**).
- Struktura komunikatów System Exclusive: bajt sterujący, a po nim dowolna liczba bajtów danych
- System MIDI rozpoznaje typ komunikatu na podstawie bajtu identyfikującego (pierwszy bajt danych).
- W zależności od operacji, która ma być wykonana, w bajtach danych podaje się dodatkowe parametry, np. numer kanału, do którego adresowany jest komunikat oraz numer urządzenia.
- Dzięki temu komunikat System Exclusive może być adresowany również do konkretnego instrumentu.



# Komunikaty systemowe

- Komunikat Manufacturer Specific System Exclusive przeznaczony jest dla producentów instrumentów muzycznych do obsługi funkcji niedostępnych w innych modelach.
- Każdy producent urządzeń MIDI, jeżeli chce używać tego komunikatu, powinien zarejestrować się w organizacji MIDI Manufacturers Association (MMA), gdzie otrzyma unikalny identyfikator dla „swoich” komunikatów System Exclusive.
- Universal Non Real Time System Exclusive obejmują grupę komunikatów o różnym znaczeniu.
- Należą do niej np. komunikaty: Sample Dump i Sample Dump Extensions przeznaczone są do przesyłania próbek dźwiękowych (plików typu wave w specjalnym formacie MIDI - Sample Dump Standard).
- Chociaż zasadniczo format MIDI nie jest przystosowany do tego typu transmisji, to dzięki tym komunikatom transmisja taką staje się możliwa. Jest ona jednak ok. 40 razy wolniejsza od standardowej transmisji MIDI.
- Do grupy Universal Non Real Time System Exclusive należą również komunikaty MIDI Time Code (tzw. kody czasowe MIDI - MTC) wykorzystywane do synchronizacji.



# Synchronizacja w systemie MIDI

- Synchronizacja w systemie MIDI może odbywać się za pomocą komunikatów: MIDI Clock lub MIDI Time Code.
- Synchronizacja za pomocą MIDI Clock zależy od tempa (24 MIDI Clock na ćwierćnutek).
- Przeznaczona jest do zastosowań ściśle muzycznych. Jeżeli tempo zwalnia się, to komunikaty MIDI Clock wysyłane są wolniej.
- Synchronizacja MIDI Time Code stosowana jest tam, gdzie wymagany jest bezwzględny pomiar czasu.
- Używa się do tego celu kodu czasowego SMPTE.
- Głównym zastosowaniem synchronizacji SMPTE była współpraca z urządzeniami wizualnymi (kino, video, TV).
- Kod czasowy SMPTE zawiera następujące parametry:
  - godzina,
  - minuta,
  - sekunda,
  - klatka (numer klatki obrazu wewnętrz sekundy).



# Synchronizacja w systemie MIDI

- Systemy filmowe różnią się między sobą prędkością przesuwu taśmy.
- Ostatnia liczba (klatka) kodu SMPTE określa przedział czasu, który jest zależny od typu stosowanej taśmy filmowej.
- Uwzględnia się przy tym cztery prędkości przesuwu taśmy:
  - 24 klatki na sekundę - prędkość kinowa w Europie,
  - 25 klatek na sekundę - prędkość telewizyjna w systemach o częstotliwości obrazu 50 Hz (systemy TV CCIR/OIRT),
  - 30 klatek na sekundę (30 frames/second non-drop frame) - prędkość TV w systemach o częstotliwości obrazu 60 Hz.
  - 29,97 klatek na sekundę (30 frames/second drop frame) - zmodyfikowana prędkość TV; aby osiągnąć średnią prędkość 29,97 klatek/sekundę co pewien odstęp czasu opuszcza się jedną klatkę.



# Synchronizacja w systemie MIDI

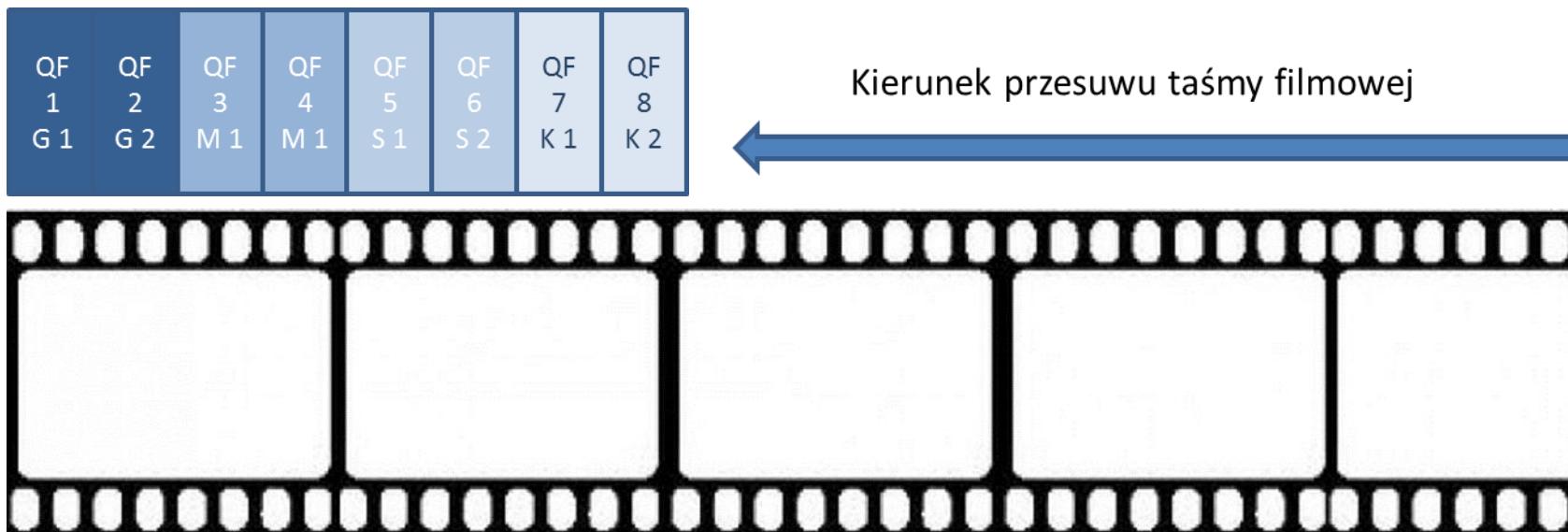
- Do synchronizacji instrumentów oraz urządzeń MIDI zgodnie ze standardem SMPTE używa się czterech rodzajów komunikatów:
- **Quarter Frame**,
- **Full Message**,
- **User Bit Message**,
- **Set-Up Message**.
- **Full Message** (Pełny Kod Czasowy) przeznaczony jest do ustawienia pozycji licznika np. po przewinięciu taśmy. Podaje kod czasowy SMPTE za pomocą jednego komunikatu. Full Message jest komunikatem typu System Exclusive z grupy Universal Real Time.
- **User Bits** wykorzystywany jest w sposób specyficzny dla danego urządzenia (System Exclusive, Universal Real Time). Na przykład stosuje się go do podawania daty lub numeru szpuli w magnetofonie taśmowym.
- **Set-Up** kierowany jest do konkretnej jednostki w systemie MIDI, np. magnetofonu wielośadowego, video, sekwensera. Uaktywnia lub dezaktywuje działanie zdarzeń przypisanych do poszczególnych chwil w momencie nadchodzenia sygnału czasu (komunikatu MTC lub SMPTE), włącza lub wyłącza tryb nagrywania w odpowiednim momencie itp. Należy do grupy komunikatów System Exclusive Non-Real Time Universal.



# Synchronizacja w systemie MIDI

## Komunikat Quarter Frame

- Używany jest w trakcie przesuwu taśmy filmowej.
- Pełna informacja o czasie w kodzie SMPTE składa się z ośmiu komunikatów Quarter Frame i zajmuje czas równy czasowi przewinięcia dwóch klatek.
- Każdy komunikat Quarter Frame nadsyłany jest co 1/4 klatki, stąd jego nazwa.
- Typ komunikatu, zawarty w bajcie danych, informuje, który z ośmiu składników kodu czasowego SMPTE jest aktualnie przesyłany.
- Komunikat należy do grupy komunikatów systemowych Common Messages.



# Struktura pliku MIDI

- Podstawową jednostką organizacji standardowego pliku MIDI jest blok (chunk), na który składają się:
  - identyfikator bloku (4 znaki),
  - liczba bajtów danych (32 bity),
  - dane.
- Plik MIDI zawiera dwa typy bloków: blok nagłówkowy i blok śladowy.
- Blok nagłówkowy zawiera minimum informacji niezbędnej do obsługi całego pliku MIDI
- Blok śladowy zawiera strumień danych MIDI, który z kolei może zawierać informacje ze wszystkich kanałów.
- W odrębnych blokach śladowych składowane są informacje z odrębnych śladów logicznych reprezentowanych w programach lub sekwencjach MIDI.
- Za pomocą różnych bloków w tym samym pliku MIDI można reprezentować różne dane dotyczące wyjść MIDI OUT, różne wzory melodyczne (np. wzorce akompaniamentu), a nawet różne utwory.

# Struktura pliku MIDI

Plik MIDI rozpoczyna się zawsze blokiem nagłówkowym, po którym następuje jeden lub więcej bloków śladowych. Ogólna struktura pliku jest następująca:

**MThd** {identyfikator bloku nagłówkowego}

**<32-bitowa długość bloku nagłówkowego>**

**<dane bloku nagłówkowego>**

**MTrk**

**<32-bitowa długość bloku śladowego>**

**<dane bloku śladowego>**

.....

**MTrk**

**<32-bitowa długość bloku śladowego>**

**<dane bloku śladowego>**



# Struktura pliku MIDI

- W bloku nagłówkowym znajdują się podstawowe informacje na temat danych zawartych w pliku. Kompletna zawartość bloku nagłówkowego ma postać:

<typ bloku> – „MThd”

<długość> – 32-bitowa reprezentacja liczby

(<format> <liczba śladów> <podzielnik>) – dane bloku nagłówkowego zajmują zawsze 6 bajtów (trzy dwubajtowe słowa)
- Część „format” określa całościową reprezentację pliku i może przyjmować jedną z trzech wartości:

0 - plik zawiera jeden wielokanałowy ślad,

1 - plik zawiera jeden lub więcej równoczesnych śladów (lub sekwencji) przeznaczonych dla różnych wyjść MIDI OUT,

2 - plik zawiera jeden lub więcej czasowo niezależnych jednośladowych wzorców.



# Struktura pliku MIDI

- <**liczba śladów**> określa liczbę bloków śladowych w pliku. Dla pliku o formacie „0” liczba śladów musi wynosić 1.
- <**podzielnik**> określa jaką częścią ćwierćnuty są interwały czasowe pomiędzy zdarzeniami. Np. jeżeli podzielnik ma wartość 192, to sekwencja bajtów w pliku postaci:

- |                 |   |
|-----------------|---|
| <b>81 48</b>    | - wartość 192 w postaci zmiennoprzecinkowej,              |
| <b>90 60 64</b> | - włącz nutę c razkreślne na pierwszym kanale logicznym,  |
| <b>81 48</b>    | - wartość 192 w postaci zmiennoprzecinkowej,              |
| <b>80 60 64</b> | - wyłącz nutę c razkreślne na pierwszym kanale logicznym, |

oznacza, iż od wcześniejszego zdarzenia powinien upływać czas jednej ćwierćnuty, następnie powinno wystąpić zdarzenie „włącz nutę”, a dalej, po upływie czasu trwania ćwierćnuty – zdarzenie „wyłącz nutę”.

Gdyby podzielnik miał wartość 384, to ta sama sekwencja bajtów określałaby brzmienie ósemki.

Uogólniając podzielnik odpowiada parametrowi PPQN (Parts Per Quarter Note) – liczba jednostek na ćwierćnuttę i determinuje rozdzielczość czasu notowanego w pliku.



# Struktura pliku MIDI

- Jeżeli podzielnik ma wartość ujemną, wówczas określa jakim ułamkiem sekundy są interwały czasowe zdarzeń.
- W tym przypadku zdarzenia określane są w czasie bezwzględnym, nie jak poprzednio – metrycznym (zależnym od tempa).
- Starszy (występujący jako pierwszy) bajt parametru podzielnik może wówczas przyjmować jedną z czterech wartości: -24, -25, -29, -30 odpowiadających czterem standardom synchronizacji SMPTE. Drugi bajt (dodatni) zawiera podzielnik wewnątrz klatki.
- Typowe wartości to: 4 (tzw. MIDI time resolution), 8, 10, 80 (tzw. Bit resolution) lub 100.
- Taki system pozwala na dokładną specyfikację czasowej podstawy śladów, a także umożliwia ustawienie rozdzielczości „milisekundowej” poprzez specyfikację 25 klatek/s oraz rozdzielcość 40 jednostek na klatkę.



# Struktura pliku MIDI

- Blok śladowy jest miejscem składowania melodii.
- Struktura bloku śladowego:

**MTrk** {nagłówek}  
**<32-bitowa długość bloku śladowego>**  
**<dane bloku śladowego>**

- Zawartość pola dane bloku śladowego składa się ze zdarzeń:

**<dane bloku śladowego>** = **<zdarzenie MTrk>** + **<zdarzenie MTrk>** + ... - dowolna liczba powtórzeń, w tym 0 oznaczające puste dane śladu

**<zdarzenie MTrk>** = **<przedział czasu>** **<zdarzenie>**

- **<przedział czasu>** reprezentuje przedział czasu jaki upłynął od poprzedniego zdarzenia, aby kolejne zdarzenie mogło być obsłużone (p. podzielnik w nagłówku pliku)

**<zdarzenie>** = **<zdarzenie MIDI>** | **<zdarzenie SysEx>** | **<metazdarzenie>**

- Zdarzenia MIDI (dowolne komunikaty kanałowe) zawierają bajty danych przeznaczone do transmisji przez złącze MIDI OUT lub do odbioru poprzez złącze MIDI IN.

# Struktura pliku MIDI

- Zdarzenie SysEx używane jest do specyfikacji komunikatu MIDI z grupy System Exclusive lub dowolnego ciągu danych, który ma być wysłany do sieci.
- Metazdarzenia służą do sterowania pracą sekwencera bądź programu komputerowego.
- Informacja zawarta w metazdarzeniach nie jest przesyłana przez złącze MIDI. Przeznaczona jest dla programu obsługującego złącze lub dla sekwencera, jest też użyteczna dla struktury pliku.
- Wszystkie metazdarzenia posiadają tę samą składnię:  
**FFh <typ> <długość> <dane>**
- Wszystkie metazdarzenia rozpoczynają się bajtem FFh, drugi bajt określa typ metazdarzenia zdefiniowany liczbą z przedziału od 0 do 127, następny bajt preczyzuje długość danych (od 0 do 4 bajtów), właściwe dane zajmujące liczbę bajtów określoną przez długość. Jeśli z danym metazdarzeniem nie są związane dane, to parametr długość zajmuje 1 bajt i przyjmuje wartość 0.

# Wybrane metazdarzenia MIDI

- W pliku MIDI w formie odrębnych metazdarzeń zapisane są informacje dotyczące tempa, metrum i tonacji.

Nazwa	Set tempo – ustawienie tempa
Typ	51
Długość	3
Format	FF 51 03 <tt tt tt>

- **Opis:**

Wielkość <tt tt tt> określa tempo na dwa sposoby:

- w mikrosekundach na ćwierćnutek,
- w liczbie dwudziestych czwartych części mikrosekundy na jeden komunikat MIDI Clock.

Reprezentacja tempa w postaci czas/ćwierćnutek zamiast odwrotnie pozwala na synchronizację długich przedziałów.



# Wybrane metazdarzenia MIDI

Nazwa	Time signature - metrum
Typ	58
Długość	4
Format	FF 58 04 <l> <m> <c> <b>

## Opis:

Metrum reprezentowane jest przez cztery liczby:

- <l> - licznik metrum
- <m> - mianownik metrum; <m> podaje wykładnik wyrażenia potęgowego o podstawie 2 określającego metrum, np. <m> = 2 oznacza, że jednostką rytmiczną jest ćwierćnuta, <m> = 3 – ósemka.
- <c> - określa pracę metronomu podając liczbę komunikatów MIDI Clock przypadających na jedno uderzenie metronomu,
- <b> wyraża liczbę trzydziestodwójkę notowanych podczas trwania ćwierćnuty MIDI (odpowiadającej czasowi trwania 24 komunikatów MIDI Clock). Trzydziestodwójka, to nuta trwająca trzydziestą drugą część całej nuty.



# Wybrane metazdarzenia MIDI

- Parametr **<b>** został dodany, gdyż istnieje wiele programów, które w czasie transmisji uwzględniają ćwierćnutek MIDI. W pliku jednak ten sam czas zapisują jako inną wartość, np. ósemkę, co później ma znaczenie przy interpretacji przez np. programy edycyjne.
- Kompletne metazdarzenie dla metrum 6/8, gdzie metronom powinien uderzać co wartość ćwierćnuty z kropką, przy 24 komunikatach MIDI Clock przypadających na ćwierćnutek (72 na takt) ma postać (szesnastkowo):

**FF 58 04 06 03 24 08**

Metrum 6/8 zakodowane jest w postaci 06 03 ( $8 = 2^3$ ) 32 komunikaty MIDI Clock przypadające na ćwierćnutek z kropką (24h w zapisie szesnastkowym) oraz 8 notowanych trzydziestodwójkę przypadających na jedną ćwierćnutek MIDI.

# Wybrane metazdarzenia MIDI

- Metazdarzenia zawierające informacje odnośnie tempa oraz metrum lokowane są w pliku MIDI w różnych miejscach w zależności od formatu pliku.
- Dla pliku w formacie 0 metazdarzenia opisujące tempo rozproszone są wzdłuż całego śladu.
- Dla pliku o formacie 1 metazdarzenia opisujące tempo muszą znajdować się w pierwszym śladzie.
- W zasadzie każdy plik MIDI powinien specyfikować metrum i tempo. Jeśli tego nie podaje, to przyjmuje się dla metrum wartość 4/4, zaś dla tempa 120 uderzeń na minutę.
- W formacie 0 te metazdarzenia powinny pojawić się przynajmniej na początku wielokanałowego śladu.
- W formacie 1 - powinny być zawarte w pierwszym śladzie.
- W formacie 2 każdy z czasowo niezależnych wzorców powinien zawierać przynajmniej początkową wartość metrum oraz tempa.



# Wybrane metazdarzenia MIDI

Nazwa:	Key signature - tonacja
Typ:	59
Długość:	2
Format:	FF 59 02 <kb> <dm>

## Opis:

Parametry oznaczają odpowiednio: <kb> – liczbę krzyżyków lub liczbę bemoli, <dm> - tonację

- |         |                |
|---------|----------------|
| kb = -7 | 7 bemoli       |
| .....   | .....          |
| kb = -1 | 1 bemol        |
| kb = 0  | C-dur (a-mol)  |
| kb = 1  | 1 krzyżyk      |
| .....   | .....          |
| kb = 7  | 7 krzyżyków    |
| dm = 0  | tonacja durowa |
| dm = 1  | tonacja molowa |



# Wybrane metazdarzenia MIDI

Nazwa:	<b>Copyright Notice – prawa autorskie</b>
Typ:	02
Długość:	Zmienna – długość komunikatu nie jest ustalona
Format:	<b>FF 02 &lt;dlugość&gt; &lt;tekst&gt;</b>

## Opis:

Notka zawierająca informację o prawach autorskich sekwencji. Powinna zawierać znaki ©, rok, właściciela praw autorskich.

Jeżeli kilka fragmentów pochodzących od różnych producentów znajduje się w tym samym pliku MIDI, to wszystkie informacje dotyczące praw autorskich powinny znajdować się razem w jednym zdarzeniu Copyright Note. Powinno to być pierwsze zdarzenie, w pierwszym bloku pliku o interwale czasowym 0. Tekst informacji kodowany jest w kodzie ASCII.



# Wybrane metazdarzenia MIDI

<b>Nazwa:</b>	Lyrics – Liryka
<b>Typ:</b>	05
<b>Długość:</b>	Zmienna
<b>Format:</b>	FF 05 <długość> <tekst>

## Opis:

Tekst przeznaczony do wykonania wokalnego. Zasadniczo każda sylaba tekstu piosenki powinna być oddzielnym metazdarzeniem Liryka.



# Techniki przetwarzania mediów cyfrowych

Wykład 6: Wprowadzenie do przetwarzania obrazów cyfrowych (3 h)



Fundusze  
Europejskie  
Wiedza Edukacja Rozwój



Politechnika Wrocławskiego



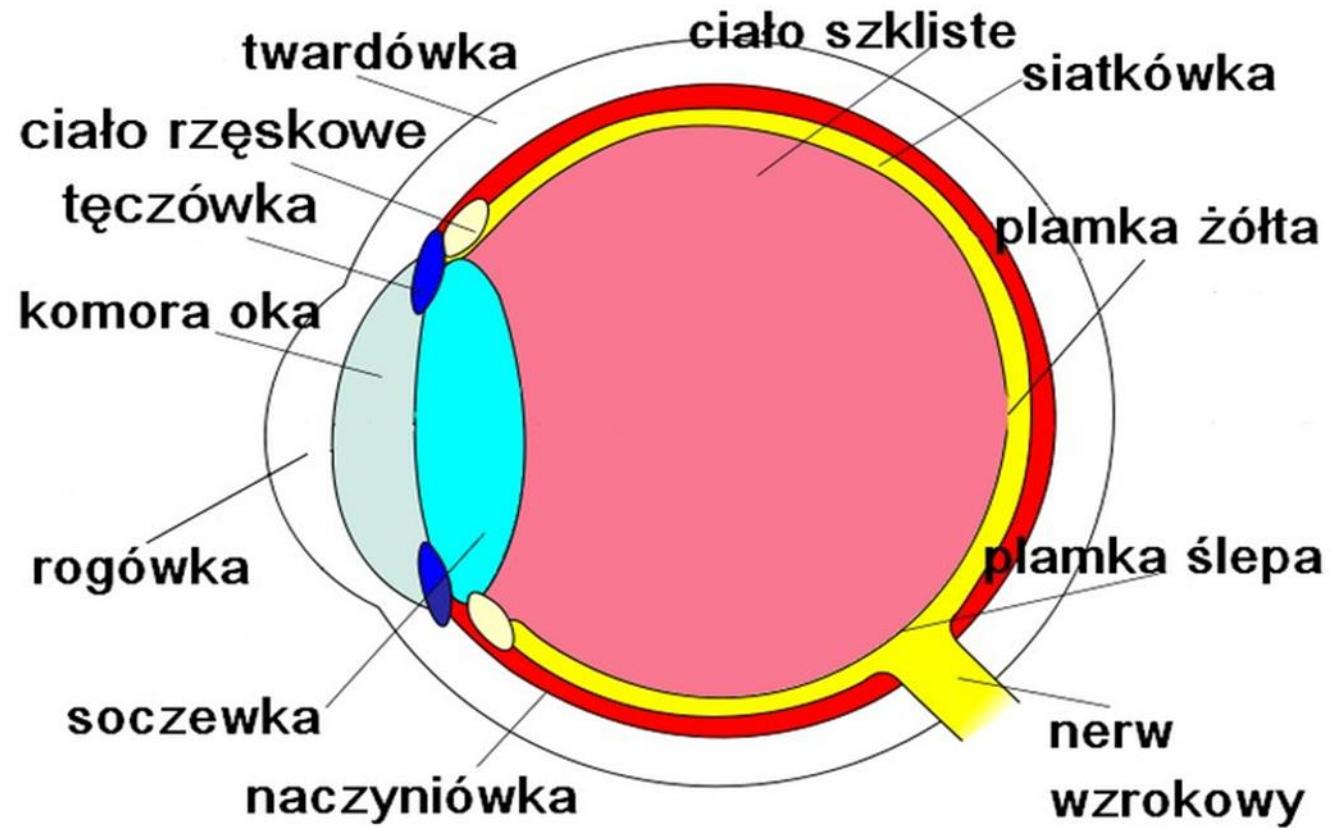
Unia Europejska  
Europejski Fundusz Społeczny

# Postrzeganie obrazów przez człowieka

- Obrazy są naturalną formą prezentacji danych, ponieważ człowiek do 90% informacji odbiera za pośrednictwem zmysłu wzroku.
- Percepcja wzrokowa jest zdolnością do rozpoznawania, różnicowania i interpretowania bodźców wzrokowych.
- Bodźce wzrokowe wywołane są przez pewien zakres promieniowania elektromagnetycznego, na które reaguje siatkówka.
- U człowieka ta część widma nazywa się światłem widzialnym.
- Zawiera się ono w przybliżeniu w zakresie długości fal 380–750 nm
- Bezpośrednio za odbieranie bodźców wzrokowych odpowiada zmysł wzroku – oko.



# Postrzeganie obrazów przez człowieka

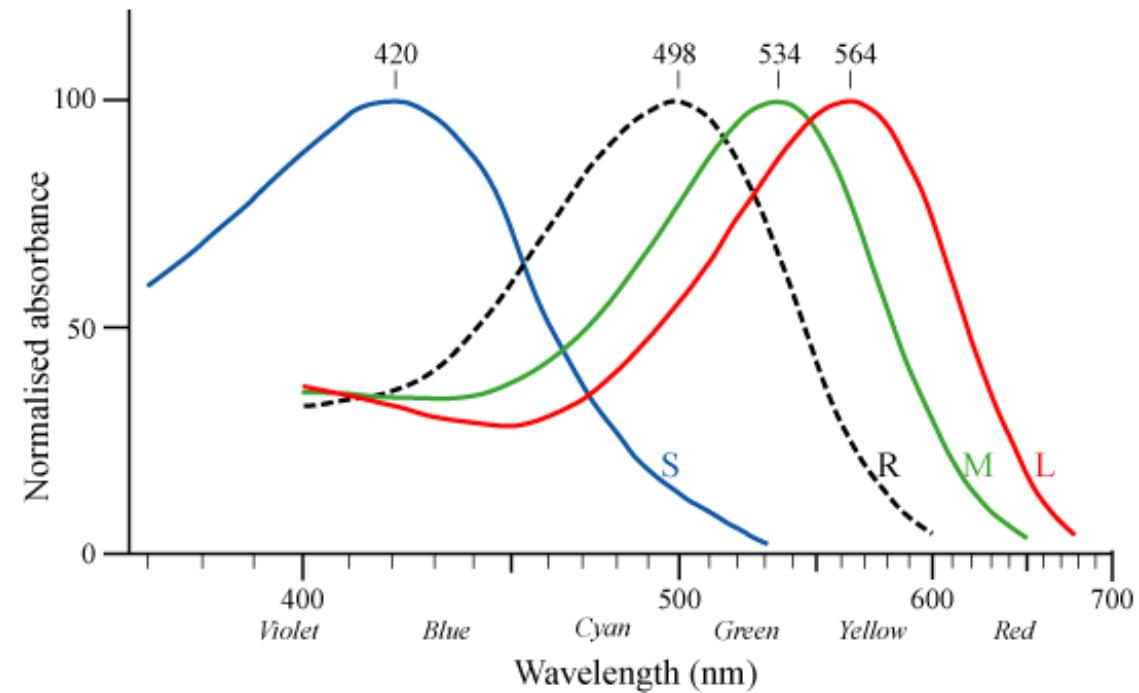


# Postrzeganie obrazów przez człowieka

- Siatkówka jest tą częścią oka, która odbiera bodźce wzrokowe.
- Siatkówka zbudowana jest z ułożonych w kilka warstw komórek nerwowych, które łączą się z mózgiem poprzez nerw wzrokowy.
- Światłoczułymi receptorami siatkówki są zmodyfikowane neurony: czopki i prećiki. Siatkówka ludzkiego oka zawiera ok. 6 mln czopków i 100 mln prećików.
- Prećiki są wrażliwe na natężenie światła i pozwalają na widzenie czarno-białe. W swojej zewnętrznej warstwie prećiki zawierają rodopsynę - światłosensitive barwnik. Prećiki nie występują w dołku środkowym, za to jest ich dużo w częściach peryferyjnych siatkówki.
- Czopki skupiają się w centralnej części siatkówki i odpowiadają za widzenie barwne. Zawierają trzy barwniki wrażliwe na światło niebieskie, zielone i czerwone. Czopki są też odpowiedzialne za ostrość widzenia. Największa ilość receptorów znajduje się w plamce żółtej, zaś w plamce ślepej nie ma ani jednego.
- Do siatkówki przylega od tyłu warstwa komórek wypełnionych czarnym pigmentem, który absorbuje nadmiar światła wpadającego do oka i zapobiega zamazywaniu (zacieraniu konturów) obrazu przez światło odbite wewnątrz oka.
- Na siatkówce powstaje obraz rzeczywisty, pomniejszony i odwrócony.

# Postrzeganie obrazów przez człowieka

Absorbcja fal świetlnych o określonej częstotliwości przez trzy rodzaje czopków i pręciki (linia czarna przerywana)



# Postrzeganie kolorów przez człowieka

- Postrzeganie kolorów jest subiektywnym wrażeniem psychicznym powstającym w mózgu człowieka.
- Z punktu widzenia fizyki oko ludzkie reaguje tylko na pewien zakres promieniowania elektromagnetycznego, które dociera do zmysłu wzroku najczęściej w postaci mieszaniny fal o różnych częstotliwościach.
- W szczególności, konkretny człowiek może odczuwać ten sam kolor odbierając wielu różnych kombinacji ilościowych fal o różnych częstotliwościach, i odwrotnie - ta sama mieszanina fal świetlnych może u tego samego człowieka wywołać odmienne wrażenia kolorystyczne w zależności od wielu czynników, jak: rodzaj oświetlenia czy obecność innych kolorów w polu widzenia.
- Powstawaniem wrażeń barwnych u człowieka, a także praktycznymi i teoretycznymi zagadnieniami dotyczącymi czynników zewnętrznych uczestniczących w tym procesie zajmuje się teoria koloru.



# Postrzeganie kolorów przez człowieka

- Ile kolorów widzi człowiek?
- Widzenie monochromatyczne charakteryzuje się całkowitą niezdolnością do postrzegania kolorów. Osoby z taką wadą widzenia reagują jedynie na intensywność światła i widzą jedną kolor w wielu odcieniach.
- Widzenie dichromatyczne (daltonizm) jest odmianą ślepoty barw. Charakteryzuje się niezdolnością do rozpoznawania koloru zielonego lub myleniem go z kolorem czerwonym. Osoba dotknięta tym schorzeniem nie potrafi rozróżnić koloru czerwonego, pomarańczowego, żółtego i zielonego, co powoduje odbieranie koloru cyjan i fiolet jako tej samej barwy. U daltonistów nie zmienia się natomiast percepcja jasności kolorów.
- Widzenie trichromatyczne (właściwe większości ludzi człowiekowi) polega na odbiorze kolorów przez 3 różne rodzaje komórek receptorowych – czopków reagujących na światło o różnej długości fali. Ten sposób widzenia pozwala na rozróżnienie od 400 000 do 1 000 000 kolorów.
- Widzenie tetrochromatyczne związane jest z bardzo rzadką mutacją chromosomu X, która powoduje, że w oku znajdują się cztery rodzaje czopków. Na skutek tego oko „rozkłada” kolory nie na trzy lecz na cztery składowe. Osoby „obdarzone” taką anomalią są w stanie dostrzec około 100 mln barw. Tetrochromatyzmem obarczone są wyłącznie kobiety.

# Postrzeganie kolorów przez człowieka

- Sposób postrzegania kolorów przez człowieka wyjaśniają dwie, z pozoru przeciwnostawne teorie widzenia barwnego: teoria trichromatyczna Younga-Helmholtza oraz teoria barw przeciwnych Ewalda Heringa.
- Thomas Young zapoczątkował badania nad widzeniem barwnym pod koniec osiemnastego i na początku dziewiętnastego wieku. W roku 1802, na podstawie wyników swoich wcześniejszych badań Young sformułował teorię o istnieniu w oku ludzkim trzech rodzajów receptorów wrażliwych na światło o trzech różnych barwach.
- Herman von Helmholtz rozwinął teorię Younga i opublikował w roku 1860. Ostatecznie nosi ona nazwę teorii widzenia trójbarwnego lub (od nazwisk jej twórców) teorią Younga – Helmholtza.
- Wyjaśnia ona widzenie barwne zakładając, że w oku istnieją trzy rodzaje receptorów wrażliwych na światło o konkretnej barwie: czerwonej, zielonej i niebieskiej.
- Helmholtz sformułował też używane do dzisiaj pojęcia związane z postrzeganiem barw: jasność, odcień i nasycenie. Ponadto wyjaśnił różnice między addytywnym a substraktywnym mieszaniem kolorów.

# Postrzeganie kolorów przez człowieka

- Druga z teorii wyjaśniających widzenie barwne zaproponowana przez Ewalda Heringa w roku 1872.
- Mówi ona o istnieniu sześciu barw podstawowych, które tworzą trzy pary barw przeciwnych: żółta - niebieska, czerwona - zielona, biała - czarna.
- Teoria Heringa początkowo wydawała się teorią przeciwną do teorii Younga – Helmholtza. Jednak w wyniku dalszych badań okazało się, że obie teorie są komplementarne.
- Teoria Younga - Helmholtza tłumaczy zasadę odbierania konkretnej barwy przez receptory oka.
- Natomiast teoria Heringa tłumaczy sposób przekazywania informacji o barwie od sensorów w oku do mózgu, gdzie ostatecznie powstaje wrażenie barwy. Dzięki temu jest ona w stanie wytłumaczyć takie efekty, jak widzenie następcze i efekt powidoku.
- Teoria Heringa nazywana teorią barw przeciwnych, jest podstawą współczesnego systemu NCS (Natural Color System), który porządkuje kolory zgodnie ze sposobem ich postrzegania przez ludzkie oko. NCS jest powszechnie stosowany np. w archeologii, pracach konserwatorskich (odnowa zabytkowych budowli), ale też np. w protetyce.

# Modele kolorów

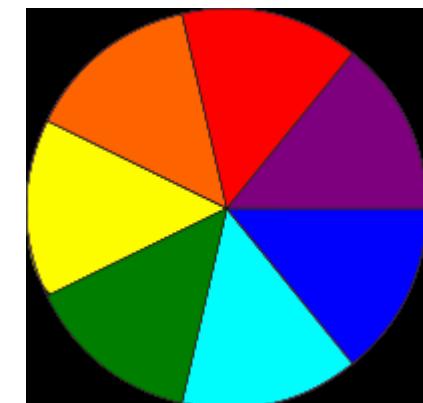
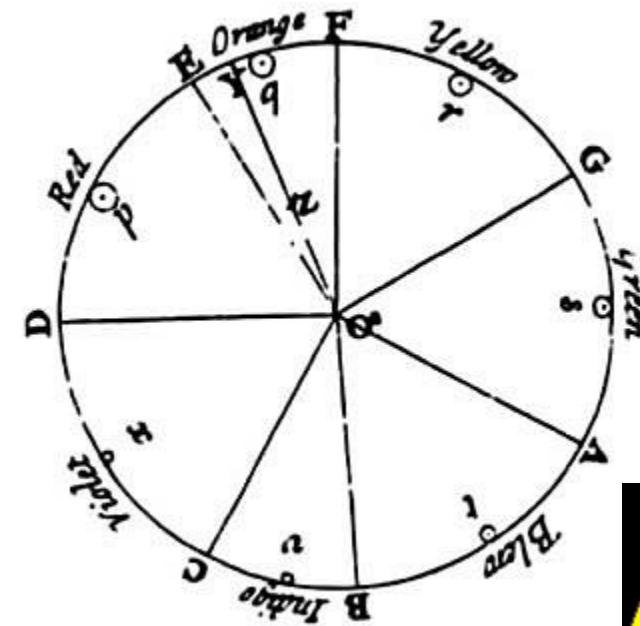
- Wszystkie modele kolorów jako podstawę przyjmują trzy odrębne parametry odbierane przez człowieka: barwę, jasność i nasycenie.
- Barwa (ang. Hue) jest wynikiem mieszania kilku znajdujących się w widmie barw podstawowych, które oko ludzkie odczuwa jako wyraźnie inne od pozostałych. Te barwy mają swoje nazwy, a barwy pośrednie określane są jako procentowy udział sąsiednich barw podstawowych.
- Jasność (ang. Brightness/Value) jest cechą koloru rozumianą intuicyjnie, zależną od ilości światła (jego energii, czyli amplitudy fali) i oznacza intensywność koloru. Wrażenie jasności odnosi się wyłącznie do tej samej barwy.
- Fotometryczną miarą jasności jest luminancja, inaczej miara wrażenia wzrokowego, które odbiera oko ze świecącej powierzchni. Jednostką luminancji jest kandela.
- Trzecim parametrem koloru jest nasycenie (ang. Saturation), wyrażające czystość koloru. Im barwa jest bardziej nasyciona, tym jest odbierana jako bardziej „żywa”. Kolory mniej nasycone, postrzegamy jako przytłumione i bliższe szarości (odcienie szarości charakteryzują się brakiem nasycenia).
- Wrażenie nasycenia odnosi się do kolorów o tej samej barwie i jasności.
- Stopień nasycenia może być traktowany jako odległość danego koloru od koloru szarego o tej samej jasności (w wybranym układzie współrzędnych).
- Chromatyczność to jakościowa cecha barwy określająca jej odcień i nasycenie. Przez odcień rozumie się „stopień, w którym bodziec może być opisany jako podobny lub różny od bodźców opisanych jako: czerwony, zielony, niebieski i żółty.



# Modele kolorów

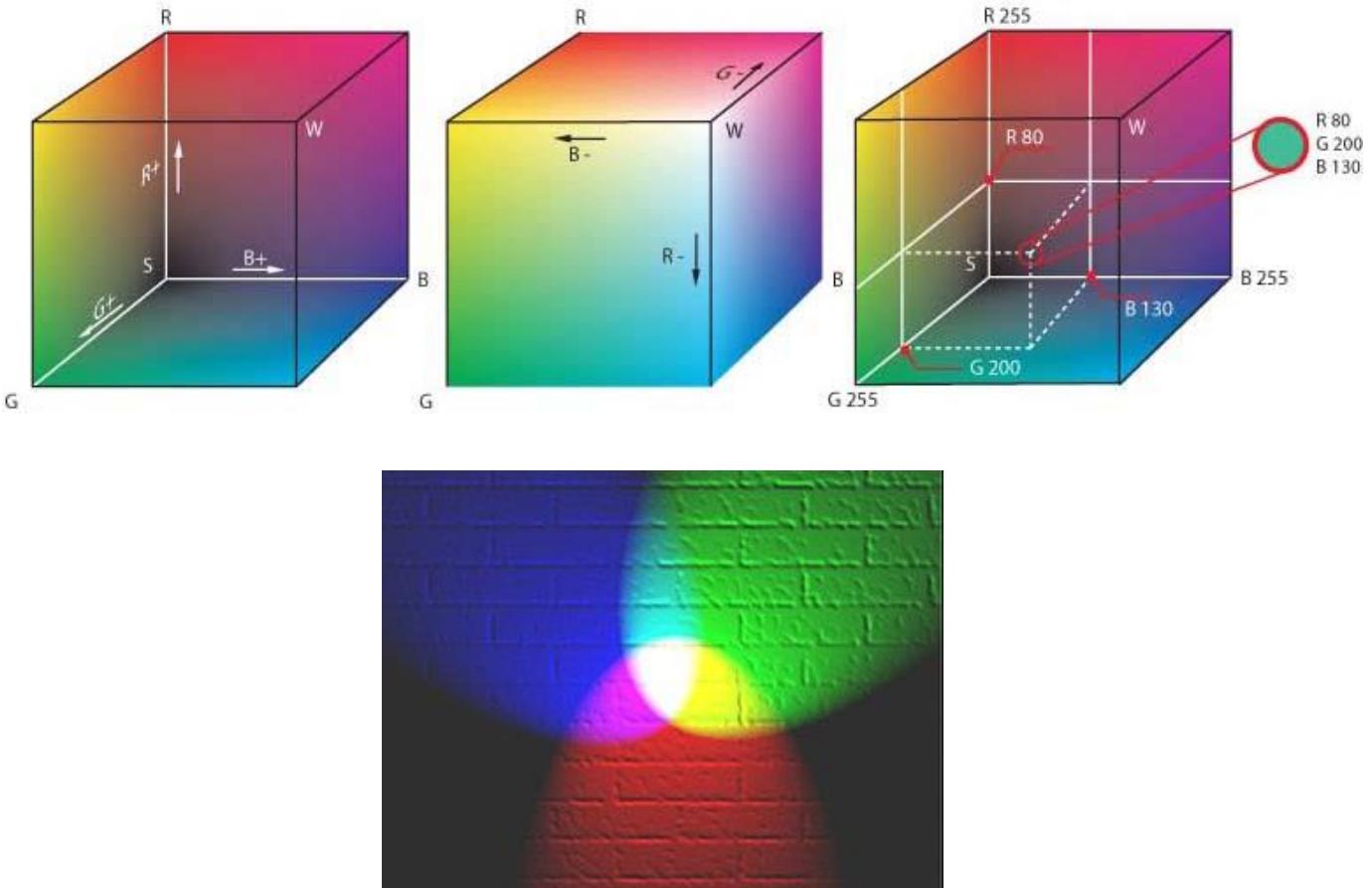
- Jednym z pierwszych uczonych, którzy zajmowali się badaniami nad postrzeganiem kolorów był Izaak Newton
- W wyniku eksperymentów z rozszczepianiem światła za pomocą pryzmatu Newton ustalił, że dowolny kolor można uzyskać mieszając w różnych proporcjach zaledwie kilka tzw. kolorów pierwotnych.
- W wydanym w 1704 r. dziele Optyka Newton opublikował tzw. koło kolorów przedstawiające geometryczną zależność między kolorami prymarnymi i ich pochodnymi.
- W tym pierwszym modelu kolorów Newtona kolorami prymarnymi są: czerwony, żółty i niebieski.
- Zmieszanie kolorów żółtego i czerwonego daje kolor pomarańczowy, czerwonego i niebieskiego daje fiolet, a niebieskiego i żółtego daje kolor zielony.
- W swoim modelu Newton uwzględnił dodatkowy kolor - indygo, chociaż współcześnie kolor ten traktuje się jako odmianę koloru niebieskiego.

Koło kolorów Newtona



# Modele kolorów - RGB

- Najbardziej znanym i powszechnie stosowanych w urządzeniach cyfrowych jest model RGB wykorzystujący ludzkie postrzeganie kolorów.
- Przestrzeń barw modelu określona jest współrzędnymi R – czerwony, G – zielony, B – niebieski
- RGB jest modelem addytywnym. Tworzenie barw polega na dodawaniu kolorów do „ciemności”.
- Każdy punkt przestrzeni jest kolorem powstającym z zsumowania składowych RGB o wartościach stanowiących współrzędne w odpowiednich wymiarach.



# Modele kolorów - RGB

- Model RGB jest wykorzystywany w wyświetlaczach cyfrowych i w cyfrowych urządzeniach analizujących.
- Liczbę kolorów możliwych do uzyskania na danym urządzeniu określa rozdzielcość każdej ze składowych:
  - R – 5 bitów, G – 6 bitów, B – 5 bitów, łącznie 16 bitów – 65 536 kolorów
  - R – 8 bitów, G – 8 bitów, B – 8 bitów, łącznie 24 bity – 16 777 216 kolorów
  - R – 8 bitów, G – 8 bitów, B – 8 bitów, 8 – bitów kanał przezroczystości alfa lub nieużywane, łącznie 32 bity – 16 777 216 kolorów
  - R – 16 bitów, G – 16 bitów, B – 16 bitów, łącznie 48 bitów –  
- 281 474 976 710 656 (281 bilionów 474 miliardy 976 milionów 710 tysięcy 656) – wyłącznie w zastosowaniach profesjonalnych

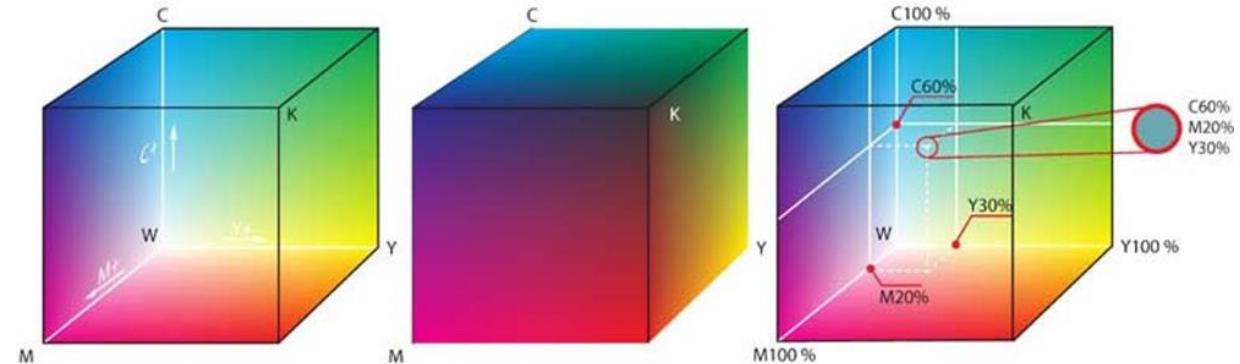
# Modele kolorów – CMY(K)

- Przedmioty w otaczającym nas świecie nie mają kolorów, a jedynie zdolność pochłaniania fal świetlnych.
- Kolor przedmiotu jest kolorem części światła odbitego od takiego przedmiotu.
- Ponieważ skład światła odbitego wynika z odejmowania składników światła białego, to taką metodę tworzenia kolorów nazywa się metodą subtraktywną. Polega ona na „odejmowaniu kolorów od bieli”.
- Modelem kolorów wykorzystującym tę metodę jest CMY (C – cyjan, M – magenta, Y – yellow).



# Modele kolorów – CMY(K)

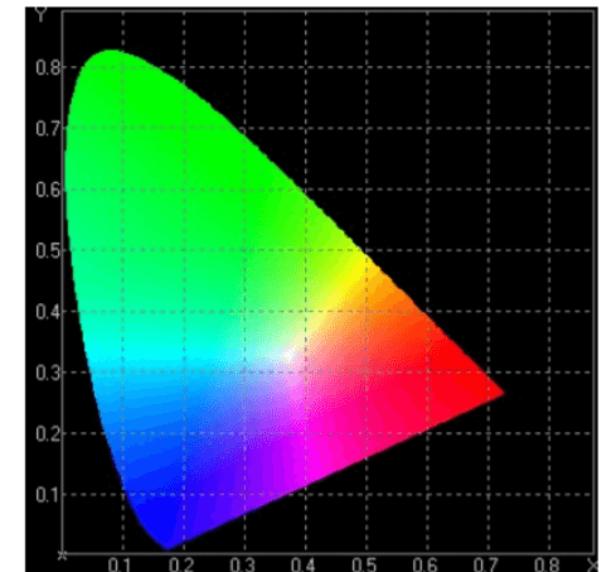
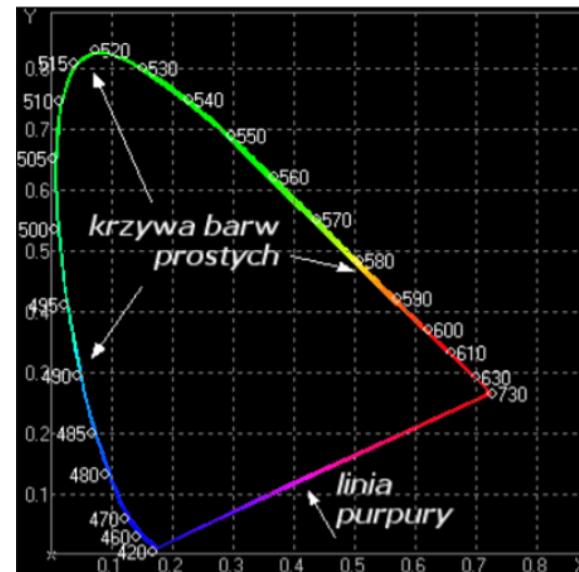
- Przestrzeń barw modelu CMY obrazuje sześcian (analogicznie do RGB).
- Punkt, w którym wszystkie składowe osiągają minimum określa kolor biały.
- W zastosowaniach praktycznych model kolorów CMY nie pozwala na utworzenie wszystkich kolorów widzianych przez ludzkie oko. Na przykład nie da się uzyskać koloru czarnego (kolor szaro-brunatny).
- Dlatego, głównie w poligrafii używa się dodatkowych farb czarnych (oznaczanych literą K - key) dla zwiększenia głębi kolorów, czytelności rysunku i do druku tekstu.
- Kolory (współrzędne) CMY można obliczyć na podstawie ich wartości w modelu RGB:
  - C = 255 – R
  - M = 255 – G
  - Y = 255 – B



# Modele kolorów – CIE XYZ (CIE 1931)

- Barwę jednoznacznie opisują współrzędne trójchromatyczne X, Y, Z, które odpowiadają procentowemu udziałowi trzech podstawowych barw R (czerwonej), G (zielonej) i B (niebieskiej).
- Podstawowy model trójwymiarowy X, Y, Z został przekształcony w dwuwymiarową przestrzeń x, y, Y, gdzie x, y określają chromatyczność barwy (jej odcień i nasycenie), zaś Y – jej jasność (luminancję).
- Przestrzeń xyY przedstawiana jest jako obszar zamknięty dwiema liniami – krzywą barw monochromatycznych i prostą linią purpury.
- Inną formą prezentacji przestrzeni xyY jest wykres chromatyczności, inaczej trójkąt barw lub trójkąt Maxwell'a

CIE XYZ jest trójwymiarowym modelem barw stworzonym w 1931 r. przez Comission Internationale de l'Eclairage (CIE) i traktowany jest jako standard dla innych modeli: CIE LUV oraz CIE Lab.



# Modele kolorów – CIE Lab

- Model CIE Lab jest transformacją przestrzeni CIE XYZ, którą opracowano w 1976 roku.
- Definiując ją założono, że barwy będące w identycznej odległości od siebie będą postrzegane jak jednakowo różniące się.
- Zakładając, że barwa nie może być jednocześnie purpurowa i niebieska i żółta lub zielona, do opisu modelu wybrano następujące składowe: L – jasność (luminacja), a – barwa od zielonej do magenty oraz b – barwa od niebieskiej do żółtej.
- Model CIE Lab umożliwia proste porównywanie barw. Różnicę pomiędzy dwiema barwami wyraża się jako  $\Delta E$ , która obliczana jest na podstawie następującej zależności:

$$\Delta E = \sqrt{(\Delta L)^2 + (\Delta a)^2 + (\Delta b)^2}$$

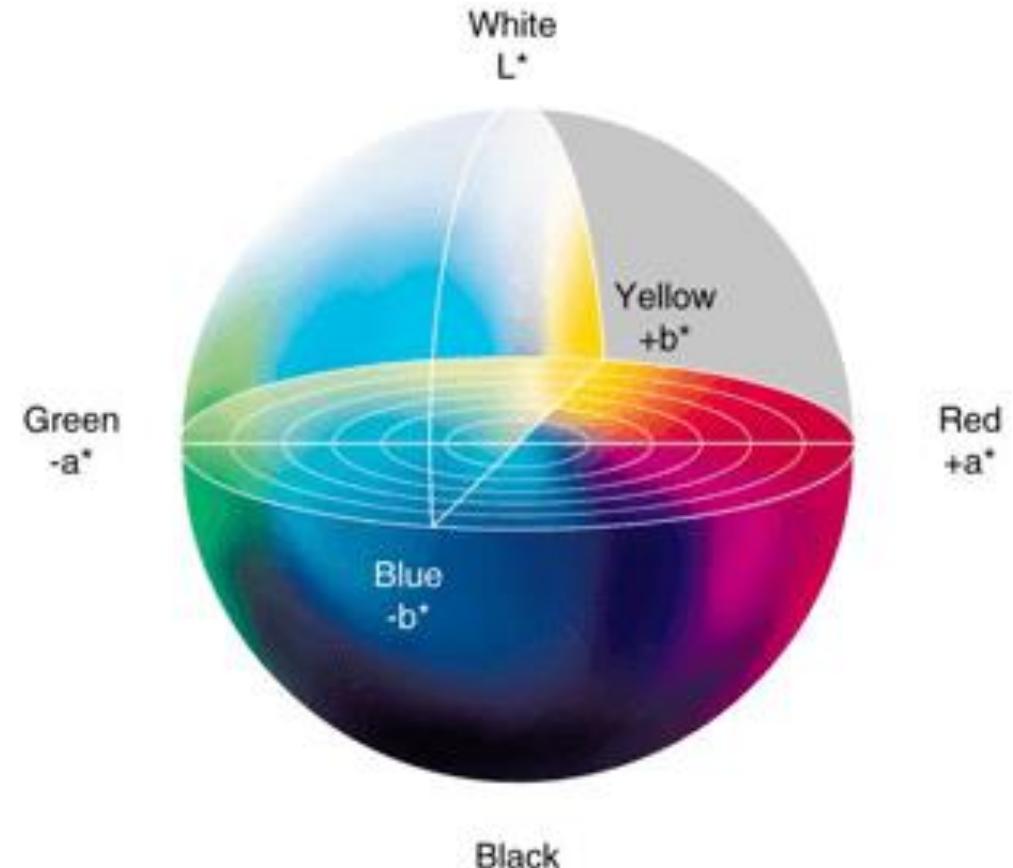
$0 < \Delta E < 1$  – nie widzi różnicę,

$1 < \Delta E < 2$  – jedynie doświadczony obserwator zauważa różnicę,

$2 < \Delta E < 3,5$  – również niedoświadczony obserwator zauważa różnicę,

$3,5 < \Delta E < 5$  – obserwator zauważa wyraźną różnicę barw,

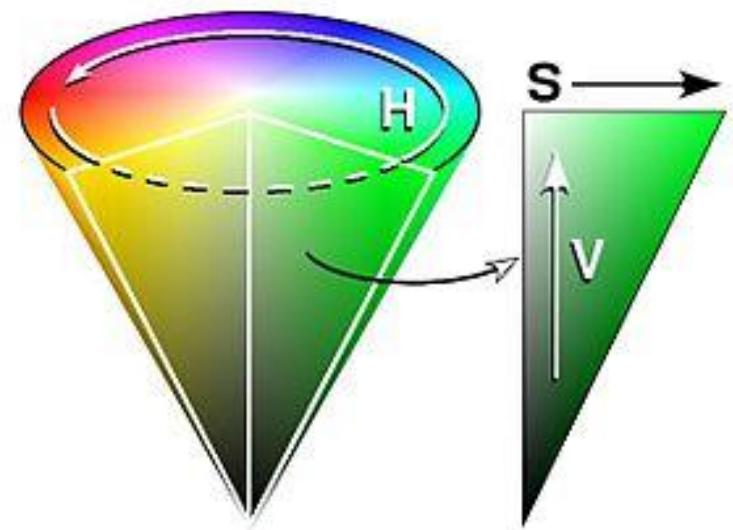
$5 < \Delta E$  – obserwator odbiera barwy jako zupełnie dwie różne.



# Model kolorów - HSV

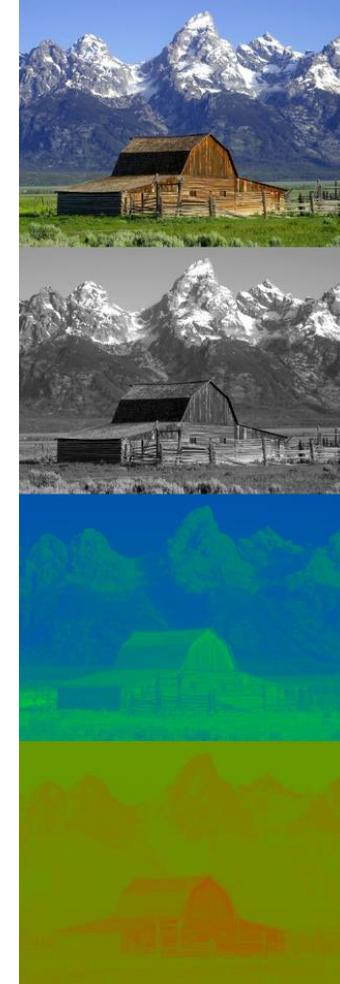
- Model zakłada, że wszystkie barwy wywodzą się ze światła białego. Część jego widma zostaje wchłonięta, a część odbita od oświetlanych przedmiotów.
- Przestrzeń barw ma kształt stożka. Do opisu koloru używa się 3 parametrów:
  - H (Hue) – barwa określana w stopniach od  $0^\circ$  do  $360^\circ$ , w nawiązaniu do koła barw
  - S (Saturation) – nasycenie, promień podstawy w przestrzeni barw, przyjmuje wartości od 0 (w środku podstawy) do 1 (na jej krawędzi).
  - V (Value) – jasność (luminancja), wysokość stożka barw o wartościach od 0 (w wierzchołku) do 1 (w środku podstawy)
- Poszczególnym częstotliwościami fal świetlnych przyporządkowano punkty (stopnie) na kole barw w modelu HSV:
  - centrum barwy czerwonej odpowiada kąt  $0^\circ$  lub  $360^\circ$
  - centrum barwy zielonej odpowiada kąt  $120^\circ$
  - centrum barwy niebieskiej odpowiada kąt  $240^\circ$
- Pozostałe barwy pośrednie dla składowej Hue są odpowiednio rozłożone pomiędzy kolorami czerwonym, zielonym i niebieskim.

Model kolorów HSV zaproponowany przez Alveya Raya Smitha w 1978 roku



# Modele kolorów - YUV

- YUV to model barw wykorzystywany w telewizji analogowej, przy przejściu z systemu telewizji „czarno-białej” do „kolorowej”
- W modelu obraz kolorowy reprezentowany jest przez:
  - Y – jasność (luminancja)
  - U i V – dwa odrębne kanały chrominancji, kodujące barwę:
    - U – przeskalowana barwa niebieska (B)
    - V – przeskalowana barwa czerwona (R)
- Wartości składowych modelu można wyliczyć na podstawie modelu barw RGB, korzystając z zależności:
  - $Y = 0,299R + 0,587G + 0,114B$
  - $U = 0,493 * (B - Y)$
  - $V = 0,877 * (R - Y)$



# Kanał przezroczystości alfa

- Kanał przezroczystości alfa obrazów jest dodatkowym kanałem, czwartym dla RGB, HSV, ... a drugim dla obrazów monochromatycznych
- Kanał alfa określa, w którym miejscu obraz jest przezroczysty (transparentny).
- Wartość alfa = 0 oznacza pełną przezroczystość; obszary obrazu, dla których ustawiono alfa = 0 są całkowicie niewidoczne.
- Wartość alfa = max oznacza całkowitą nieprzezroczystość obrazu; wartość kanału alfa nie wpływa na obraz, barwy pozostają niezmienione
- Wartość ( $0 < \text{alfa} < \text{max}$ ) znaczy, że w obrazie wynikowym pojawi się „część” obrazu w stopniu określonym przez wartość alfa.
- W 24 bitowym formacie zapisu obrazu (r, G, B, alfa), kanał alfa ma rozdzielcość 8 bitów, co pozwala określić stopień przezroczystości
- Nie wszystkie formaty graficzne pozwalają na zapisanie wartości kanału alfa, np. nie ma takiej możliwości w JPEG. W formacie GIF alfa może przyjmować jedynie dwie wartości: pełna przezroczystość lub pełna barwa.
- Format PNG zapewnia pełną obsługę kanału alfa z rozdzielcością 8 bitową (256 możliwych wartości).



# Obraz rastrowy (bitmapa) – obraz wektorowy

- Obraz rastrowy reprezentowany jest w postaci tablicy dwu- lub trójwymiarowej o wymiarach  $M \times N$  (lub  $M \times N \times O$ )
- Komórki macierzy to piksele (lub woksele dla obrazu trójwymiarowego)
- Każdy piksel zawiera wartość lub wektor wartości opisujących jego kolor:
  - Pojedyncza wartość odnosi się do obrazów binarnych
  - Wektor jednowymiarowy - do obrazów monochromatycznych
  - Trójwymiarowy dla obrazów barwnych
- Formaty obrazów rastrowych: BMP, DIB, GIF, IMG, JPG, JPEG, PCX, PNT, MAC, MSP, PIC, PNG, TGA, TIFF, WMF
- W obrazie wektorowym obiekty reprezentowane są przez prymitywy (punkty, proste, krzywe, kształty lub wieloboki)
- Prymitywy graficzne opierają się na równaniach matematycznych.
- W obrazach wektorowych możliwe jest dowolna zmiana rozdzielczości obrazu (skalowanie) bez utraty informacji.
- Przed wyświetleniem obrazu konieczna jest jego zamiana na obraz rastrowy.
- Proces odwrotny nazywa się wektoryzacją.
- Formaty obrazów wektorowych: SVG, CDR, CGM, DRW, DWF, DXB, DXF, EPS, PIC, WMF.

# Konwersja cyfrowa obrazu

- Cyfrowy obraz rastrowy powstaje w procesie konwersji (digitalizacji) obrazu analogowego na jego postać cyfrową, czytelną dla komputera.
- Proces konwersji obejmuje dwa etapy:
  1. próbkowanie
  2. kwantowanie (kwantyzację)
- Próbkowanie jest przekształceniem sygnału obrazu z dziedziny ciągłej na dyskretną. W przypadku obrazów próbkowanie ma charakter dwuwymiarowy i teoretycznie polega na pobieraniu wartości „koloru” tylko w ścisłe określonych miejscach na płaszczyźnie XY obrazu, np. w węzłach siatki prostokątnej.
- W efekcie powstaje funkcja jasności  $J(x,y)$  (dla obrazów monochromatycznych) lub funkcja „koloru” (dla obrazów barwnych) określona na płaszczyźnie dyskretnej, tzn. dla  $x=nT_x$ ,  $y=mT_y$ , gdzie  $n$  i  $m$  są liczbami całkowitymi, a  $T_x$  i  $T_y$  są odstępami (współczynnikami) próbkowania wzdłuż osi X i osi Y. Współczynniki próbkowania determinują rozdzielcość przestrzenną obrazu.
- Kwantowanie obrazów sprowadza się do odwzorowania wartości ciągłych funkcji jasności (lub funkcji „koloru”) na ich skończoną reprezentację dyskretną. Dokonuje się tego dzieląc cały zakres jasności na przedziały i przypisując każdemu z nich pewną wartość dyskretną. Liczba przedziałów określa rozdzielcość kwantowania. Z kolei rozdzielcość kwantowania definiuje głębię kolorów w obrazie cyfrowym.
- Przez głębię koloru (głębię bitową) rozumie się liczbę bitów przeznaczoną na zapisanie informacji o kolorze pojedynczego piksela w danym modelu barw.

# Konwersja cyfrowa obrazu

- Im więcej bitów jest przeznaczonych na zapisanie informacji o kolorze, tym lepsza jest jakość obrazu wynikowego.
- W obrazach binarnych (czarno-białych) do zapisania informacji o kolorze wystarczy 1 bit, np. 0 – piksel biały, 1- piksel czarny.
- W obrazach binarnych często przyjmuje się, że wartości 0 i 1 nie odpowiadają poziomom jasności, lecz oznaczają przynależność pikseli do obiektów znajdujących się na obrazie lub do tła, np. 1 - piksel obiektu, 0 – piksel tła.
- W obrazach monochromatycznych (np. w odcieniach szarości lub sepii) zwykle stosuje się 8 bitów, co daje 256 odcieni. Taki obraz człowiek postrzega jako ciągły w sensie jasności.
- W obrazach kolorowych istotna jest liczba bitów przypadających na kanał w danym modelu kolorów, np. różne rozdzielczości kwantowania dla prezentowanego wcześniej modelu RGB.

# Przetwarzanie obrazów cyfrowych

- Przez przetwarzanie obrazów cyfrowych rozumie się stosowanie szeregu przekształceń zmieniających lub poprawiających jakość obrazów, podkreślających ich składowe pod kątem lepszej obserwacji, analizy i rozpoznawania oraz doprowadzenie do postaci wygodnej do kodowania.
- Techniki poprawy jakości obrazu są głównie procedurami heurystycznymi używanymi do manipulowania obrazem w celu uzyskania pozytywnego wrażenia u człowieka.
- Obejmują one różnorodne operacje, jak zmianę rozkładu poziomów szarości – poprawę kontrastu, redukcję poziomu szumów, wyostrzanie krawędzi i poprawę ostrości, filtrację przestrzenną, interpolację i powiększanie, pseudokoloryzację itd.

# Techniki przetwarzania mediów cyfrowych

Wykład 7: Operacje bezkontekstowe na obrazach (3 h)



Fundusze  
Europejskie  
Wiedza Edukacja Rozwój



Politechnika Wrocławskiego

Unia Europejska  
Europejski Fundusz Społeczny



# Operacje bezkонтекstowe na obrazach

- Operacje bezkontekstowe przekształcają jeden obraz w drugi:

$$O(x, y) \Rightarrow O'(x, y)$$

- Operacje przeprowadza się na pojedynczych pikselach obrazu.
- Wynik operacji zależy wyłącznie od wartości piksela.
- Otoczenie piksela nie wpływa na rezultat operacji.
- Wszystkie piksele o tej samej wartości wejściowej traktowane są jednakowo i po wykonaniu operacji przyjmują tę samą wartość.
- Operacje bezkontekstowe stosowane są w celu:
  - Poprawy jakości obrazu (jasności, kontrastu)
  - Zmiany histogramu
  - Zmiany koloru
  - Wydobycia i/lub uwypuklenia cech obrazu
  - Korekcji Gamma
  - Przeprowadzenia operacji arytmetycznych na obrazach



# Histogram obrazu

- Histogram jest jednym z podstawowych narzędzi wykorzystywanych w przetwarzaniu i analizie obrazów. Stosuje się go w korekcji jakości wizualnej obrazu, procedurze kwantyzacji, kompresji, segmentacji, wyszukiwania obrazów, etc.
- Histogram jest funkcją przyporządkowującą możliwym poziomom jasności lub możliwym kolorom liczbę odpowiadających im pikseli w obrazie.
- Zwykle przedstawia się go w postaci wykresu. Na osi X histogramu znajdują się wszystkie możliwe wartości kolorów (np. od 0 do 255 dla obrazów monochromatycznych), wartość na osi Y jest proporcjonalna do liczności pikseli o danej barwie.
- Histogram opisuje obraz globalnie, tj. podaje liczbę wystąpień pikseli określonego koloru, ale nie podaje informacji o ich rozmieszczeniu w przestrzeni
- W praktyce jest prawie niemożliwe odtworzenie obrazu na podstawie histogramu.
- Jest jednym z podstawowych narzędzi wykorzystywanych w przetwarzaniu i analizie obrazów. Jest wykorzystywany m. in. jako etap procedur korekcji jakości wizualnej obrazu, kwantyzacji, kompresji, segmentacji, wyszukiwania obrazów, etc.

# Histogram obrazu

- W zastosowaniach praktycznych wykorzystuje się histogram skumulowany i histogram znormalizowany.
- Histogram skumulowany powstaje poprzez zsumowanie dla każdej wartości koloru liczby pikseli o wartościach mniejszych lub równych od aktualnie rozpatrywanej:

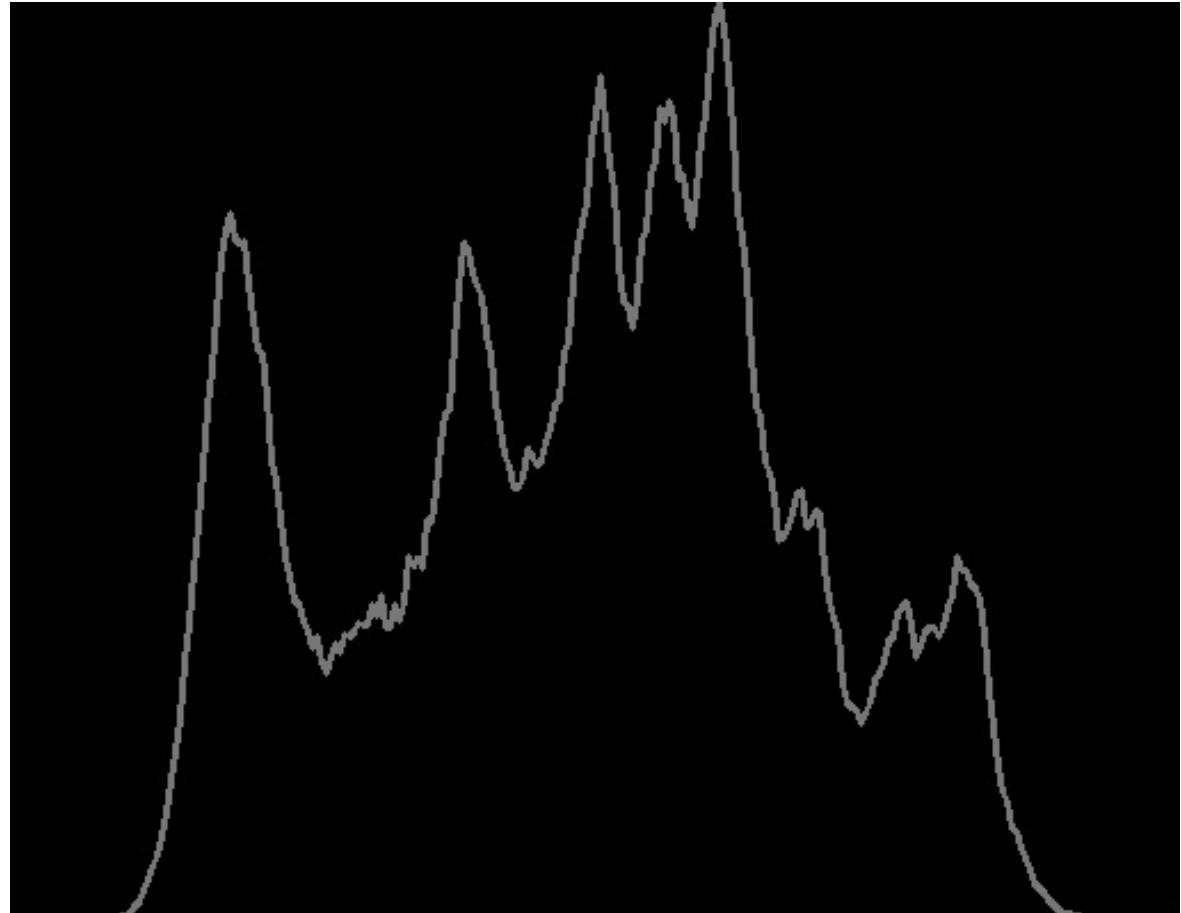
$$h(k) = \sum_{i=1}^k n(i)$$

- Histogram znormalizowany (unormowany) powstaje poprzez uśrednienie kolejnych „prążków” histogramu względem całkowej liczby pikseli:

$$h(k) = \frac{n_k}{N},$$

gdzie  $n_k$  - liczba pikseli koloru  $k$ ,  $N$  – liczba wszystkich pikseli w obrazie.

# Histogram obrazu - przykład



Fundusze  
Europejskie  
Wiedza Edukacja Rozwój



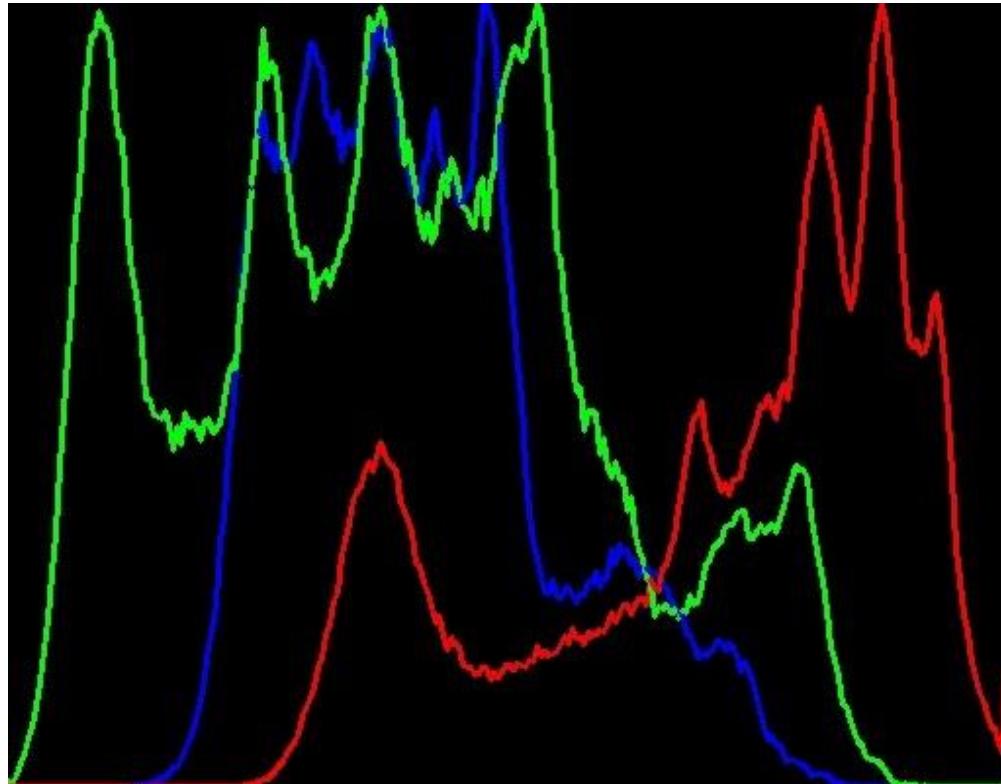
Politechnika Wrocławska



Unia Europejska  
Europejski Fundusz Społeczny

# Histogram obrazu kolorowego - przykład

Histogram obrazu kolorowego wylicza się dla każdej składowej koloru oddzielnie, np. RGB (rys. poniżej) lub HSV



Fundusze  
Europejskie  
Wiedza Edukacja Rozwój



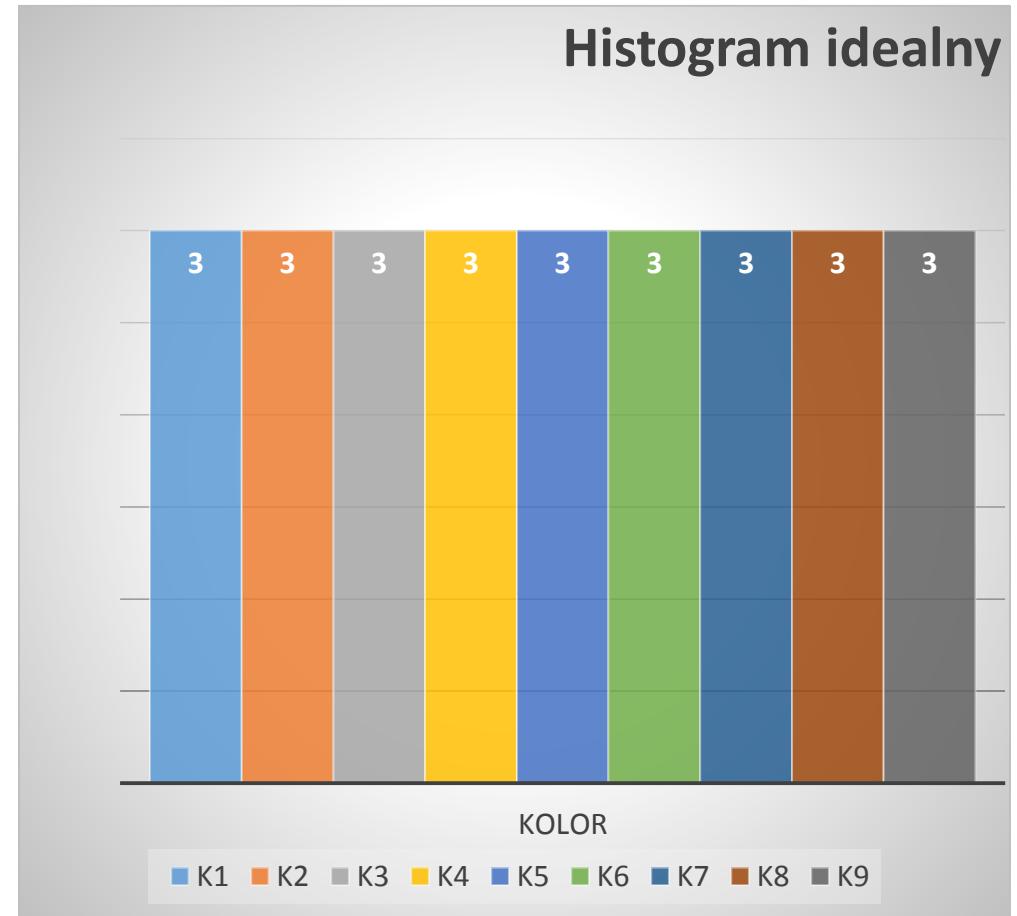
Politechnika Wrocławska



Unia Europejska  
Europejski Fundusz Społeczny

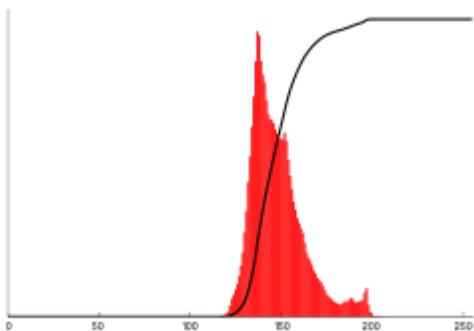
# Wyrównanie histogramu

- Wyrównanie histogramu polega na takim jego przekształceniu, aby wszystkie kolory obrazu były jednakowo liczne.
- Dąży się do tego by jak najbardziej „spłaszczyć” histogram.
- Wyrównanie histogramu służy poprawie jakości obrazu:
  - Kontrastu, tj. współczynnika określającego różnicę między jasnością najciemniejszego i najjaśniejszego piksela
  - Ostrości, tj. stopnia odwzorowania drobnych szczegółów w obiekcie na ekranie
- Ostrość i kontrast to pojęcia ze sobą związane – ostrość obrazu wynika z występowania w nim wysokiego kontrastu.
- Operację wyrównania histogramu przeprowadza się w oparciu o histogram skumulowany

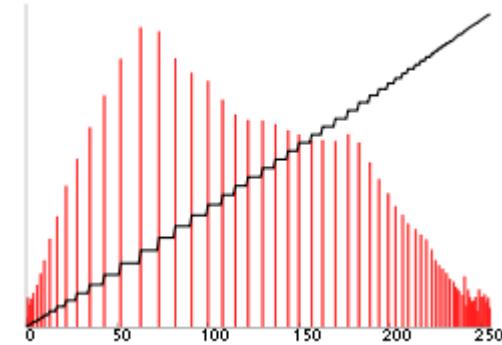


# Wyrównanie histogramu

Obraz bez wyrównania histogramu i odpowiadający mu histogram (wykres czerwony) oraz histogram skumulowany



Ten sam obraz po wyrównaniu histogramu wraz z histogramem oraz histogramem skumulowanym



Fundusze  
Europejskie  
Wiedza Edukacja Rozwój



Politechnika Wrocławska

Unia Europejska  
Europejski Fundusz Społeczny



# Wyrównanie histogramu obrazu kolorowego

- Operację wyrównania histogramu dla obrazów kolorowych przeprowadza się dla każdego kanału obrazu oddzielnie.
- Wygodniej jest wcześniej przejść z kolorów RGB do prezentacji YCbCr i wyrównać histogram składowej luminancji Y.



Fundusze  
Europejskie  
Wiedza Edukacja Rozwój



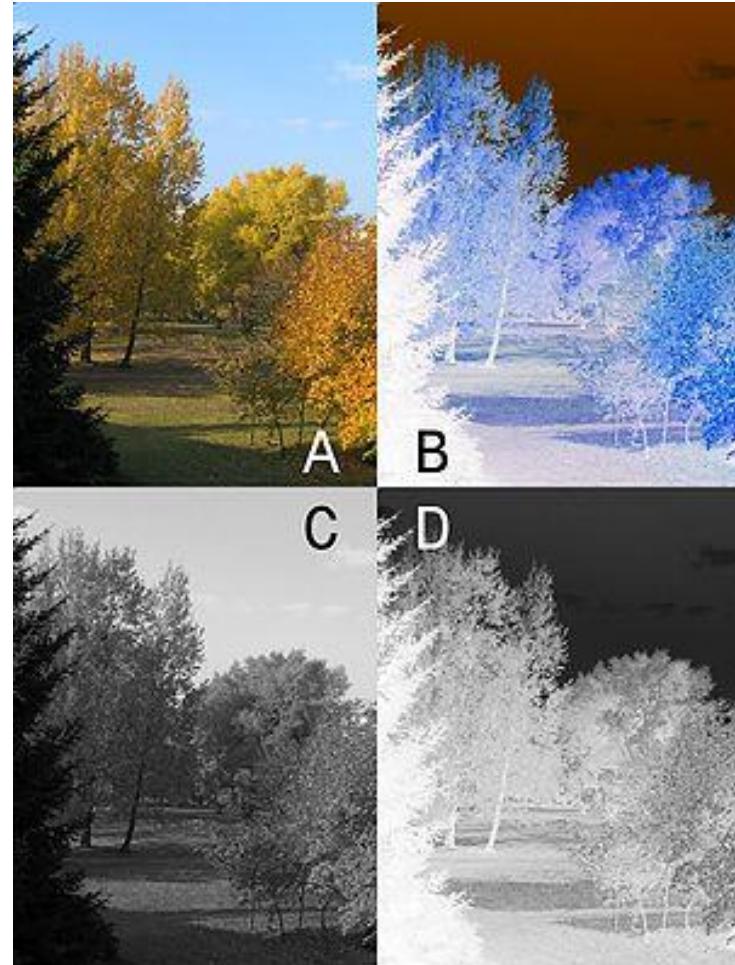
Politechnika Wrocławska

Unia Europejska  
Europejski Fundusz Społeczny



# Zmiana wartości pikseli – LUT (LookUpTable)

- LUT jest tablicą, która definiuje operacje bezkontekstowe
- Tablica podaje wartości pikseli po wykonaniu operacji.  
 $O(x, y) \Rightarrow \text{LUT}(O(x, y))$
- Na przykład, aby uzyskać negatyw obrazu monochromatycznego LUT definiuje się następująco:  
 $\text{LUT}(k + 1) = 255 - k,$   
gdzie  $k \in \langle 0, 255 \rangle$
- Analogiczna operacja dla obrazów kolorowych RGB każdy kanał „odwracany” jest oddzielnie.
- Operacja ta jest w pełni odwracalna



# Progowanie

- Operację progowania definiuje się następująco:
- $O(x, y) \Rightarrow \begin{cases} 0 \text{ jeżeli } O(x, y) < \text{próg} \\ 255 \text{ jeżeli } O(x, y) \geq \text{próg} \end{cases}$
- Progowanie pozwala uzyskać obraz binarny na podstawie obrazu monochromatycznego lub kolorowego.
- Wartość progu może być stała lub zmienna, adaptowana do charakterystyki obrazu.
- Progowanie stosuje się często, aby oddzielić obiekty obrazu od tła



# Progowanie

- Wartość progu można ustalić automatycznie używając:
- Na podstawie kształtu histogramu – analizuje się piki, doliny oraz krzywizny wygładzonego histogramu,
- Metody oparte na grupowaniu – próbki obrazu monochromatycznego są grupowane w dwóch klastrach jako tło i obiekt albo modelowane jako mieszanina dwóch rozkładów Gaussa,
- Metody oparte na entropii – analizuje się entropię tła i obiektów pierwszoplanowych oraz entropię krzyżową między obrazem oryginalnym a jego obrazem binarnym,
- Metody oparte na atrybutach obiektów – używa się miar podobieństwa między obrazami w skali szarości a obrazami binarnymi takich, jak podobieństwo rozmytych kształtów (ang. fuzzy shape similarity), czy zbieżność krawędzi (ang. edge coincidence) itp.
- Metody przestrzenne – używa się rozkładów prawdopodobieństw wyższych rzędów i/lub korelacji między pikselami,
- Metody lokalne – wartość progu dla każdego piksela adaptuje do lokalnej charakterystyki obrazu, w tej metodzie poszczególnym pikselom obrazu mogą być przypisane różne wartości progowe.



# Progowanie

- Operacja progowania dotyczy również obrazów kolorowych
- Dla każdej ze składowych modelu RGB ustala się wartości progowe, a następnie łączy się je za pomocą operacji logicznej AND
- W analogiczny sposób można przeprowadzić progowanie w modelu CMYK
- Częściej jednak wykorzystuje się modele HSV i HSL

# Zwiększenie jasności obrazu

- Operacja dodawania stałej wartości do wszystkich pikseli obrazu powoduje zwiększenie jasności obrazu.
- Należy przy tym uważać na przekroczenie zakresu wartości maksymalnych.
- Przykładowa LUT dla operacji zwiększania jasności obrazu np. o 30:  
$$\text{LUT}(k + 1) = \min(255, k + 30),$$
  
gdzie  $k \in \langle 0, 255 \rangle$
- W obrazie wynikowym nie wystąpią piksele o wartościach mniejszych niż 30.
- Wszystkie wartości z obrazu wynikowego większe lub równe 225 będą miały wartość maksymalną 255.
- W wyniku tej operacji histogram obrazu „przesunie” się o 30 w kierunku wartości większych.

# Zmniejszenie jasności obrazu

- Analogicznie można przeprowadzić operację zmniejszenia jasności obrazu.
- Operację tę przeprowadza się odejmując od obrazu oryginalnego pewną wartość stałą.
- Należy przy tym uważać, by nie wyjść poza zakres wartości minimalnych
- Przykładowa LUT dla operacji odejmowania stałej 30 od obrazu:  
$$\text{LUT}(k + 1) = \max(0, k - 30),$$
  
gdzie  $k \in \langle 0, 255 \rangle$
- W tym przypadku histogram obrazu „przesunie się” o 30 w kierunku małych wartości.
- Piksele o wartościach mniejszych lub równych 30 będą miały wartość wynikową 0.
- Piksele najjaśniejsze w obrazie oryginalnym (o wartościach większych lub równych 225) w obrazie wynikowym przyjmą wartość 225
- Operacja zmiany jasności obrazu powoduje utratę informacji i w związku z tym jest nieodwracalna.



# Dodawanie stałej do obrazu RGB

- Operacja dodawania stałej do obrazu kolorowego powoduje zmianę koloru:
- Dodanie do obrazu wartości np. [30, 0, 0] spowoduje zwiększenie jasności w kanale R.
- W efekcie obraz wynikowy „zaczerwieni się”
- Analogicznie dodanie [0, 30, 0] spowoduje zwiększenie jasności w kanale G i „zazielenienie” obrazu
- Wreszcie dodanie [0, 0, 30] spowoduje zwiększenie jasności w kanale B i wyraźny niebieskawy odcień na całym obrazie.

# Odejmowanie stałej do obrazu RGB

- Odejmowanie stałej wartości od wartości obrazu powoduje zmianę koloru (analogicznie jak dodawanie).
- Odjęcie wartości w kanale R, np. (50, 0, 0), powoduje dodanie barwy cyjan do obrazu.
- Odjęcie (0, 50, 0) w kanale G, sprawia, że obraz wynikowy jest równoważny dodaniu składowej magenta
- Odjęcie wartości (0, 0, 50) od wartości kanału B, daje w efekcie większy udział barwy żółtej w obrazie wynikowym (obraz jest „zażółcony”)



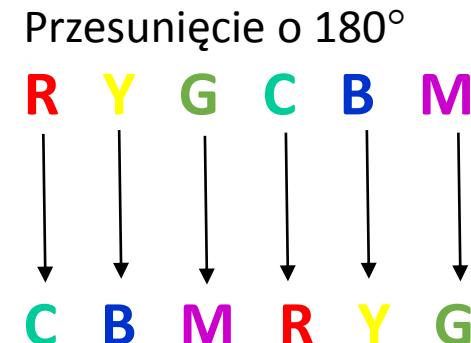
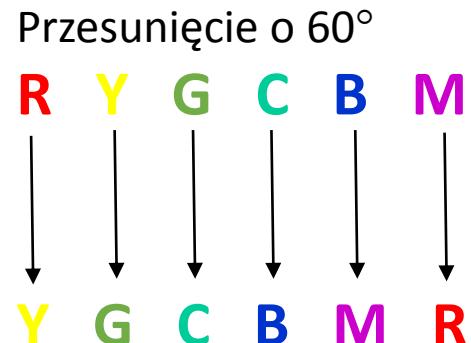
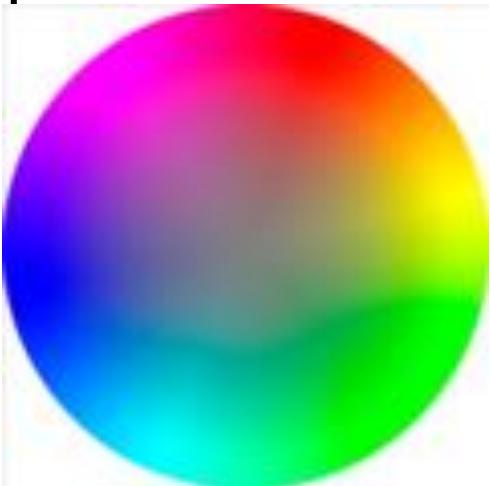
# Dodawanie i odejmowanie stałej do/od obrazu HSV

- W obrazach HSV dodawanie stałej do składowej V jest równoważne rozjaśnieniu obrazu, barwa i nasycenie nie ulegają zmianie.
- Odejmowanie stałej w kanale V, powoduje przyciemnienie obrazu wynikowego. Pozostałe parametry obrazu nie zmieniają się.
- Operacje dodawania i odejmowania stałej w kanale S, zwiększają i zmniejszają nasycenie barw w obrazie wynikowym.
- W szczególności zmniejszenie saturacji do 0, daje w rezultacie obraz monochromatyczny w odcieniach szarości.



# Dodawanie i odejmowanie stałej do/od obrazu HSV

- Wynikiem operacji przeprowadzanych na kanale H (Hue) jest zmiana barwy obrazu.
- W szczególności dodawanie wartości będących wielokrotnością  $60^\circ$ , sprawia, że obraz wynikowy „przechodzi” w kolejne barwy podstawowe z koła barw, np.:



# Arytmetyka obrazów

- Arytmetyka obrazów dotyczy operacji przeprowadzanych na dwóch lub więcej obrazach, zwykle binarnych lub w monochromatycznych, w skali szarości.
- Ogólnie wszystkie operacje arytmetyczne realizowane są wg schematu:

$$O'(x, y) = O_1(x, y) \text{ operator } O_2(x, y)$$

- Operator jest symbolem:
  - Dodawania
  - Odejmowania
  - Mnożenia
  - Dzielenia
  - Operacji logicznych (AND, OR, XOR, NOT)

# Dodawanie obrazów

- Dodawanie obrazów:  $O'(x, y) = O_1(x, y) + O_2(x, y)$
- W przypadku dodawania może pojawić się problem wyjścia poza zakres wartości maksymalnych w obrazie wynikowym. Rozwiązuje się go poprzez:
- Przeskalowanie obrazu wynikowego poprzez uśrednienie wyników:

$$O'(x, y) = \frac{O_1(x, y) + O_2(x, y)}{2}$$

- Przypisanie wszystkim wartośćom większym lub równym od maksimum wartości maksymalnej:  $O'(x, y) = \min(\max(O_1(x, y), O_2(x, y)))$
- Zastosowanie operacji modulo wartość maksymalna (cykliczne przypisywanie wartości):  $O'(x, y) = ((O_1(x, y) + O_2(x, y)) - 256)$
- Operację dodawania obrazów stosuje się do:
  - Nakładania tekstur
  - Usuwania szumu przez dodawanie kolejnych obrazów tej samej sceny i uśrednianie wartości
  - Uwypuklenia cech obrazu, np. wyostrzanie obrazu przez dodanie do niego wcześniej wyodrębnionych krawędzi
  - Uzyskania efektów artystycznych

# Odejmowanie obrazów

- Odejmowanie obrazów:  $O'(x, y) = O_1(x, y) - O_2(x, y)$
- Operacja nieprzemienna, tzn.  $O_1(x, y) - O_2(x, y) \neq O_2(x, y) - O_1(x, y)$
- Odejmowanie jednego obrazu od drugiego może prowadzić do sytuacji, w której wartość wynikowa będzie mniejsza od minimum zakresu zmienności koloru. Rozwiązaniem może być:
- Przypisanie wartościom różnicy ich wartości bezwzględnych:

$$O'(x, y) = |O_1(x, y) - O_2(x, y)|$$

- Przypisanie wartościom ujemnym liczby 0, tj.

$$O'(x, y) = \max(0, O_1(x, y) - O_2(x, y))$$

- Zastosowanie operacji modulo:  $O'(x, y) = ((O_1(x, y) - O_2(x, y)) + 256)$
- Odejmowanie obrazów stosuje się:

- Do tworzenia obrazów różnicowych, tj. wyodrębniania różnic między tymi samymi obiekty na różnych obrazach lub między obrazami podobnych obiektów (głównie w celu wykrywania ruchu)
- Do odejmowania od obrazu niejednorodnego tła – w sytuacji gdy np. oświetlenie przy akwizycji obrazu było nierówne



# Obraz różnicowy - przykład

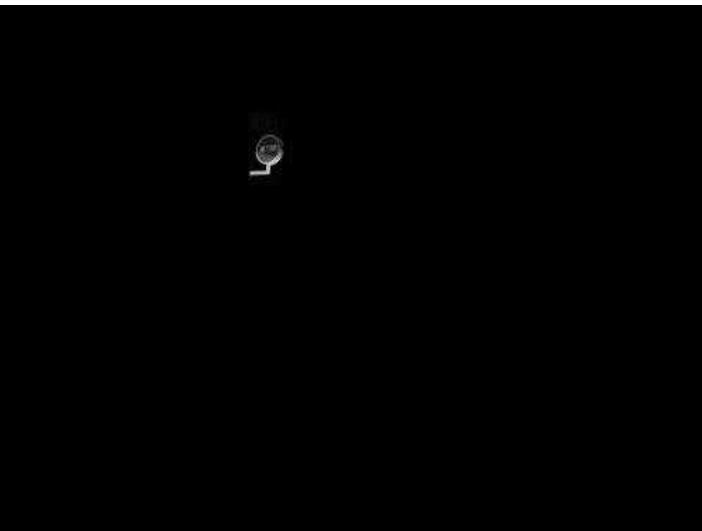
Obraz 01



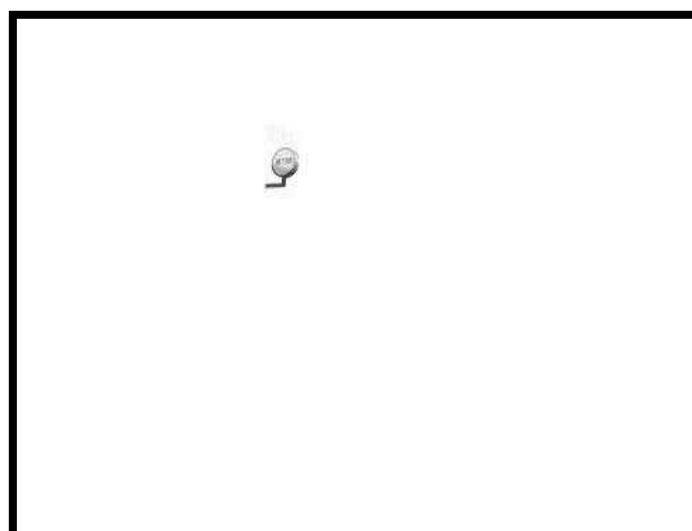
Obraz 02



Obraz  
różnicowy



Negatyw obrazu  
różnicowego



Fundusze  
Europejskie  
Wiedza Edukacja Rozwój



Politechnika Wrocławska

Unia Europejska  
Europejski Fundusz Społeczny



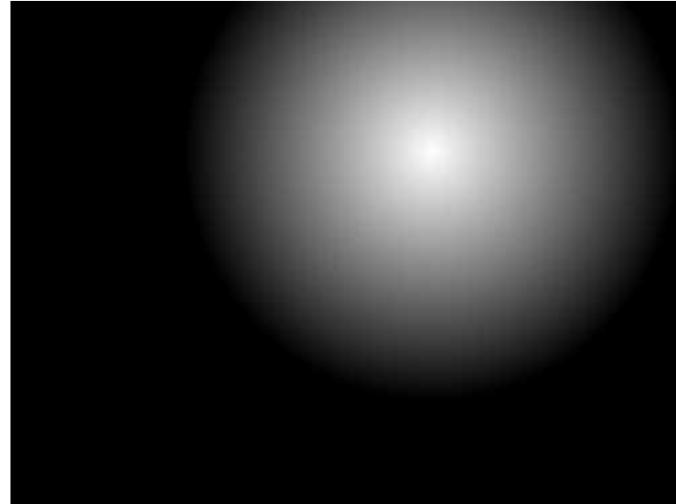
# Mnożenie obrazów

- Mnożenie obrazów:  $O'(x, y) = O_1(x, y) * O_2(x, y)$
- Przekroczenie zakresu występuje częściej niż przy dodawaniu. Stosuje się takie same rozwiązania.
- Mnożenie dwóch obrazów rzeczywistych rzadko stosuje się w praktyce.
- Wyjątkiem jest mnożenie przez obraz binarny. W tym przypadku alternatywnym rozwiązaniem jest operacja logiczna AND, która daje takie same rezultaty.
- Mnożenie obrazu przez stałą powoduje jego rozjaśnienie. Efekt operacji jest wyraźniej widoczny niż w przypadku dodawania stałej.

# Mnożenie obrazów - przykład



X



=



Fundusze  
Europejskie  
Wiedza Edukacja Rozwój



Politechnika Wrocławska

Unia Europejska  
Europejski Fundusz Społeczny



# Potęgowanie obrazów - korekcja gamma

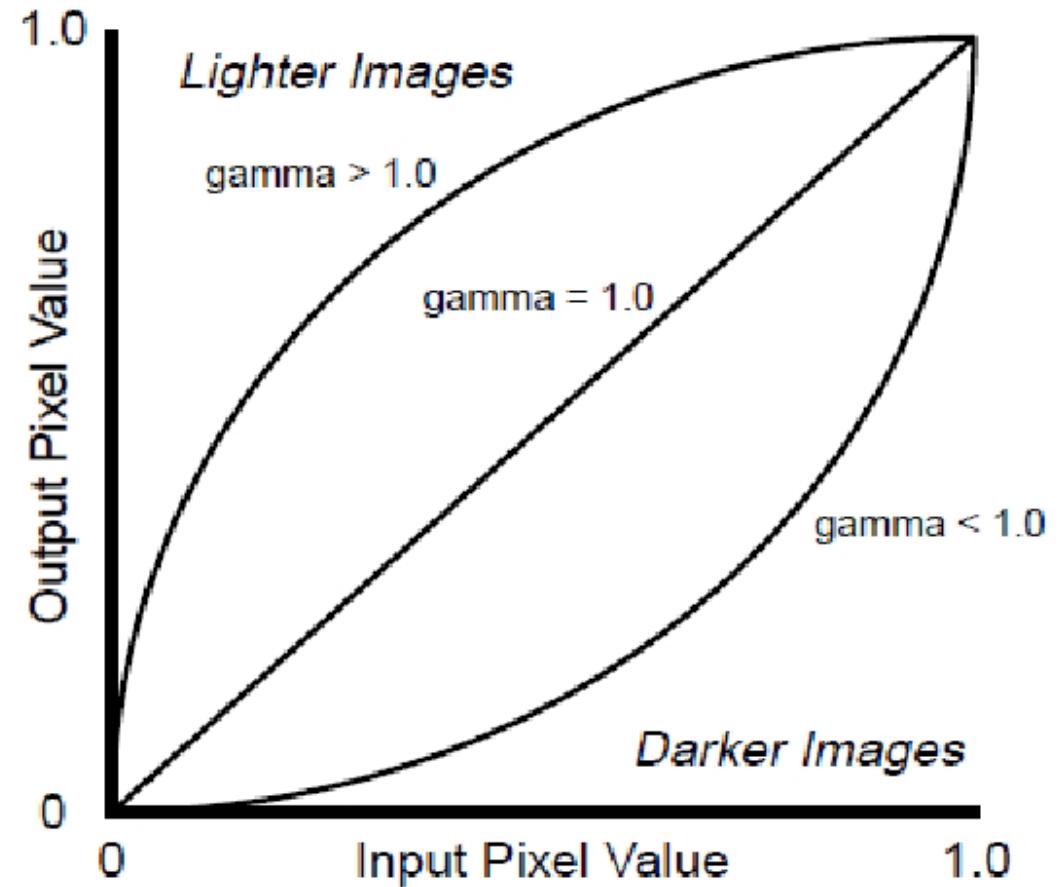
Korekcja gamma jest operacją wykonywaną na obrazie, bazującą na krzywej gamma:

$$O'(x, y) = O(x, y)^\gamma,$$

Gdzie  $\gamma \in \mathbb{R}$

$\gamma > 1$  powoduje rozjaśnienie obrazu

$\gamma < 1$  powoduje jego przyciemnienie



# Dzielenie obrazów

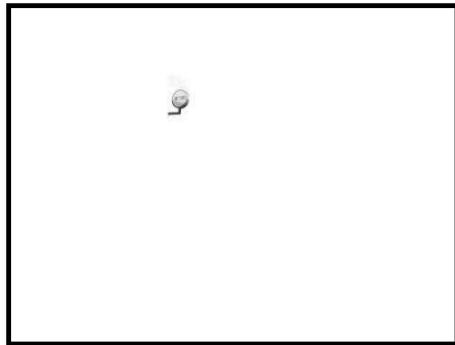
- Dzielenie obrazów:  $O'(x, y) = O_1(x, y)/O_2(x, y)$
- Operacja nieprzemienna. Nieokreślona dla  $O_2(x, y) = 0$
- Zastosowanie dzielenia jest podobne do odejmowania. W przypadku dzielenia obrazu przez tło obraz wynikowy jest wyraźniejszy. Podobnie jest przy wykrywaniu tutaj. Tutaj dodatkowo nie potrzeba odwracać negatywu.



/



=



Fundusze  
Europejskie  
Wiedza Edukacja Rozwój



Politechnika Wrocławska

Unia Europejska  
Europejski Fundusz Społeczny



# Operacje logiczne na obrazie

- Operacje logiczne przeprowadza się najczęściej na obrazach binarnych.
- Dla większości operacji pojedynczy obraz wynikowy powstaje jako rezultat operacji przeprowadzonej na dwóch obrazach wejściowych.
- Najczęściej wykorzystuje się operatory logiczne:
- NOT – operacja jednoargumentowa, w wyniku uzyskuje się negatyw obrazu
- AND – operacja dwuargumentowa, zwykle stosowana do maskowania części obrazu i „wycinania” z niego określonej części
- NAND – negacja iloczynu obrazów
- OR – operacja dwuargumentowa, stosowana między innymi w fotomontażu
- NOR – negacja sumy obrazów
- XOR – suma rozłączna stosowna do wykrywania zmian w obrazie
- NXOR – równoważność logiczna obrazów

# Normalizacja obrazu

- Przez normalizację obrazu rozumie się proces, który zmienia zakres wartości pikseli:

$$O(x, y) \rightarrow O'(x, y), \text{ gdzie } O(x, y) \in [min, max], \\ z \text{ kolej } O'(x, y) \in [min', max']$$

- Normalizacja liniowa obrazu z gradacją kontrastu (monochromatycznego w odcieniach szarości realizowana zgodnie z poniższą regułą:

$$O'(x, y) = \frac{(O(x, y) - min) * (max' - min')}{(max - min) + min'}$$



# Techniki przetwarzania mediów cyfrowych

Wykład 8: Operacje kontekstowe na obrazach



Fundusze  
Europejskie  
Wiedza Edukacja Rozwój



Politechnika Wrocławskiego



Unia Europejska  
Europejski Fundusz Społeczny

# Operacje kontekstowe – filtracja obrazu

- Operacje kontekstowe (filtry) biorą pod uwagę otoczenie piksela.
- Ogólnie można je zdefiniować jako funkcję obrazu, której argumentem jest dany piksel obrazu i jego otoczenie:

$$O'(x, y) = f(O(x + i, y + j)),$$

gdzie  $i = -m, \dots, m$  oraz  $j = -n \dots n$

Obszar  $(2m + 1, 2n + 1)$  tworzy otoczenie piksela (jego kontekst K)

- Ze względu na rodzaj funkcji przekształcającej filtry mają charakter:
  - Liniowy
  - Nieliniowy
- Operacje kontekstowe wykorzystuje się przede wszystkim do:
  - Poprawy jakości obrazu (usuwanie szumu, poprawa ostrości)
  - Rozmycie obrazu
  - Wykrywania cech w obrazach (np. krawędzie, linie, rogi, regiony)

# Operacja konwolucji (splotu)

- Podstawą filtrów przestrzennych jest operacja konwolucji.
- Splot jest działaniem określonym dla dwóch funkcji, w wyniku którego powstaje trzecia funkcja, postrzegana jako modyfikacja dwóch poprzednich. Splot to również wynik tej operacji.
- W przetwarzaniu obrazów wartości pikseli w obrazie można traktować jako wartości pewnej dwuwymiarowej, przestrzennej funkcji obrazu.
- Dla dwuwymiarowego obrazu dyskretnego splot określony jest zależnością:

$$O'(x, y) = \sum_{i,j \in K} O(x - i, y - j) * M(i, j) \text{ lub } O' = O * M$$

Gdzie:

$K$  – otoczenie piksela o współrzędnych  $(x, y)$ , obszar określony przez punkty:

$$((-m; m), (-n; n))$$

$M(i, j)$  - wagi pikseli otaczających piksel o współrzędnych  $(x, y)$ , maska (jądro) konwolucji  
 $O(x, y)$  - dyskretna funkcja obrazu źródłowego



# Własności splotu

1. Konwolucja jest przemienna:

$$f * g = g * f$$

Kolejność splatania funkcji nie ma znaczenia

2. Splot jest operacją łączną:

$$f * (g * h) = (f * g) * h$$

Kolejność wykonywania operacji splotu nie ma znaczenia, to ważna własność z punktu widzenia filtracji obrazu wieloma maskami – można je zastąpić jedną, większą lub odwrotnie – jedną dużą maskę można rozdzielić na dwie mniejsze

3. Konwolucja jest rozdzielna względem dodawania:

$$f * (g + h) = f * g + f * h$$

4. Elementem neutralnym splotu („jedynka”) jest [1]:

$$[1] * f = f * [1] = f$$



# Wynik konwolucji

- Podobnie, jak w przypadku operacji bezkontekstowych, tak i przy filtracji może się zdarzyć, że wynik splotu wykroczy poza zakres zmienności piksela.
- Stosuje się dwa rozwiązania, zmierzające do normalizacji wyniku:

1. Jeżeli suma współczynników maski  $\sum_{i,j} M(i,j) > 1$ , to wynik konwolucji normalizuje się dzieląc go przez tę sumę:

$$O'(x,y) = \frac{\sum_{i,j \in K} M(i,j) * O(x-i, y-j)}{\sum_{i,j} M(i,j)}$$

2. Innym rozwiązaniem jest normalizacja współczynników maski, tak by spełniały one warunek:

$$\sum_{i,j} M(i,j) = 1 \text{ i często dodatkowo } \Lambda_{i,j} M(i,j) \leq 1$$

Na przykład:

$$\begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix} \Rightarrow \begin{bmatrix} 1/9 & 1/9 & 1/9 \\ 1/9 & 1/9 & 1/9 \\ 1/9 & 1/9 & 1/9 \end{bmatrix} \quad \text{albo} \quad \begin{bmatrix} 0 & -2 & 0 \\ 1 & 3 & 1 \\ 0 & -2 & 0 \end{bmatrix}$$



# Filtry liniowe

- Obraz uzyskany w wyniku filtracji liniowej składa się z pikseli, które z kolei są kombinacją liniową pikseli obrazu wejściowego oraz ich otoczenia.
- Niech  $O(x, y)$  będzie obrazem wejściowym, zaś  $M(i, j)$  będzie maską (jądrem) filtru. Procedura filtracji przebiega w następujących krokach:
  1. Do każdego piksela obrazu wejściowego  $O(x, y)$  przykłada się maskę  $M(i, j)$
  2. Wartość piksela w obrazie wynikowym  $O'(x, y)$  oblicza się jako kombinację liniową pikseli z otoczenia piksela  $(x, y)$  i współczynnikami z macierzy  $M(i, j)$
  3. Na ogół procedura rozpoczyna się od lewego górnego rogu obrazu i kończy w prawym dolnym rogu.

$f(x - 1, y - 1)$	$f(x - 1, y)$	$f(x - 1, y + 1)$		$M(-1, -1)$	$M(-1, 0)$	$M(-1, 1)$		
$f(x, y - 1)$	$f(x, y)$	$f(x, y + 1)$		$M(0, -1)$	$M(0, 0)$	$M(0, 1)$		$O'(x, y)$
$f(x + 1, y - 1)$	$f(x + 1, y)$	$f(x, +1 y + 1)$		$M(1, -1)$	$M(1, 0)$	$M(1, 1)$		

# Elementy brzegowe

- W obliczaniu splotu obrazów problemem są piksele brzegowe, których kontekst jest „niepełny” i dla których w związku z tym nie można obliczyć odpowiadającej im wartości w obrazie wynikowym.
- Szerokość takiej „ramki” w obrazie zależy od rozmiaru maski i jest równa jej promieniowi.
- W niektórych przypadkach można nie liczyć wartości brzegowych. W pozostałych powiększa się obraz w obu kierunkach o promień maski a nowo powstały pikselom przypisuje się wartości na trzy alternatywne sposoby:
  - Wszystkim pikselom wykraczającym poza obszar obrazu oryginalnego przypisuje się wartość 0
  - Dodanym pikselom przypisuje się kopie wartości brzegowych
  - Nowym pikselom obrazu rozszerzonego  $O''$  przypisuje się wartości wynikające z „zawijania” obrazu, tj. wartości pikseli obrazu o współrzędnych określonych zależnością:

$$O''(X + p, Y + q) = O((X + p) \bmod X; (Y + q) \bmod Y)$$



# Filtry dolnoprzepustowe

- Filtr dolnoprzepustowy tłumia składowe widma o dużej częstotliwości, a pozostawia bez zmian składowe o małej częstotliwości.
- Typowym zastosowaniem filtrów dolnoprzepustowych jest wygładzanie obrazu oraz usuwanie szumów i zakłóceń impulsowych.
- Najprostszym filtrem dolnoprzepustowym jest filtr uśredniający z jądrem:

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

- Wartości pikseli w obrazie wynikowym są średnią arytmetyczną wartości pikseli z otoczenia.
- Filtr uśredniający usuwa drobne zakłócenia obrazu, wygładza niewielkie zawirowania krawędzi, usuwa efekty falowania jasności na obiektach i w tle.
- Wadą filtru jest:
  - Rozmycie krawędzi
  - Utrata ostrości
  - Zamazanie detali (szczególnie widoczne przy dużych maskach)
- Niepożądany efekt działania filtru jest tym wyraźniejszy im większy rozmiar ma jego maska.

# Filtr uśredniający – przykłady

we wszystkich przykładach wykorzystano maski o współczynnikach 1



# Filtr Gaussa

- Filtr Gaussa jest szczególnym rodzajem filtrów dolnoprzepustowych.
- Współczynniki maski filtru Gaussa są zwykle dyskretnym przybliżeniem funkcji gęstości dwuwymiarowego rozkładu normalnego:

$$f(x, y) = \frac{1}{2\pi\sigma^2} \exp\left(-\frac{x^2 + y^2}{2\sigma^2}\right)$$

- W filtrach gaussowskich największą wagę otrzymuje współczynnik odpowiadający elementowi centralnemu maski, a współczynniki mają tym mniejszą wartość im większa jest ich odległość od elementu centralnego.
- Rozmiar maski powinien uwzględniać wariancję funkcji Gaussa  $\sigma$ . Ogólnie prawdziwa jest zależność: im większa jest wariancja, tym większy rozmiar maski, ale też większe rozmycie w obrazie wynikowym.
- Przykładowe maski filtru Gaussa dla różnych wartości odchylenia standardowego  $\sigma$ :

$$\begin{bmatrix} 1 & 4 & 1 \\ 4 & 32 & 4 \\ 1 & 4 & 1 \end{bmatrix}$$

$$\begin{bmatrix} 1 & 3 & 4 & 3 & 1 \\ 3 & 12 & 19 & 12 & 3 \\ 4 & 19 & 32 & 19 & 4 \\ 3 & 12 & 19 & 12 & 3 \\ 1 & 3 & 4 & 3 & 1 \end{bmatrix}$$

$$\begin{bmatrix} 1 & 2 & 1 \\ 2 & 4 & 2 \\ 1 & 2 & 1 \end{bmatrix}$$

$$\begin{bmatrix} 1 & 2 & 4 & 2 & 1 \\ 2 & 4 & 8 & 4 & 2 \\ 4 & 8 & 16 & 8 & 4 \\ 2 & 4 & 8 & 4 & 2 \\ 1 & 2 & 4 & 2 & 1 \end{bmatrix}$$



# Filtr Gaussa - przykład

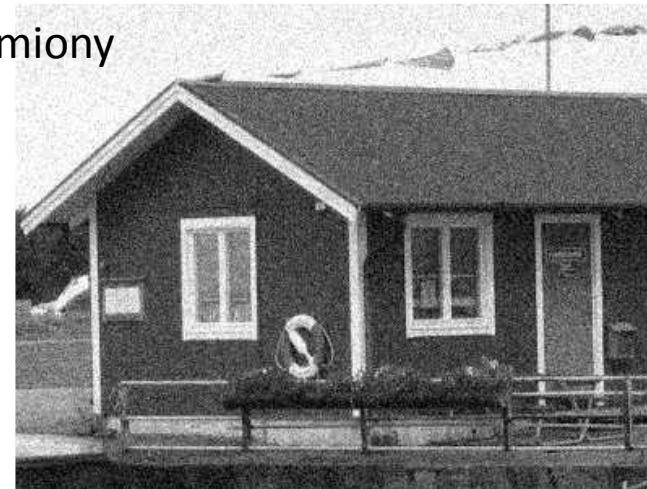
Obraz oryginalny



$$\begin{bmatrix} 1 & 2 & 1 \\ 2 & 4 & 2 \\ 1 & 2 & 1 \end{bmatrix}$$



Obraz zaszumiony



$$\begin{bmatrix} 1 & 2 & 4 & 2 & 1 \\ 2 & 4 & 8 & 4 & 2 \\ 4 & 8 & 16 & 8 & 4 \\ 2 & 4 & 8 & 4 & 2 \\ 1 & 2 & 4 & 2 & 1 \end{bmatrix}$$



Fundusze  
Europejskie  
Wiedza Edukacja Rozwój



Politechnika Wrocławska

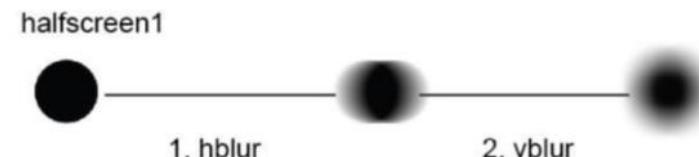


Unia Europejska  
Europejski Fundusz Społeczny

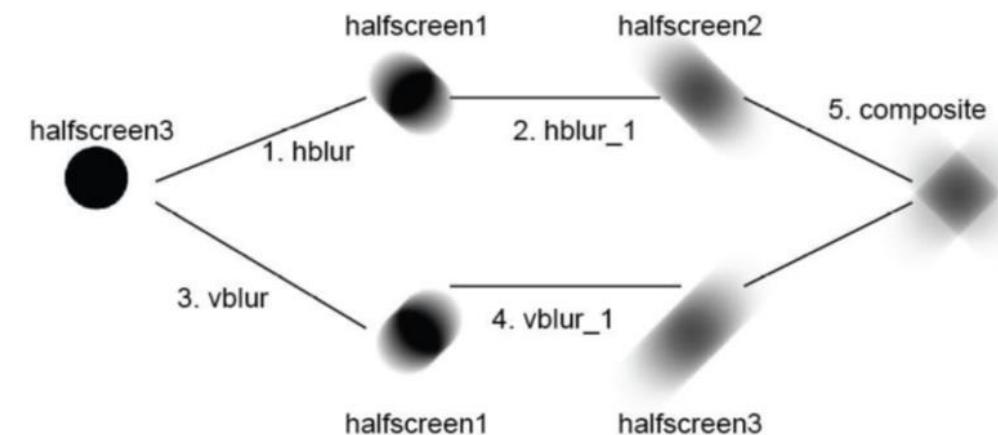
# Filtr Gaussa – rozmycie kierunkowe

- Jednym z efektów ubocznych stosowania filtru Gaussa rozmycie krawędzi (tzw. rozmycie gaussowskie).
- Symetryczność maski filtru powoduje, że obraz jest rozmywany we wszystkich kierunkach w przybliżeniu w takim samym stopniu (rys. A).
- Efekt ten można wykorzystać do tworzenia tzw. rozmycia kierunkowego w jednym zadanym kierunku (np. pionowo, poziomo, czy po przekątnych kwadratu).
- Sumując takie rozmycia można uzyskać ciekawe efekty (np. rozbłyski w kształcie gwiazdek – rys. B).

A. rozmycie gaussowskie



B. rozmycie w dwóch kierunkach



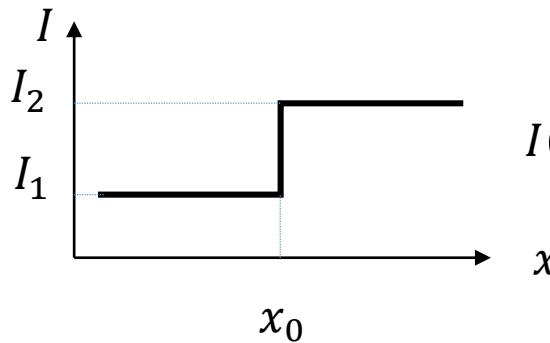
# Filtryle górnoprzepustowe

- Filtr górnoprzepustowy wzmacnia części obrazu o większych częstotliwościach i tłumia partie o małych częstotliwościach.
- Filtry tego typu używane są do podkreślenia elementów, które charakteryzują się dynamiczną zmianą jasności – kontury, krawędzie, drobne elementy faktury itp.
- Na obrazie wynikowym prawie wszystkie elementy należące do tła i wewnętrznych części obiektów zostają usunięte (wyzerowane), a punkty należące do krawędzi oraz granic między obiektem i tłem zostają odpowiednio wzmacnione.
- Zasada przeprowadzania obliczeń jest identyczna jak w filtrach dolnoprzepustowych. Różnica tkwi w innych wartościach współczynników maski (występują w nich również wartości ujemne).



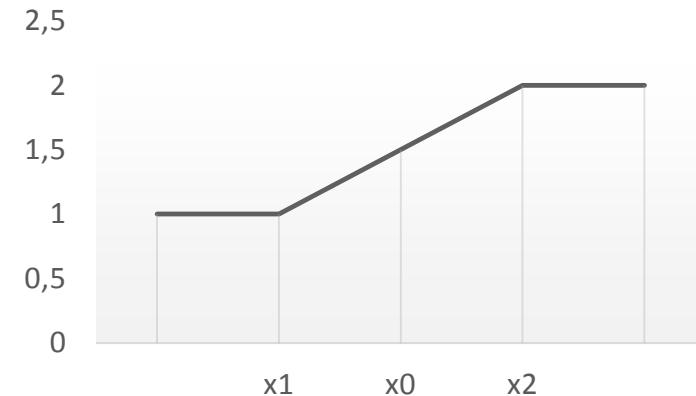
# Krawędzie w obrazie

- Krawędź jest granicą między obszarami obrazu o różnych odcieniach jasności.



$$I(x) = \begin{cases} I_1 & \text{jeśli } x < x_0 \\ I_2 & \text{jeśli } x \geq x_0 \end{cases}$$

- W rzeczywistych obrazach zmiana intensywności obrazu nie jest punktowa.
- Między jednym a drugim poziomem jasności występuje strefa przejściowa

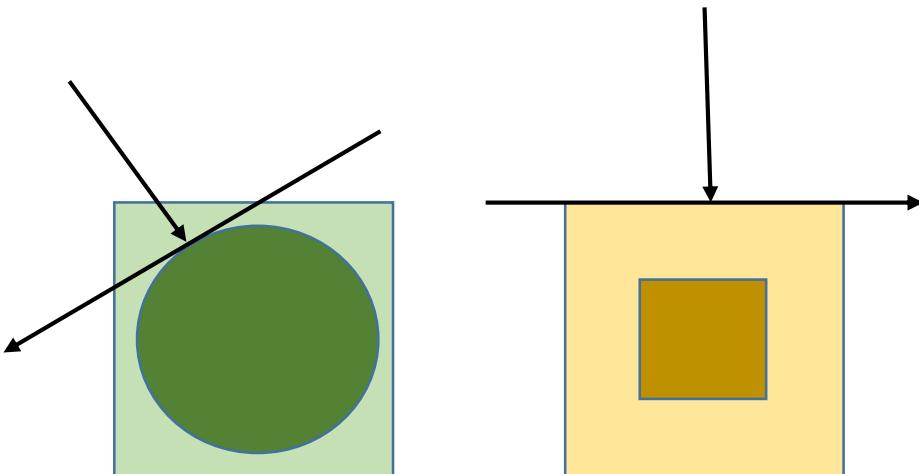


$$I(x) = \begin{cases} I_1 & \text{jeśli } x \leq x_1 \\ I_1 + (x - x_1) \frac{I_2 - I_1}{x_2 - x_1} & \text{jeśli } x_1 < x < x_2 \\ I_2 & \text{jeśli } x \geq x_2 \end{cases}$$



# Krawędzie dwuwymiarowe

- Krawędź dwuwymiarowa występuje na granicy dwóch obiektów albo obiektu i tła.
- Może być prosto- lub krzywoliniowa



- Na styku obiektu i tła (dwóch obiektów) można wyróżnić dwa wzajemnie prostopadłe kierunki:
  - Prostopadły do krawędzi – w tym kierunku jasność obrazu zmienia się gwałtownie, jak w przypadku krawędzi jednowymiarowej
  - Styczny do krawędzi – w tym kierunku jasność obrazu lokalnie pozostaje stała, tak jakby w tym miejscu nie było krawędzi

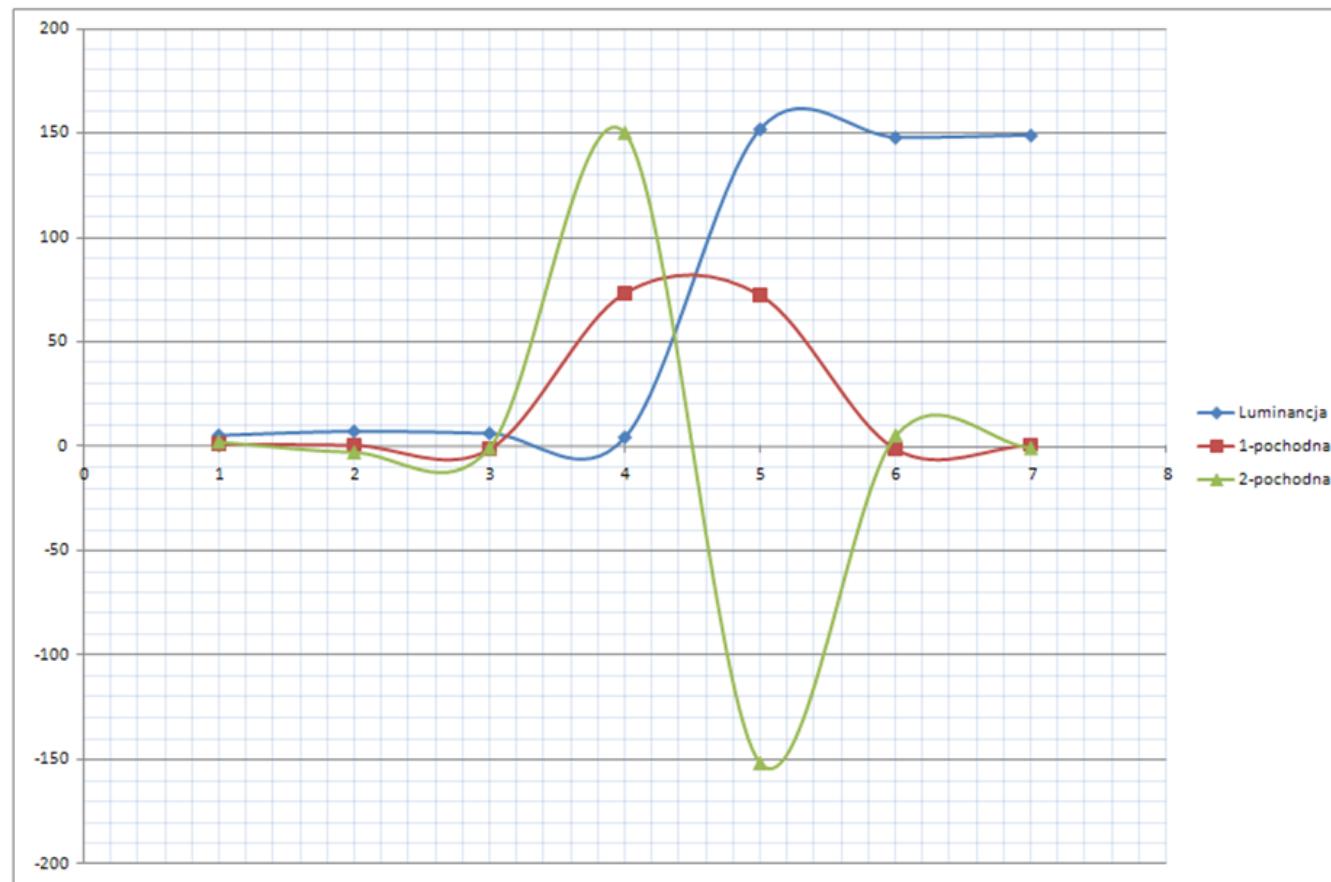
10 10 10 : 50 50 50 50	10 10 20 30 40 50 50
10 10 10 : 50 50 50 50	10 10 20 30 40 50 50
10 10 10 : 50 50 50 50	10 10 20 30 40 50 50
10 10 10 10 10 10 10	
... ... ... ... ... ...	
50 50 50 50 50 50 50	
50 50 50 50 50 50 50	

# Wykrywanie krawędzi

- Wykrywanie krawędzi polega na znalezieniu nagłych zmian intensywności w obrazie.
- Problemy w wykrywaniu krawędzi:
  - Zmiana intensywności nie powinna być punktowa, ponieważ podobne własności ma szum w obrazie (zwłaszcza szum impulsowy np. szum biały lub szum typu pieprz i sól)
  - W miejscu występowania krawędzi naga zmiana intensywności występuje tylko w jednym kierunku. W kierunku prostopadłym krawędź jest „niewidoczna”
- Trudno jest precyjnie zlokalizować krawędź, zwłaszcza jeśli w obrazie występuje strefa przejściowa.
  - W takich przypadkach za miejsce występowania krawędzi przyjmuje się:
  - Geometryczny środek zbocza
  - Przekroczenie wartości progowej, np.  $\frac{I_1 + I_2}{2}$  albo  $30\%(I_2 - I_1)$
  - Maksimum lokalne pierwszej pochodnej obrazu
  - Miejsce przejścia przez zero (zmiana znaku) drugiej pochodnej obrazu



# Wykres luminancji i pochodnych obrazu



Fundusze  
Europejskie  
Wiedza Edukacja Rozwój



Politechnika Wrocławska



Unia Europejska  
Europejski Fundusz Społeczny

# Gradient obrazu

- Z matematycznego punktu widzenia gradient  $\nabla f$  rzeczywistej funkcji wielu zmiennych  $f(x, y, \dots)$  jest wektorem, który zawiera wszystkie pierwsze pochodne cząstkowe tej funkcji:

$$\nabla f = \begin{bmatrix} \frac{\partial f}{\partial x} \\ \frac{\partial f}{\partial y} \\ \vdots \end{bmatrix}$$

- Wartościami funkcji  $\nabla f$  są wektory.
- W przestrzeni, w której określona jest funkcja  $f$  w dowolnym punkcie o współrzędnych  $(x, y, \dots)$  zwrot wektora  $\nabla f(x, y, \dots)$  wskazuje kierunek najszybszego wzrostu funkcji.
- Ponadto, wektor gradientu  $\nabla f(x, y, \dots)$  jest prostopadły do konturu stałej wartości funkcji  $f$ , który przechodzi przez punkt  $(x, y, \dots)$

# Gradient obrazu

- Ponieważ krawędź wyznacza gwałtowna zmiana wartości funkcji jasności, to do jej wyodrębniania z obrazu często stosuje się metody gradientowe.



# Techniki przetwarzania mediów cyfrowych

Wykład 5: Morfologia matematyczna  
w przetwarzaniu obrazów

# Plan wykładu

- Wprowadzenie
- Operacje morfologiczne na obrazach:
  - Dylatacja,
  - Erozja,
  - Otwarcie i zamknięcie morfologiczne,
  - Transformacja „Hit and Miss” (HMT).
- Wybrane algorytmy morfologiczne.
  - Ekstrakcja konturu,
  - Szkieletyzacja,
  - Ścienianie i pogrubianie,
  - Wykrywanie narożników.

# Wprowadzenie

- **Morfologia** jest nauką o formie i strukturze.
- **Morfologia matematyczna** jest teorią i techniką stosowaną do analizy i przetwarzania struktur geometrycznych wykorzystującą teorię zbiorów, teorię krat, topologię i funkcje losowe.
- Pozwala na zdefiniowanie w przestrzeniach ciągłych i dyskretnych takich pojęć topologicznych i geometrycznych, jak rozmiar, kształt, wypukłość, spójność czy odległość.
- Morfologia matematyczna rozwinęła się na podstawie prac Hermanna Minkowskiego z początku XX wieku.
- Podstawowe koncepcje i operacje morfologiczne na obrazie zdefiniowali Georges Matheron i Jean Serre, którzy sformułowali podstawowe koncepcje i zdefiniowali morfologiczne operacje na obrazie.
- Zalety morfologii matematycznej: prostota zapisu operacji na obrazie, proste sposoby wykonania tych operacji.
- Morfologia znalazła szerokie zastosowanie zwłaszcza w dziedzinie analizy kształtu i badaniu różnych obszarów obrazu.

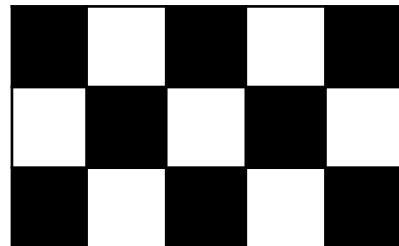
# Wprowadzenie

- Operacje morfologiczne najprościej wyjaśnić na obrazach binarnych, chociaż mają one zastosowanie również dla obrazów z gradacją kontrastu i obrazów kolorowych.
- Z założenia obraz binarny przedstawia czarny obraz (obiekt) na białym tle.
- **Obiekt** ten (a tym samym i cały obraz) można przedstawić jako zbiór zawierający wszystkie piksele o wartości „1” (piksele czarne).
- Taka reprezentacja pozwala wykonać na obrazie operacje znane z teorii zbiorów: dodawanie, odejmowanie, przekrój, zawieranie się zbiorów itp.

# Wprowadzenie

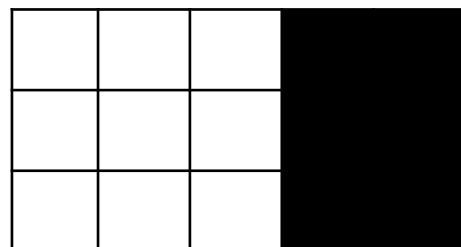
**Przykład:**

Obraz binarny A



$$A = \{(0; 2), (2; 2), (4; 2), (1; 1), (3; 1), (0; 0), (2; 0), (4; 0)\}$$

Obraz binarny Z



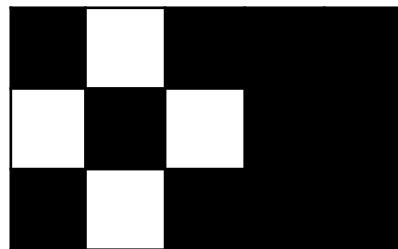
$$Z = \{(3; 2), (4; 2), (3; 1), (4; 1), (3; 0), (4; 0)\}$$

# Wprowadzenie

## Przykład c.d.

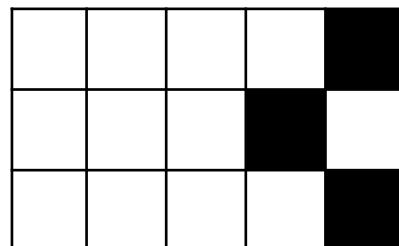
Suma obrazu A i Z

$$A \cup Z = \{(0; 2), (2; 2), (3; 2), (4; 2), (1; 1), (3; 1), (4; 1), (0; 0), (2; 0), (3; 0), (4; 0)\}$$



Przecięcie (iloczyn) obrazów A i Z

$$A \cap Z = \{(4; 2), (3; 1)(4; 0)\}$$



# Wprowadzenie

- Współczesna morfologia matematyczna bazuje na geometrii całkowej Minkowskiego.
- W podstawowych operacjach morfologicznych na obrazach wykorzystuje się operacje dodawanie i odejmowania Minkowskiego zdefiniowane następująco:
  - Suma Minkowskiego:

$$A \oplus B = \{a + b : a \in A, b \in B\}$$

- Różnica Minkowskiego:

$$A \ominus B = (A' \oplus B^*)'$$

Gdzie:

$$A' = \{a' : a' \notin A\}$$

- dopełnienie zbioru  $A$ ,

$$B^* = \{-b : b \in B\}$$

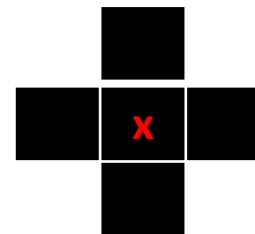
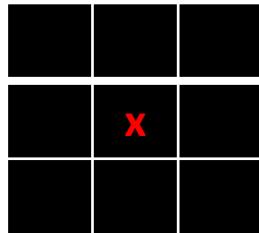
- odbicie zbioru  $B$  względem środka układu współrzędnych.

# Wprowadzenie

- Podstawowym pojęciem przekształceń morfologicznych jest **element strukturalny obrazu (wzorzec)**.
- Element strukturalny to pewien wycinek obrazu (w reprezentacji dyskretnej pewien podzbiór jego elementów) z wyróżnionym jednym punktem, zwanym **elementem centralnym (punktem centralny)**.
- Aby zdefiniować konkretny element strukturalny należy określić jego kształt oraz położenie punktu centralnego.
- W rozważaniach teoretycznych najczęściej stosowanym elementem strukturalnym jest koło o promieniu jednostkowym.
- W przetwarzaniu obrazów cyfrowych element strukturalny pełni rolę podobną do maski w operacji konwolucji.

# Wprowadzenie

- W rzeczywistych zastosowaniach kształt elementu strukturalnego wynika z siatki kwadratowej wykorzystanej do reprezentacji obrazu.
- Często (ale nie zawsze) elementem strukturalnym jest kwadrat o boku  $2n+1$  i wszystkich wartościach równych „1” (piksele czarne = obiekt).
- Punkt centralny może (ale nie musi) leżeć w środku elementu strukturalnego



# Wprowadzenie

- Ogólny algorytm przekształcenia morfologicznego przebiega następująco:
  1. Przyóż punkt centralny elementu strukturalnego do pierwszego punktu obrazu,
  2. Sprawdź czy lokalna konfiguracja punktów obrazu (obszar pokryty elementem strukturalnym) odpowiada układowi zapisanemu w elemencie strukturalnym,
  3. Jeśli tak, to wykonaj operację związaną z danym przekształceniem.
  4. W przeciwnym przypadku, przesuń element strukturalny do kolejnego piksela i przejdź do kroku 2
- Przesuwanie elementu strukturalnego o obrazie odbywa się po wierszach i kolumnach, począwszy od lewego górnego rogu

# Wprowadzenie

- Przekształcenia morfologiczne różnią się od innych operacji na obrazach cyfrowych tym, że:
  - Punkty obrazu źródłowego przekształcane są w taki sam sposób, niezależnie od wartości pikseli sąsiednich (inaczej niż w operacjach kontekstowych, np. w filtrach medianowych).
  - Przekształcane są tylko te punkty obrazu, których otoczenie jest zgodne z elementem strukturalnym (w filtrach przelicza się wszystkie punkty, nawet gdy operacje filtracji nie zmieniają wartości pikseli).
- Przekształcenia morfologiczne pozwalają na dokonywanie bardziej subtelnych przekształceń obrazów.

# Operacja dylacji (dylatacji)

- Dylacja (dylatacja) jest operacją dwuargumentową wykorzystującą sumę Minkowskiego
  - Pierwszym argumentem jest obraz
  - Drugim – element strukturalny
- Algorytm przekształcenia:
  - Do każdego punktu obrazu przykłada się punkt centralny elementu strukturalnego
  - Jeżeli choć jeden piksel z otoczenia pokrytego elementem strukturalnym ma kolor obiektu (czarny  $\equiv 1$ ), to zmień wartość bieżącego piksela na 1

# Operacja dylacji (dylatacji)

## Własności:

- Obiekt zwiększa rozmiar i staje się wyraźniejszy
- Zanikają detale obrazu oraz „dziury”
- Kształt obiektu na obrazie wynikowym zależy od kształtu elementu strukturalnego, jego rozmiaru oraz położenia punktu centralnego
- Kształt najbardziej zbliżony do kształtu oryginalnego obiektu uzyskuje się stosując element strukturalny zbliżony o koła

# Operacja erozji

- Erozja jest operacją dwuargumentową wykorzystującą różnicę Minkowskiego
  - Pierwszym argumentem jest obraz
  - Drugim – element strukturalny
- Algorytm przekształcenia:
  - Do każdego punktu obrazu przykłada się punkt centralny elementu strukturalnego
  - Jeżeli choć jeden piksel z otoczenia pokrytego elementem strukturalnym ma kolor tła (biały  $\equiv 0$ ), to zmień wartość bieżącego piksela na 0

# Operacja erozji

## Właściwości:

- Obiekt zmniejsza rozmiar i znika (przy erozji wielokrotnej)
- Zanikają detale obrazu, ale w inny sposób niż w dylatacji
- „Dziury” powiększają się
- Kształt obiektu na obrazie wynikowym zależy od kształtu elementu strukturalnego, jego rozmiaru oraz położenia punktu centralnego
- Kształt najbardziej zbliżony do kształtu oryginalnego obiektu uzyskuje się stosując element strukturalny zbliżony o koła

# Zastosowania dylatacji i erozji

- Wyodrębnianie granic obiektu w obrazie oryginalnym  $O$ :
  - Granica „zewnętrzna” obiektu =  $D(O) - O$ , gdzie  $D(O)$  – dylataция obrazu  $O$
  - Granica „wewnętrzna” obiektu =  $O - E(O)$ , gdzie  $E(O)$  – erozja obrazu  $O$

# Cyfrowe Przetwarzanie Obrazów

## Digital Image Processing

### Wykład 1



Aurelio A. Heckert

Copyright (c) 2008 Cezary Bołdak

Permission is granted to copy, distribute and/or modify this document under the terms of the GNU Free Documentation License, Version 1.2 or any later version published by the Free Software Foundation; with no Invariant Sections, no Front-Cover Texts, and no Back-Cover Texts. A copy of the license is included in the section entitled "GNU Free Documentation License"



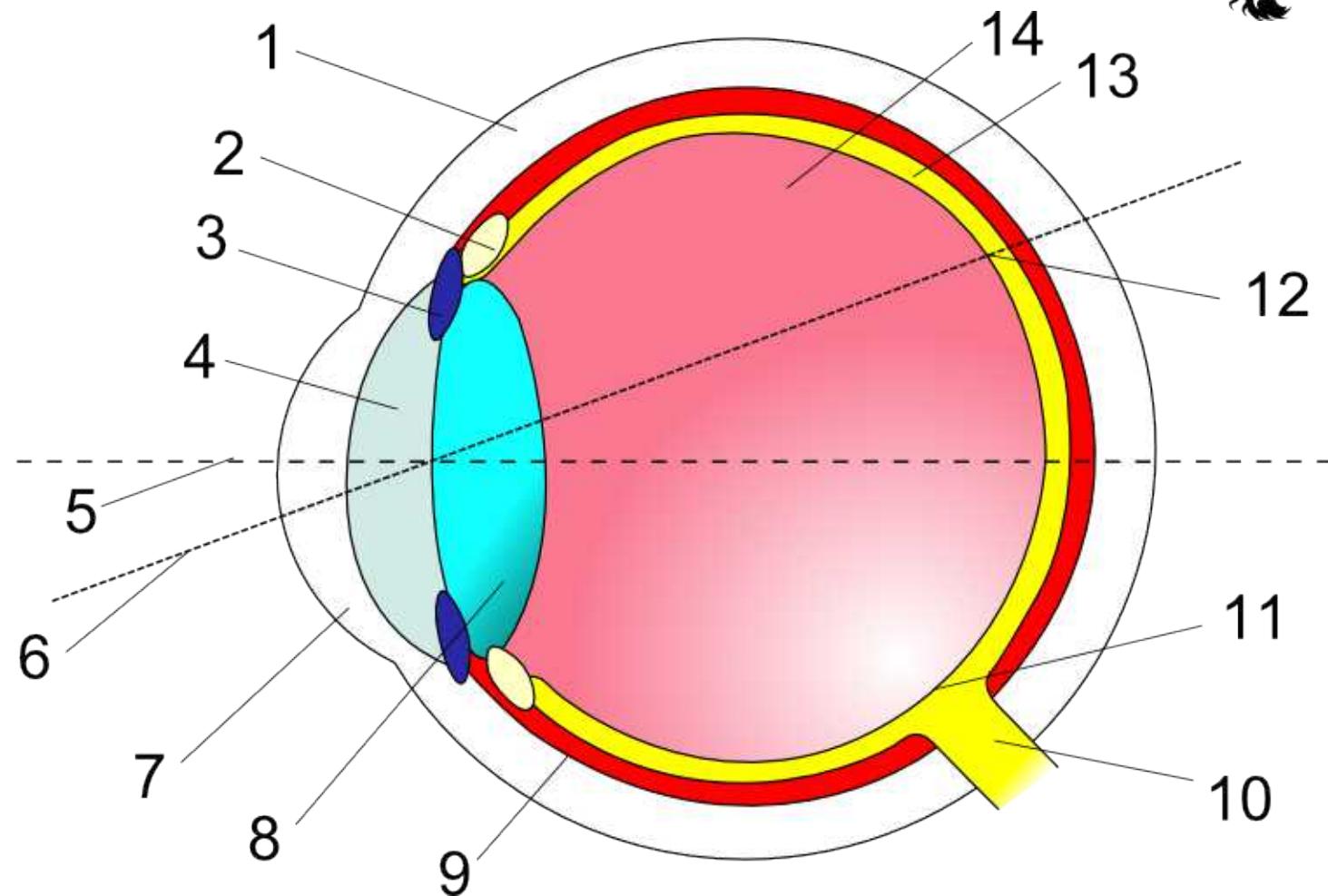
## Tematyka wykładu

- postrzeganie obrazu przez człowieka
- obraz cyfrowy - reprezentacja, pozyskiwanie
- zmiana parametrów obrazu (jasność, nasycenie, kontrast)
- cyfrowa filtracja obrazu (konwolucja)
- przetwarzanie obrazu w dziedzinie częstotliwości, FFT
- wyodrębnianie krawędzi
- morfologia matematyczna
- segmentacja obrazów
- kształty w obrazie (momenty geometryczne, transformata Hough'a , SSA)
- tekstury w obrazach
- zmiana rozdzielczości obrazów
- wyrównywanie obrazów, przekształcenia geometryczne
- detekcja i rozpoznawanie obiektów
- kompresja obrazów
- szum w obrazach

# Budowa oka człowieka



- 1 - twardówka
- 2 - ciało rzęskowe
- 3 - tęczówka
- 4 - ciecz wodnistą
- 5 - oś optyczna
- 6 - oś widzenia
- 7 - rogówka
- 8 - soczewka
- 9 - naczyniówka
- 10 - nerw wzrokowy
- 11 - plamka ślepa
- 12 - dołek środkowy  
(plamka żółta)
- 13 - siatkówka
- 14 - ciało szkliste

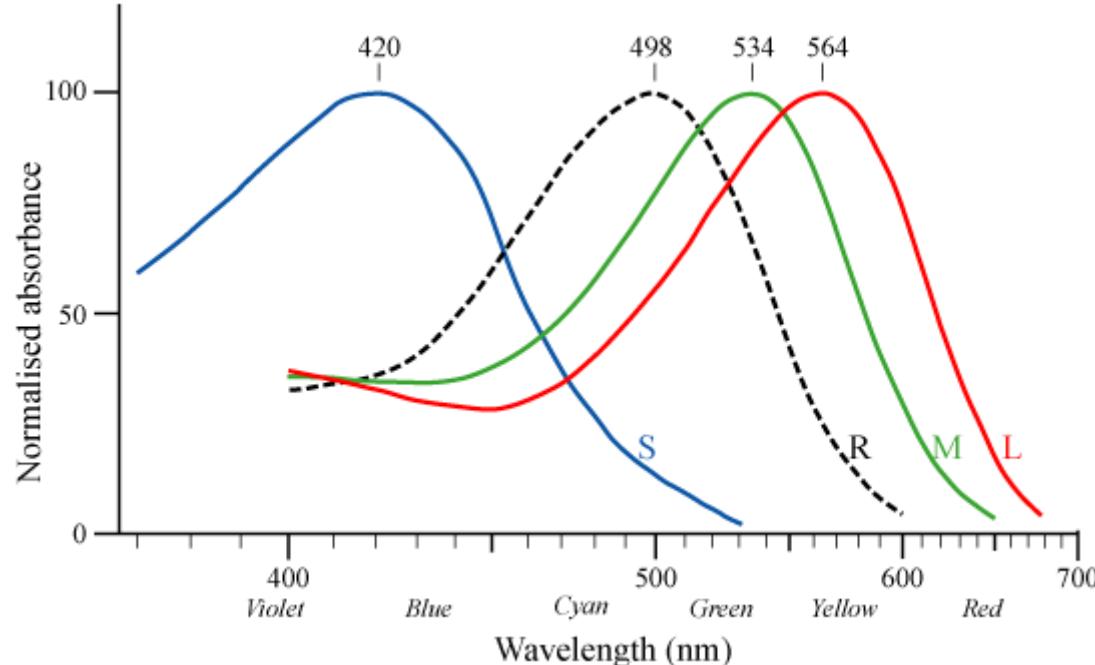


autor: MesserWoland, Wikipedia.pl



## Siatkówka (ang. retina)

- czopki (ang. cones)
  - 3 rodzaje, reagujące na 3 długości fal, ok. 6 mln, skupione wokół żółtej plamki, niewielka czułość
- pręciki (ang. rods)
  - 1 rodzaj, reakcja na 1 długość fal, ok. 100 mln, na całej siatkówce, duża czułość nawet na słabe światło
- inspiracja dla modelu RGB



Absorbcja częstotliwości przez 3 rodzaje czopków (S, M L) i pręciki

Bowmaker J.K. and Dartnall H.J.A., Visual pigments of rods and cones in a human retina; J. Physiol 298: pp.501-51

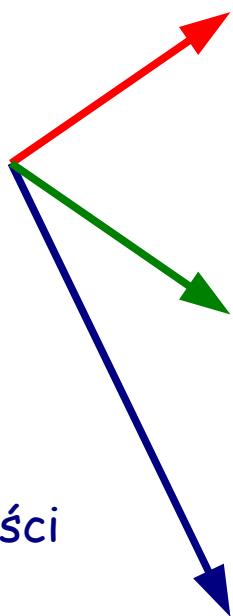
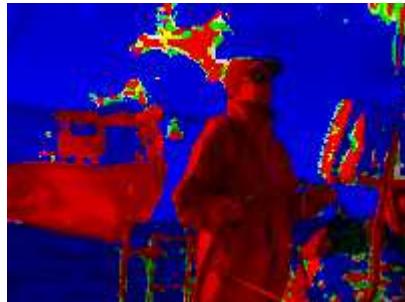


## Model koloru RGB

Każdy kolor jest reprezentowany przez 3 składowe:  
czerwoną (R), zieloną (G), niebieską (B)



zachowany kanał  
o maksymalnej wartości



lub



lub



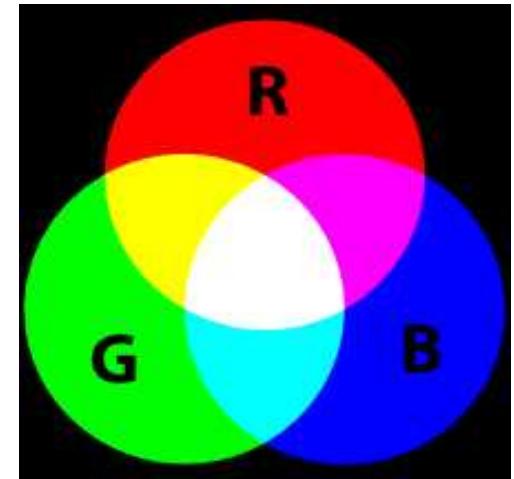
lub



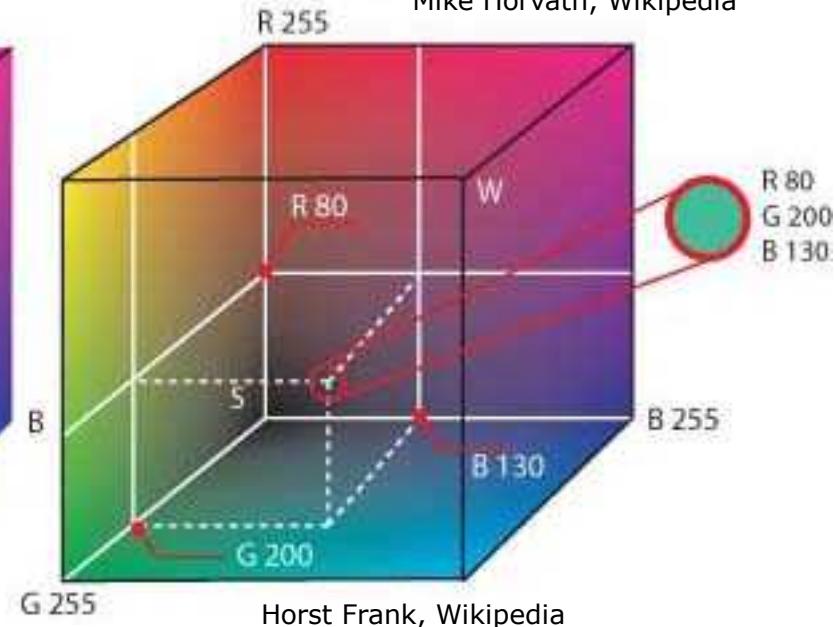
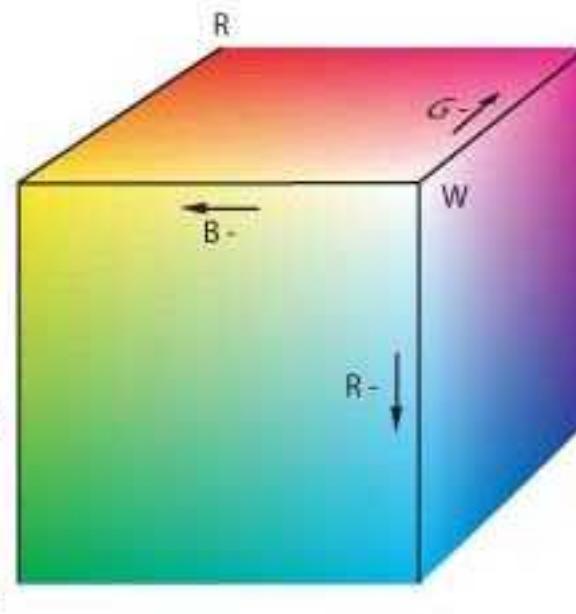
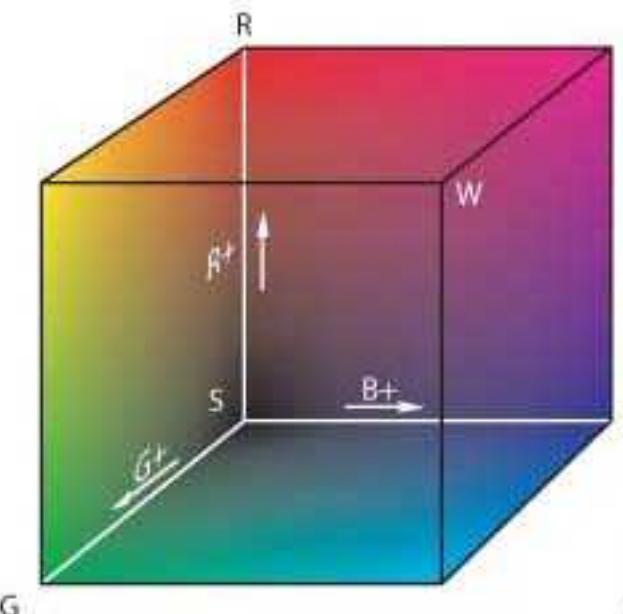


## Model koloru RGB

- inne kolory tworzą się na zasadzie dodawania (model addytywny)
- liczbę możliwych kolorów określa poziom zmienności każdej ze składowych
  - 16 bitów (5-6-5) – 65536 kolorów
  - 24 bity (8-8-8) – ok. 16,7 mln kolorów
  - 32 bity, jak wyżej, 8 bitów nieużywane lub kanał alfa
  - 48 bitów (16-16-16), zastosowania profesjonalne



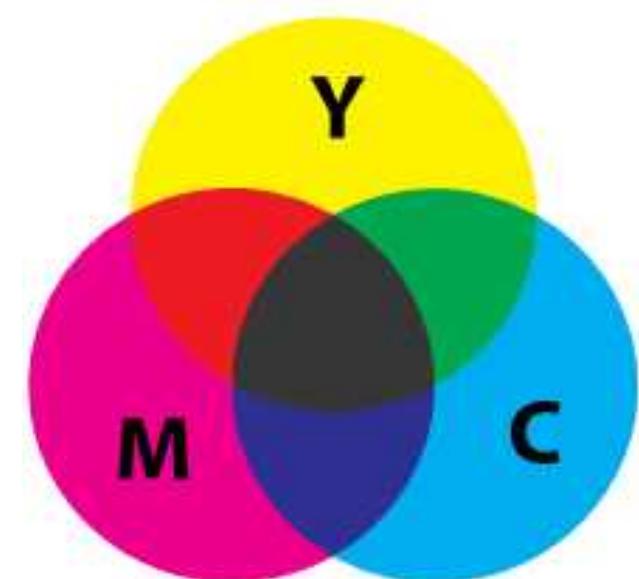
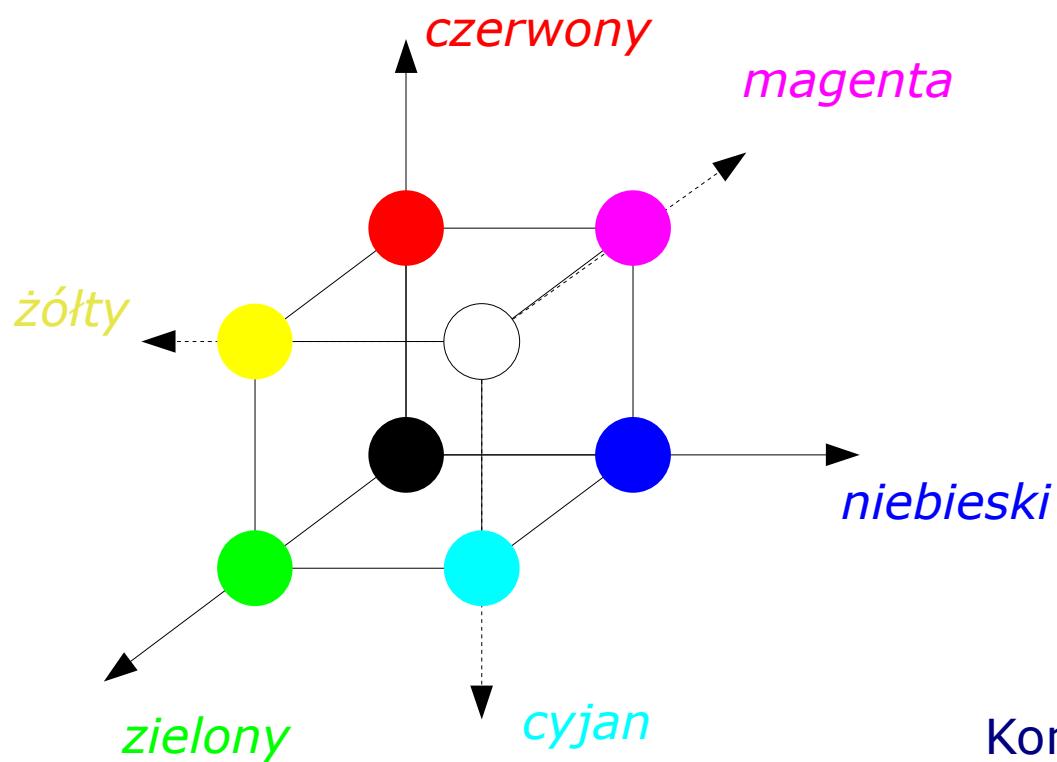
Mike Horvath, Wikipedia





## Model koloru CMY(K)

- inne kolory tworzą się na zasadzie odejmowania (model substrakcyjny)
- 4 kolor (czarny) dodany ze względów praktycznych
- używany powszechnie w poligrafii



Mike Horvath, Wikipedia

Konwersja:

$$C = 1 - R$$

$$M = 1 - G$$

$$Y = 1 - B$$

$$R = 1 - C$$

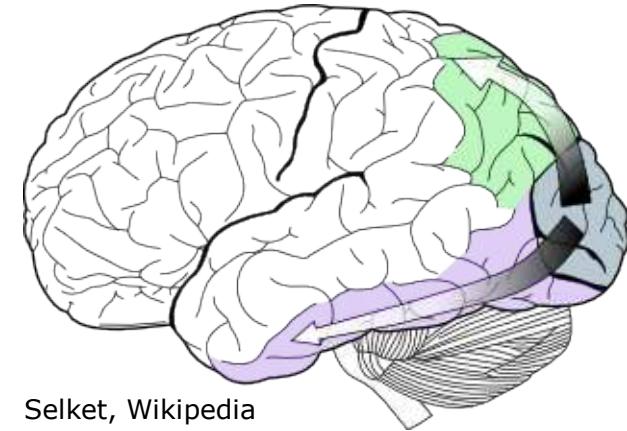
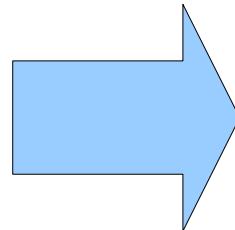
$$G = 1 - M$$

$$B = 1 - Y$$

# Percepcja koloru przez człowieka



Benjamin Crowell



Selket, Wikipedia

R



Luminancja

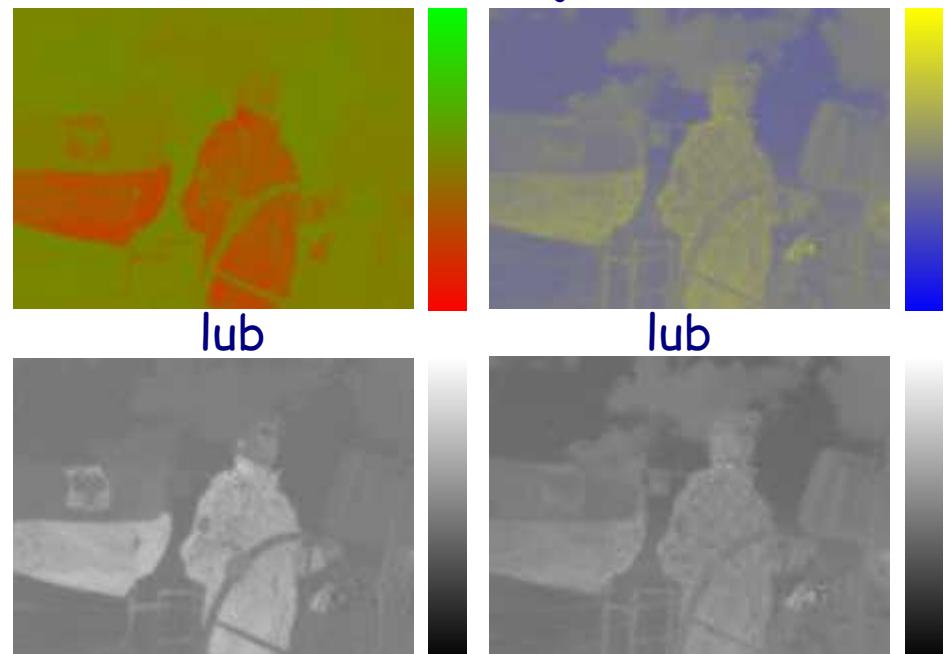
G



B



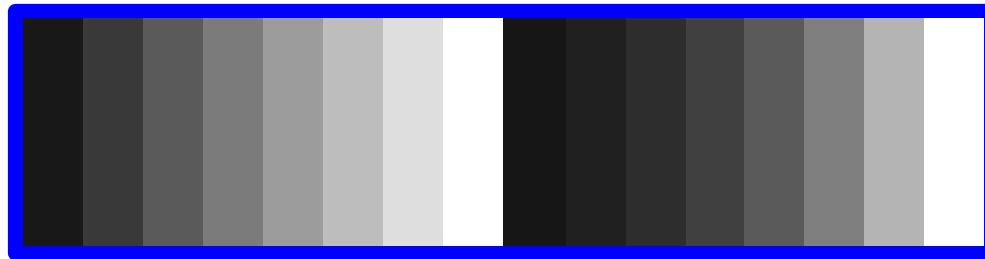
Chrominancja





## Percepcja jasności przez człowieka

Prawo Webera-Fechnera – człowiek dostrzega różnicę kolorów na podstawie **stosunku** ich intensywności a nie **różnicy**

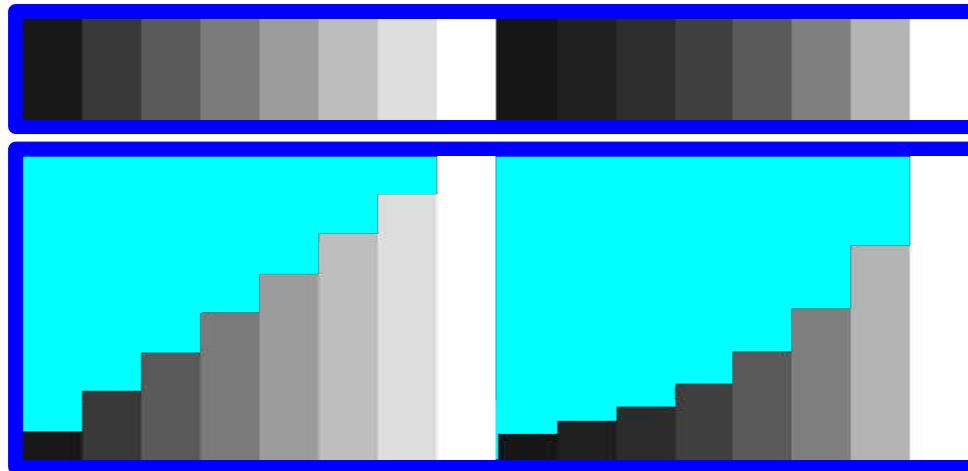


Która skala jest liniowa, a która logarytmiczna ?



## Percepcja jasności przez człowieka

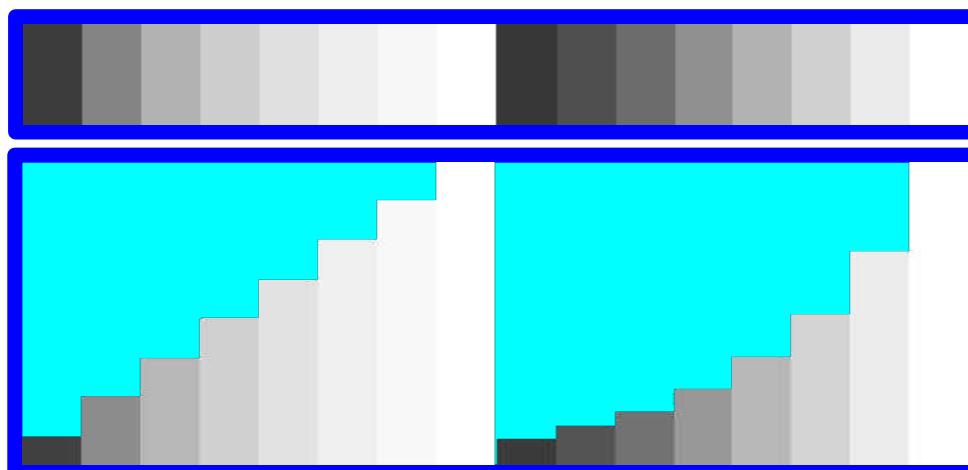
Po lewej skala liniowa !!! A po prawej logarytmiczna !!!



???

???

Trzeba pamiętać o korekcji gamma monitorów !!!  
Oto obraz po odwróceniu korekcji gamma monitora





## Pułapki w percepcji jasności (1)

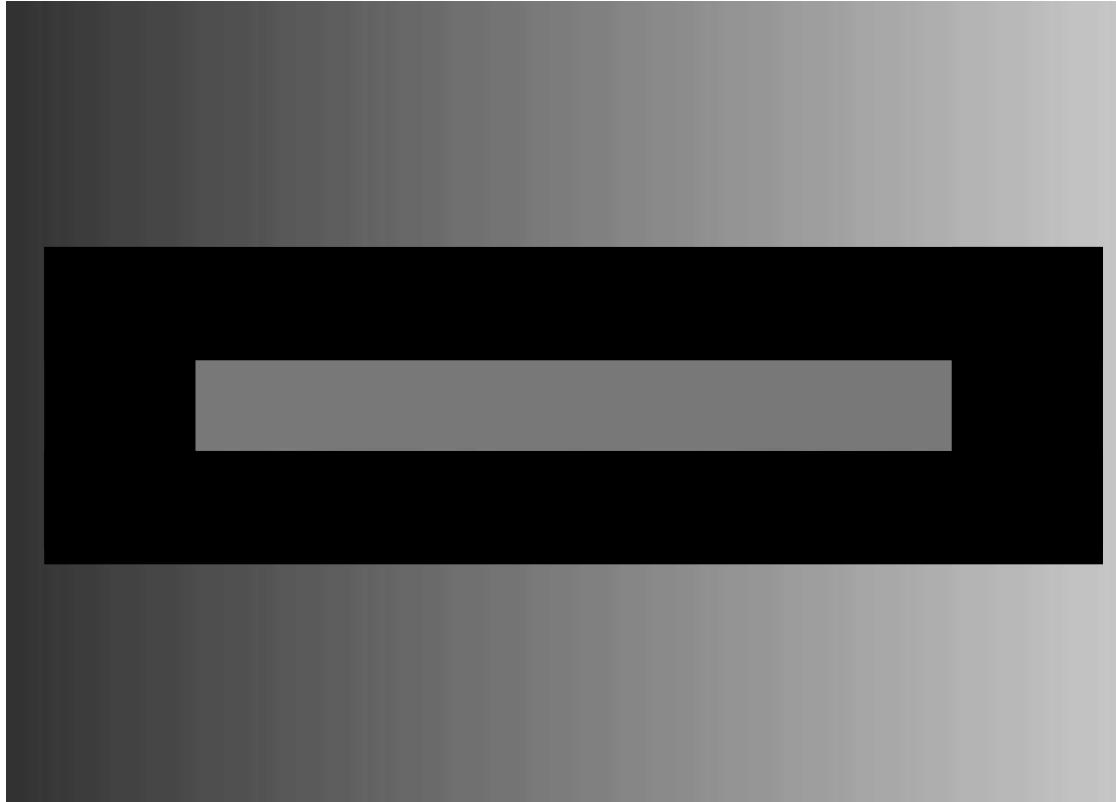


Dodek, Wikipedia

Czy wewnętrzny prostokąt zmienia jasność ?



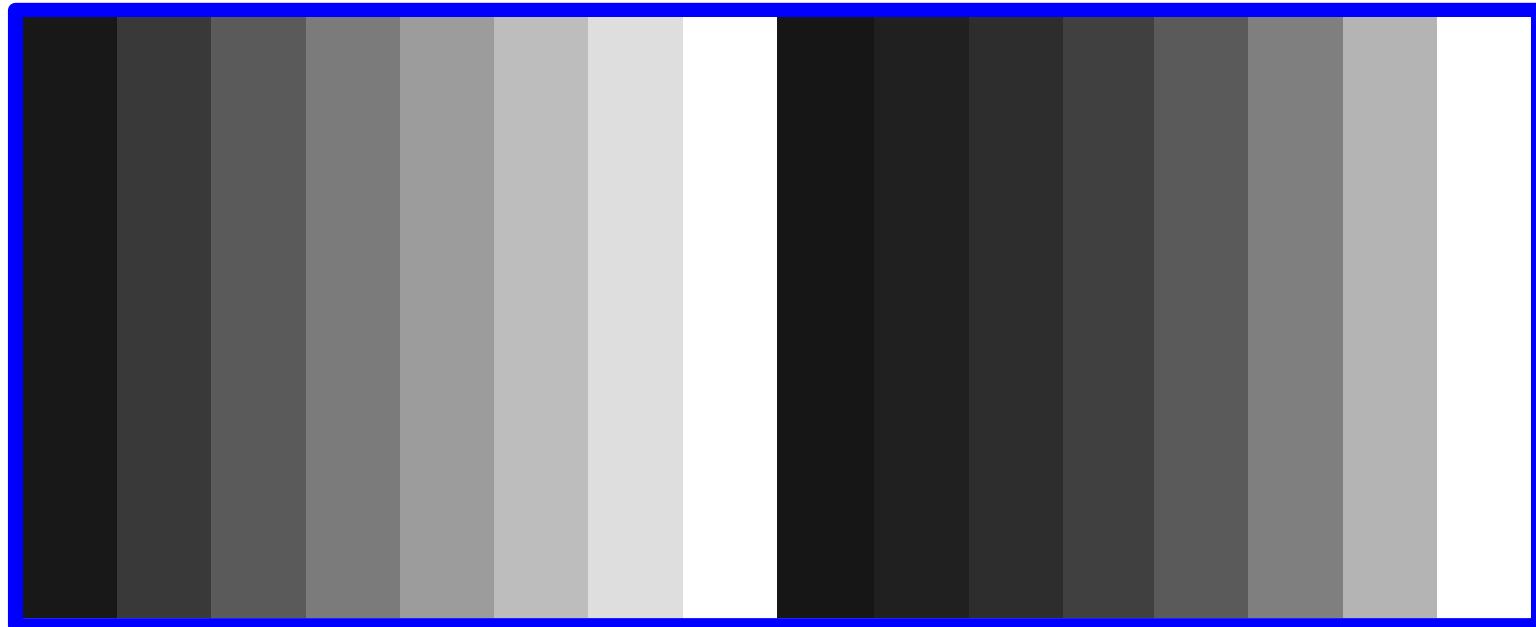
## Pułapki w percepcji jasności (1)



Mózg postrzega jasność w kontekście otoczenia



## Pułapki w percepcji jasności (2)



Granica jasności stwarza iluzję jasnych pasków po lewej stronie pasków i ciemnych pasków po prawej



## Pułapki w percepcji jasności (3)



Actual luminance distribution



Perceived luminance distribution

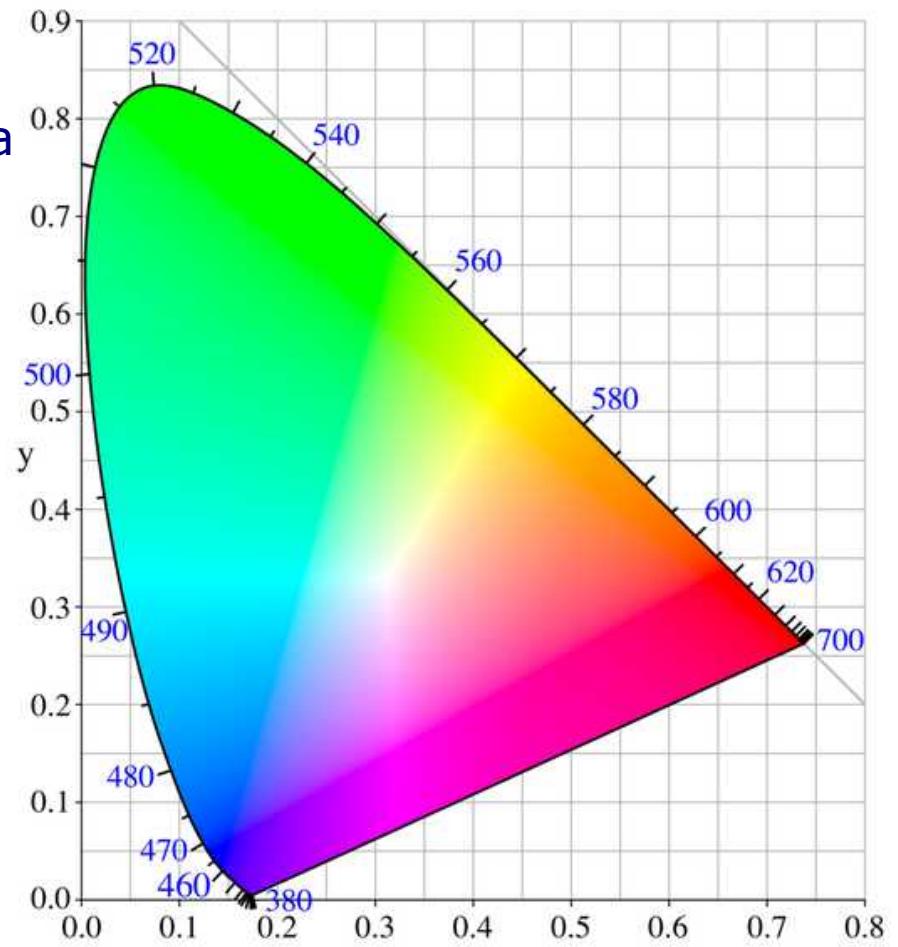


Efekt łańcu (ang. *cornsweet*)



## Model koloru CIE XYZ 1931

- model odniesienia stworzony przez Commission Internationale de l'Éclairage
- oparty na percepcji kolorów przez człowieka
- wywiedziony z modelu RGB, zmodyfikowany eksperymentalnie
- model 3D, współrzędne X, Y, Z „z grubsza” odpowiadają R, G, B
- współrzędna Y odpowiada luminancji
- znormalizowane współrzędne x i y odpowiadają chrominancji  
 $x=X/(X+Y+Z)$   $y=Y/(X+YZ)$
- często modyfikuje się model na CIE xyY
- obwód wykresu odpowiada światłu monochromatycznemu o podanej długości fali
- prosta część obwodu nie odpowiada rzeczywistej barwie



PAR, Wikipedia



## Model koloru HSV (HSB)

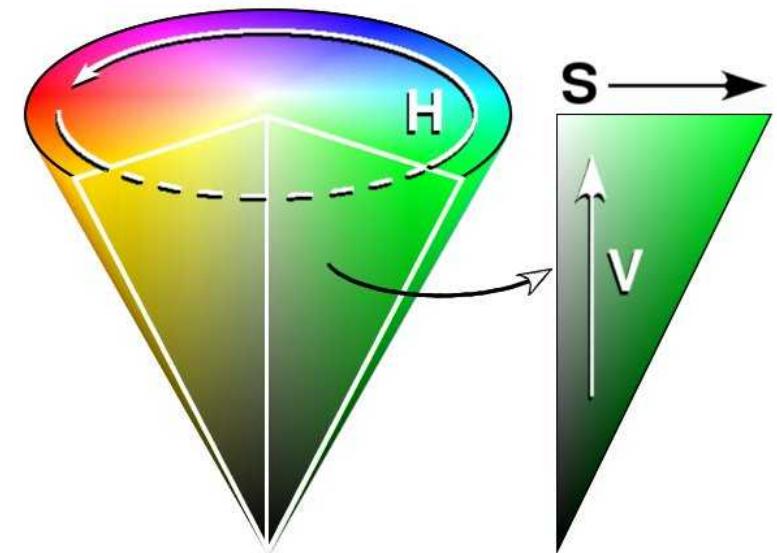
- przestrzeń kolorów jest stożkiem
- Hue* – barwa, liczona w stopniach (0, 360)
- Saturation* – nasycenie (0,1)
- Value* lub *Brightness* – jasność, wartość (0,1)

Konwersja ( $r,g,b$  w zakresie  $[0,1]$ ):

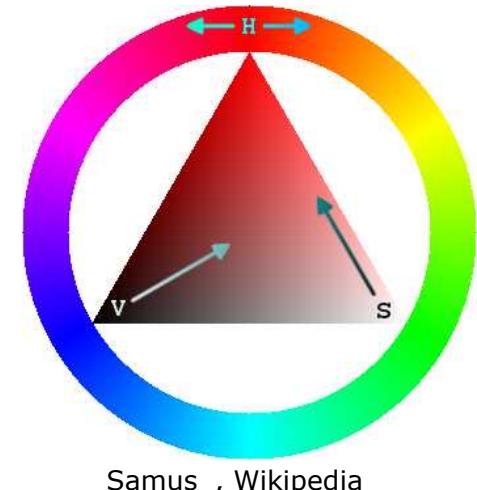
$$\max = \max(r, g, b)$$

$$\min = \min(r, g, b)$$

$$H = \begin{cases} \text{niezdefiniowany} & \text{jeśli } \max = \min \\ 60 \cdot \frac{g-b}{\max - \min} + 0^\circ & \text{jeśli } \max = r \text{ i } g \geq b \\ 60 \cdot \frac{g-b}{\max - \min} + 360^\circ & \text{jeśli } \max = r \text{ i } g \leq b \\ 60 \cdot \frac{b-r}{\max - \min} + 120^\circ & \text{jeśli } \max = g \\ 60 \cdot \frac{r-g}{\max - \min} + 240^\circ & \text{jeśli } \max = b \end{cases}$$
$$S = \begin{cases} 0 & \text{jeśli } \max = 0 \\ \frac{\max - \min}{\max} & \text{w przeciwnym wypadku} \end{cases}$$
$$V = \max$$



Wapcaplet, Wikipedia



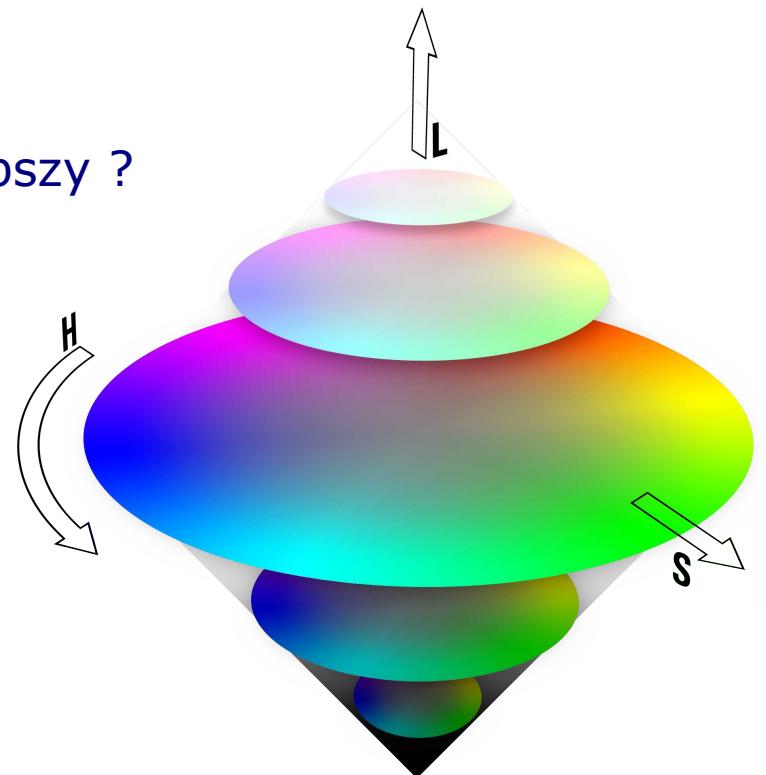
Samus\_, Wikipedia



## Model koloru HSL (HSI)

- przestrzeń kolorów jest bryłą złożoną z 2 stożków
- ta sama wartość barwy *Hue*
- inne sformułowanie nasycenia *Saturation*
- *Lightness* – średnie światło białe
- konkurencja do HSV, ale który model jest lepszy ?

$$S = \begin{cases} 0 & \text{jeśli } L=0 \text{ lub } \max=\min \\ \frac{\max-\min}{2L} & \text{jeśli } 0 < L \leq \frac{1}{2} \\ \frac{\max-\min}{2-2L} & \text{jeśli } L > \frac{1}{2} \end{cases}$$
$$L = \frac{\max+\min}{2}$$



Alexandre Van de Sande, Wikipedia

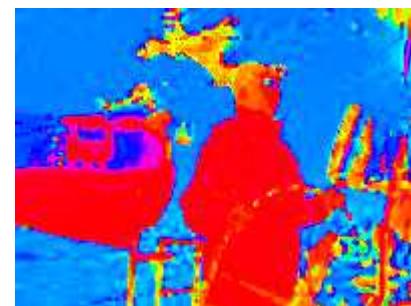


## HSL kontra HSV

- kwestia gustu, upodobań
- różne aplikacje używają jednego z nich
- lub obu alternatywnie



Hue



HSL

Saturation



Lightness/Intensity



Saturation



Value/Brightness



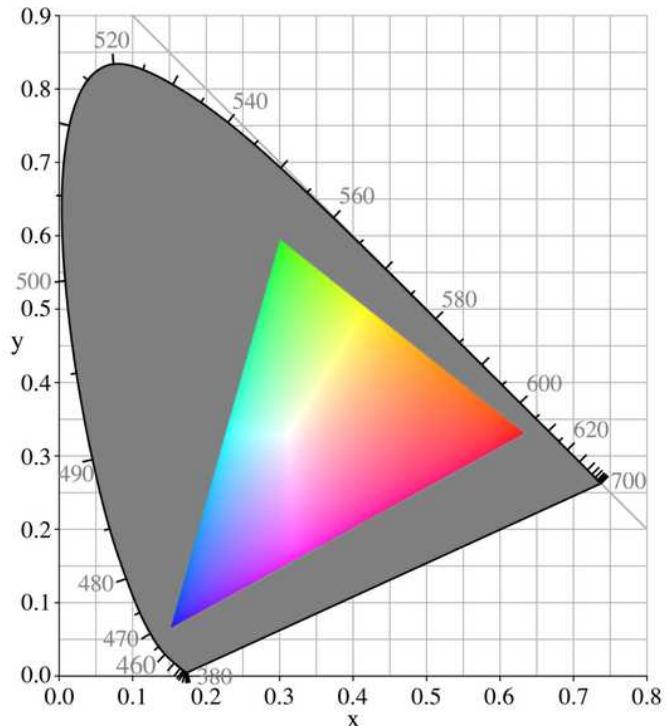
HSV

- biały kolor – nasycenie 0, kolor bez znaczenia
- *Value* jaśniejsza niż *Lightness*
- kanał *Hue* ma nasycenie = 1,
- *Saturation* w HSL ma mniejszy kontrast niż *Saturation* w HSV



## Gamut

- w przestrzeni chrominancji CIE 1931 XYZ dowolne 3 punkty pozwalają zbudować poprzez ich zmieszanie dowolny kolor z trójkąta
- trójkąt ten nosi nazwę *gamut*
- żadne urządzenie cyfrowe nie może oddać wszystkich kolorów widzianych przez człowieka



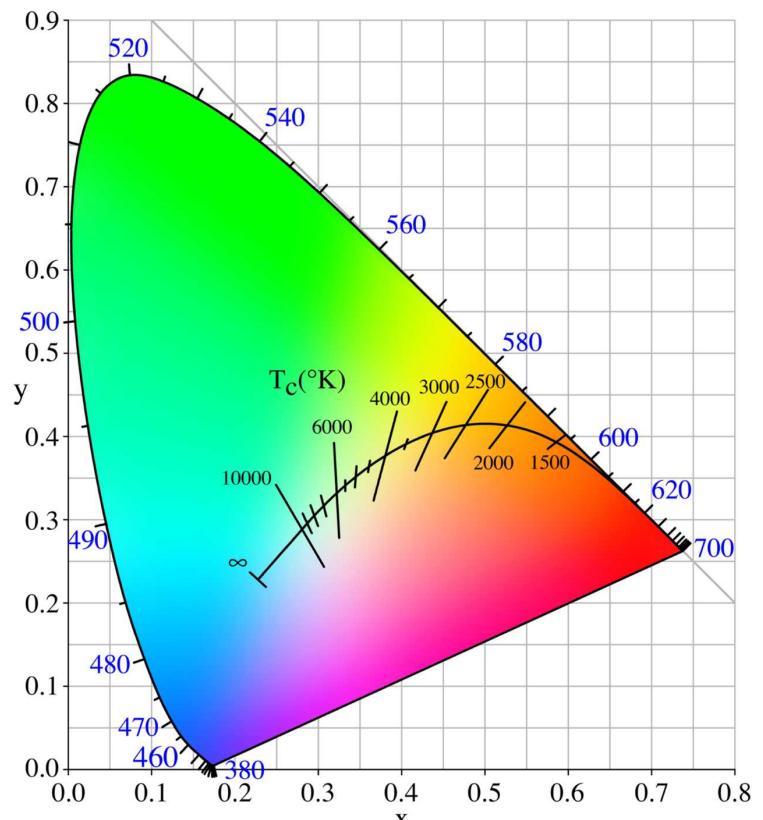
**Gamut monitora CRT**

Hankwank, Wikipedia



## Temperatura koloru

- źródło światła emitujące światło monochromatyczne (o określonej długości fali) można scharakteryzować temperaturą tej barwy w kelvinach (K)
- odnosi się to do promieniowania ciała doskonale czarnego w tej temperaturze
- ciało doskonale czarne to pojęcie abstrakcyjne (nie istnieje w rzeczywistości); nie odbija ono żadnego światła, które na nie pada, wszystko pochłania, zatem całe emitowane światło pochodzi z promieniowania temperaturowego
- barwa ta odnosi się do maksimum spektrum emitowanego przez to ciało

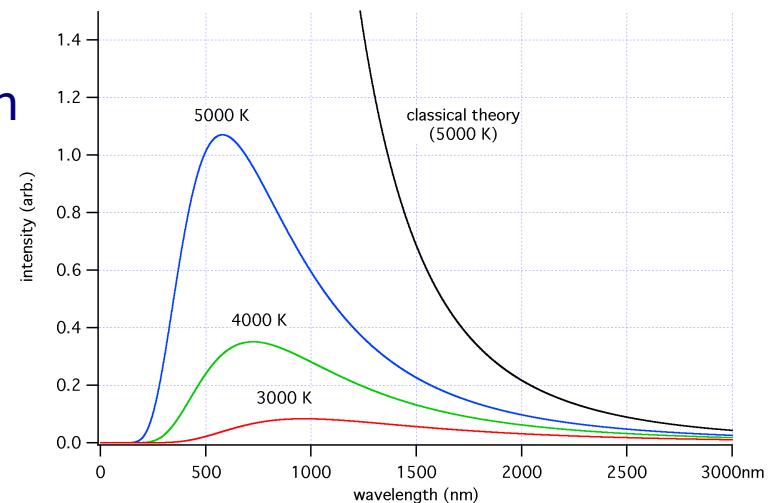


Barwa (chrominancja)  
w zależności od temperatury  
w układzie CIE XYZ 1931



## Barwa koloru

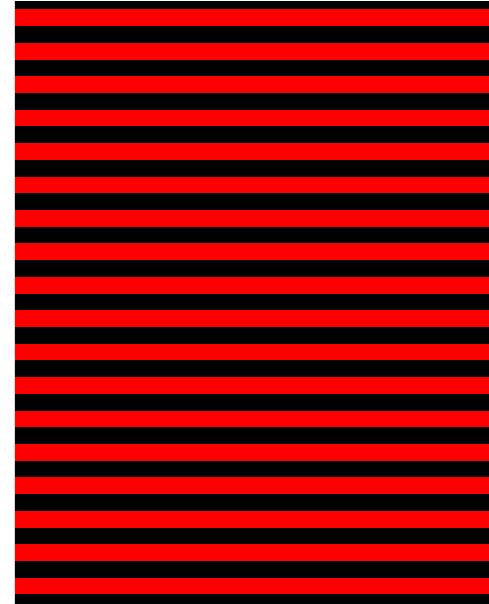
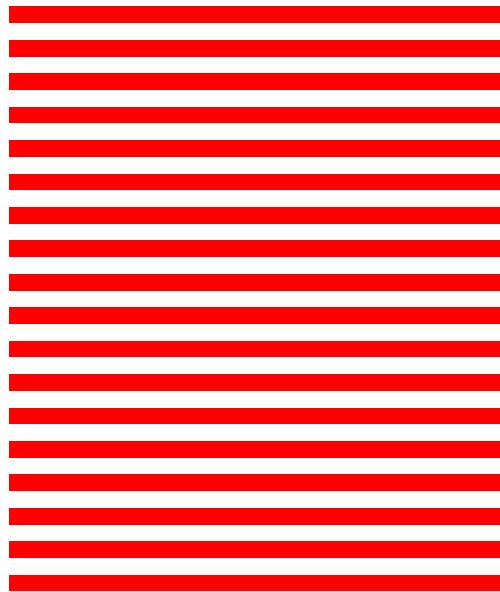
Drphysics, Wikipedia



- barwa o temperaturze  $T$  odnosi się do maksimum spektrum fal emitowanego przez to ciało
- niektóre temperatury popularnych barw:
  - 1700 K: zapałka
  - 1850 K: świeca
  - 2800 K: żarówka wolframowa
  - 3400 K: lampy studyjne
  - 4100 K: światło księżyca
  - 5000 K: typowe światło dzienne w ciepły dzień
  - 5500–6000 K: typowe światło dzienne w chłodny dzień, lampy błyskowe
  - 6420 K: lampy ksenonowe
  - 6500 K: wzorcowe światło dzienne (D65)
  - 9300 K: analogowy ekran TV
- charakterystyka monitorów – jaką biel mogą wyświetlić



## Pułapki w percepcji kolorów (1)

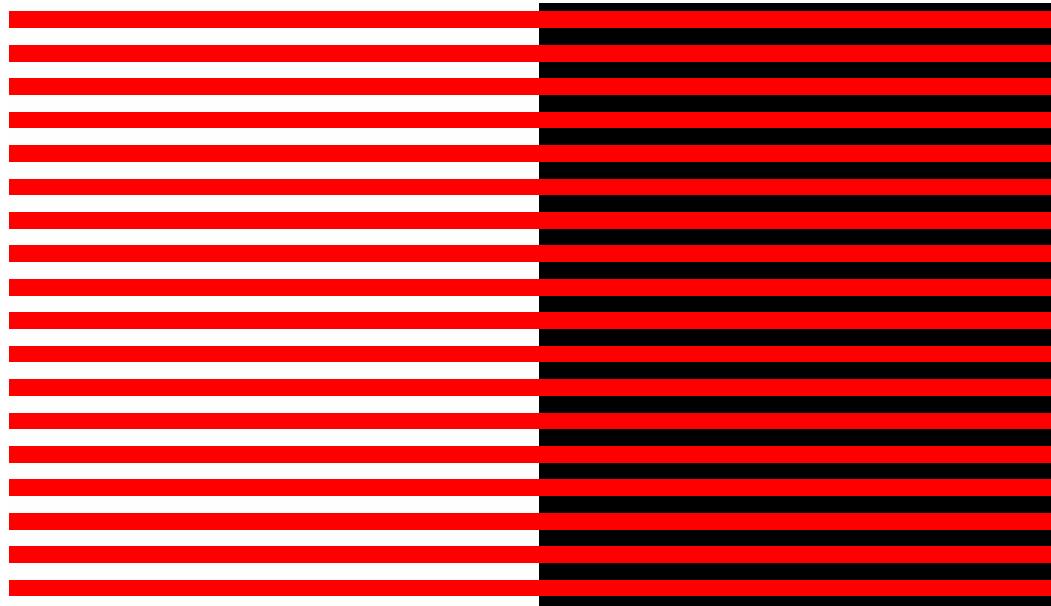


Lockal, Wikipedia

Efekt Bezolda – różnica w intensywności koloru w zależności od otoczenia



## Pułapki w percepcji kolorów (1)



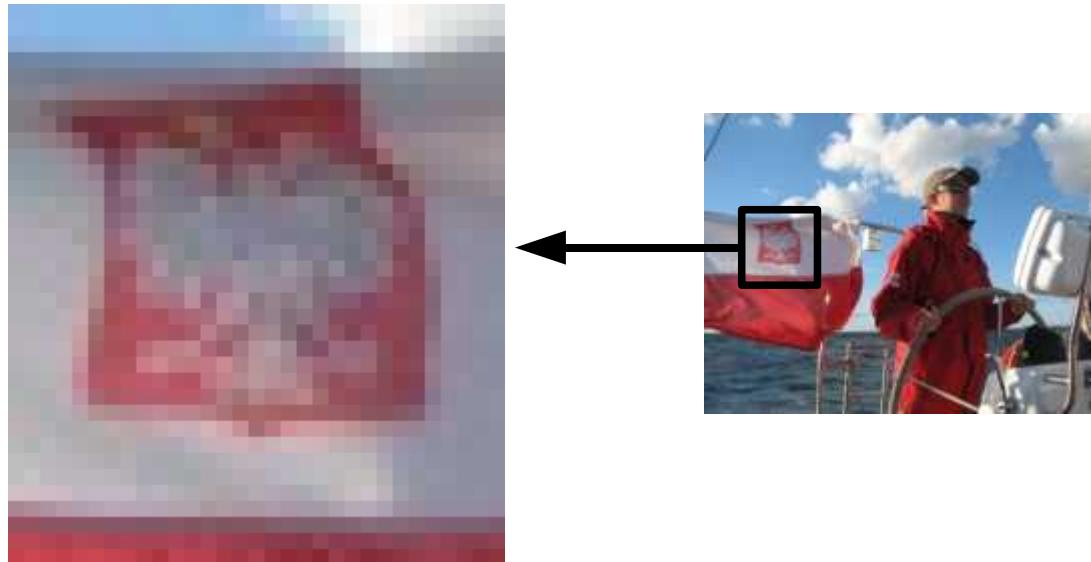
Lockal, Wikipedia

Efekt Bezolda – różnica w intensywności koloru w zależności od otoczenia



## Obraz cyfrowy rastrowy (bitmapa)

- macierz  $M \times N$  (lub  $M \times N \times O$ )



- komórki macierzy nazywane są **pikselami (wokselami w 3D)**
- każdy piksel zawiera wektor (kanałów) opisujący kolor:
  - 1-bitowy dla obrazów binarnych
  - 1-elementowy dla obrazów monochromatycznych
  - 3-elementowy dla obrazów barwnych ([R G B], [H S V], [L a b], ...)
  - wieloelementowy dla obrazów multymodalnych
- różne są zakresy tych kanałów, np: [0-255] lub [0-1]



## Obrazy indeksowane (z paletą kolorów)

- nie zawsze wszystkie kolory są potrzebne
- ze względu na objętość obrazu używa się **palety** – tablicy kolorów użytych w obrazie, ponumerowanych i w pełnej formie
- obraz zawiera wtedy ich indeksy w palecie
- palety można wymieniać bez ingerencji w sam obraz
- palety systemowe

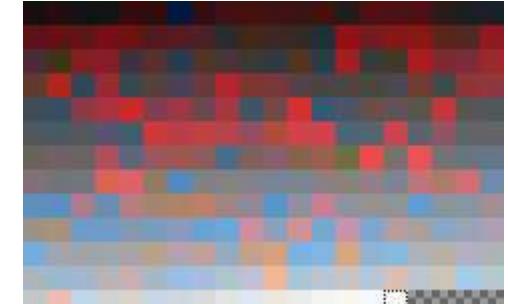
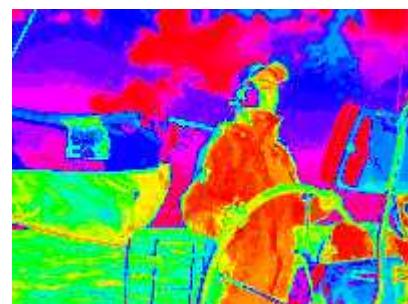
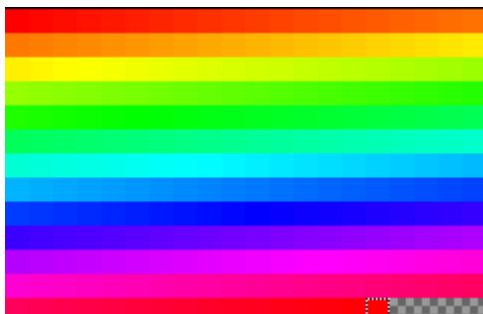
obraz cyfrowy indeksowany

1	1	2	3	4	4
1	2	2	3	4	4
1	1	1	1	3	4
1	1	1	2	3	4

paleta kolorów

Indeks	R	G	B
1	245	23	12
2	12	33	122
3	123	123	123
4	0	0	255

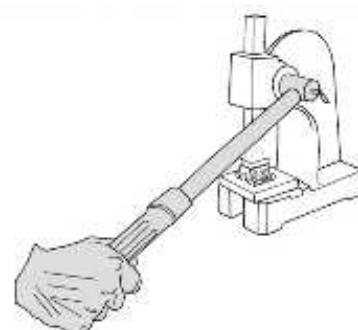
ten sam obraz zwizualizowany z z różnymi paletami kolorów





## Obraz rastrowy a obraz wektorowy

- obiekty w obrazie reprezentowane są przez prymitywy (proste, krzywe, punkty) o rzeczywistych współrzędnych/wymiarach
- możliwa jest dowolna zmiana rozdzielczości bez utraty informacji
- zastosowania:
  - fonty TrueType
  - obrazy CorelDraw, SVG
- na tym wykładzie obraz cyfrowy to głównie obraz **rastrowy**
- przed wyświetleniem zmiana na obraz rastrowy
- odwrotny proces to wektoryzacja



Yyy, Wikipedia



7x Magnification

Vector



Bitmap

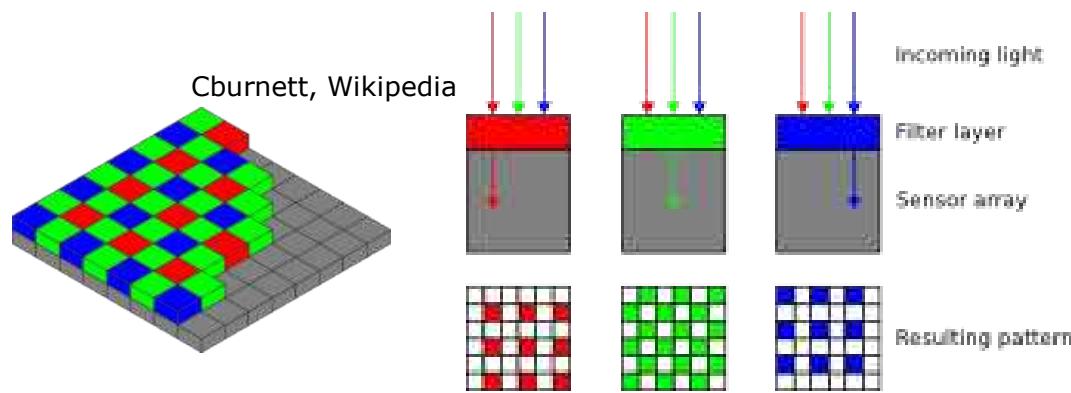


Mahlum, Wikipedia



## Powstawanie obrazu cyfrowego

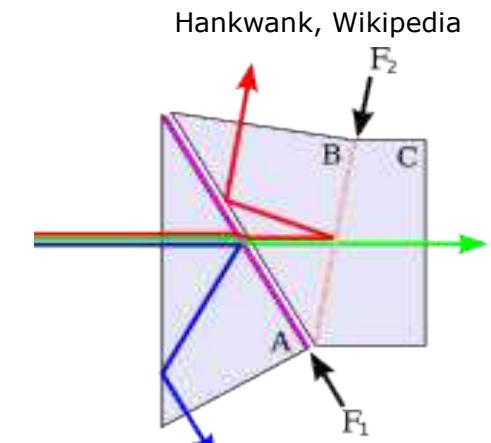
- pozyskiwanie obrazu w formie cyfrowej (kamery cyfrowe, ...)
- digitalizacja obrazów analogowych (skanery, ...)
- w obu przypadkach źródłem jest „w przybliżeniu” coś analogowego  
światło jest reprezentowane przez fotonu – dualizm korpuskularno-falowy
- matryce światłoczułe(główne CCD lub CMOS)
  - złożone, 3 kolory na 1 chipie, tańsze
  - monochromatyczne (po rozdzieleniu kolorów), droższe, wyższa jakość obrazu



Filtr Bayera na matrycy złożonej.

Kolor zielony przenosi większość informacji o luminancji.

Obraz finalny jest interpolowany z takiej matrycy.



Pryzmat trichroiczny rozdzielający światło na 3 składowe.

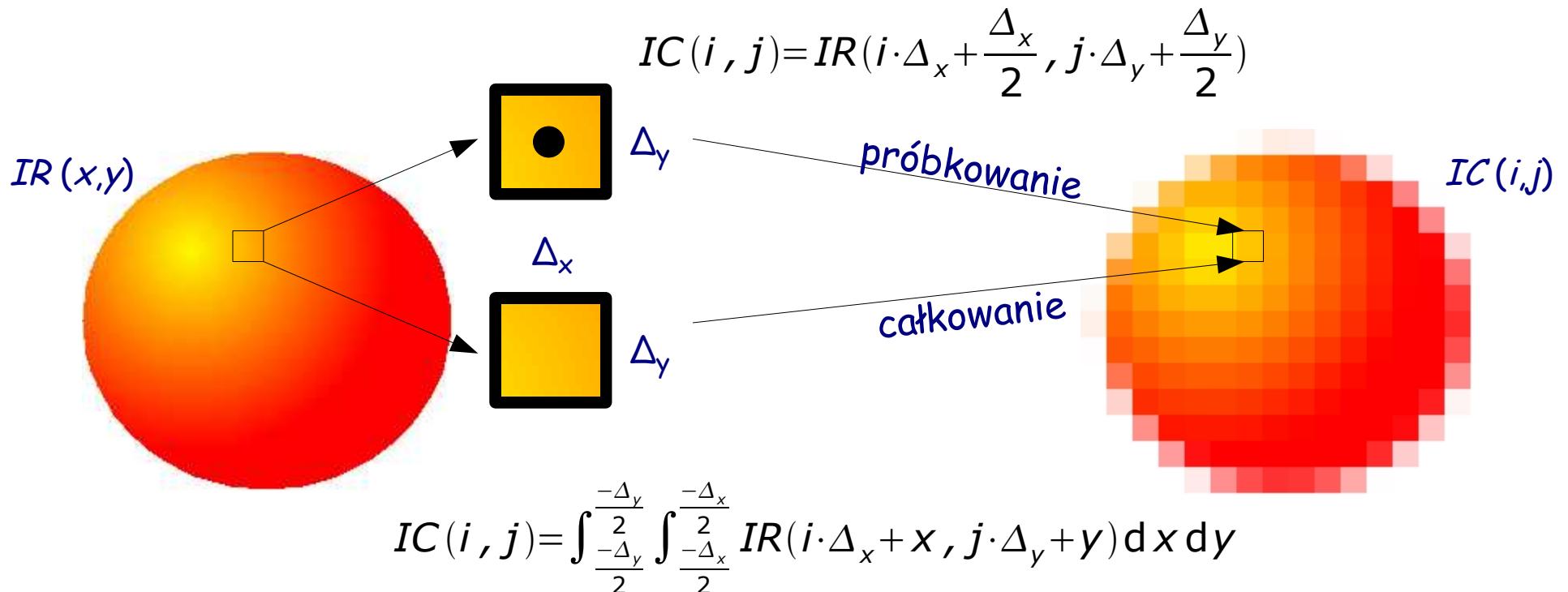
F1 - filtr dolnoprzepustowy

F2 - filtr górnoprzepustowy



## Pikselizacja

- przejście z obrazu o **praktycznie** nieskończonej rozdzielczości na obraz cyfrowy
- $\Delta_x, \Delta_y, i=0, \dots, M-1, j=0, \dots, N-1$  definiują rozdzielczość i rozmiar obrazu cyfrowego
- próbkowanie (ang. *sampling*) kontra całkowanie





## Inne zastosowania pikselizacji

- ukrycie pewnych fragmentów obrazu

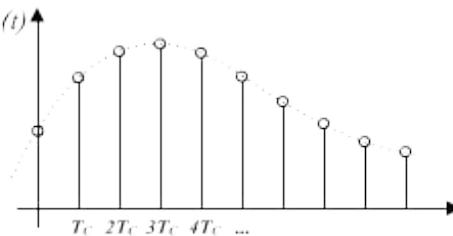
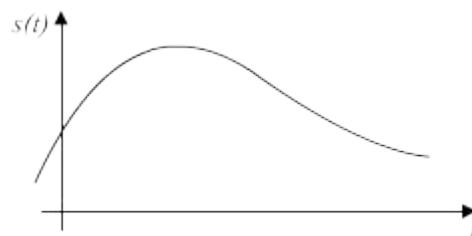


- zmniejszenie obrazu, oszczędność pamięci



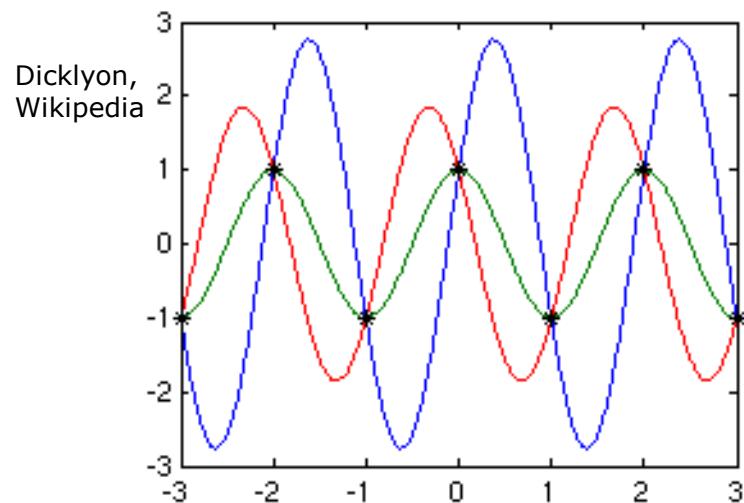


## Problemy z próbkowaniem



Ivan.Romero, Wikipedia

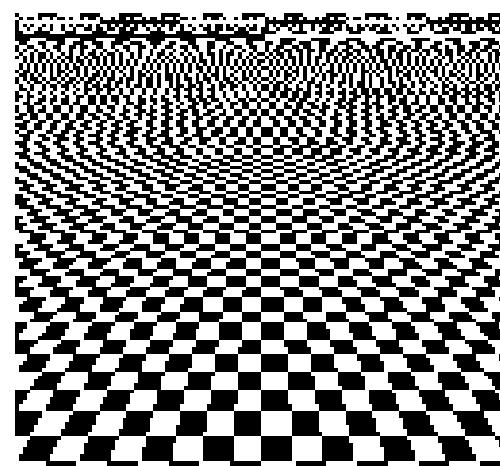
- minimalna częstotliwość próbkowania musi być 2 razy większa niż częstotliwość Nyquist'a aby zapewnić jednoznaczność rekonstrukcji
- próbkowanie z mniejszą częstotliwością (ang. *subsampling*) w przypadku obrazów cyfrowych prowadzi do **aliasingu**



Dicklyon,  
Wikipedia

$$f_{\text{próbkowania}} = 2 * f_{\text{Nyquist'a}}$$

trzy sygnały dopasowane do sygnału cyfrowego

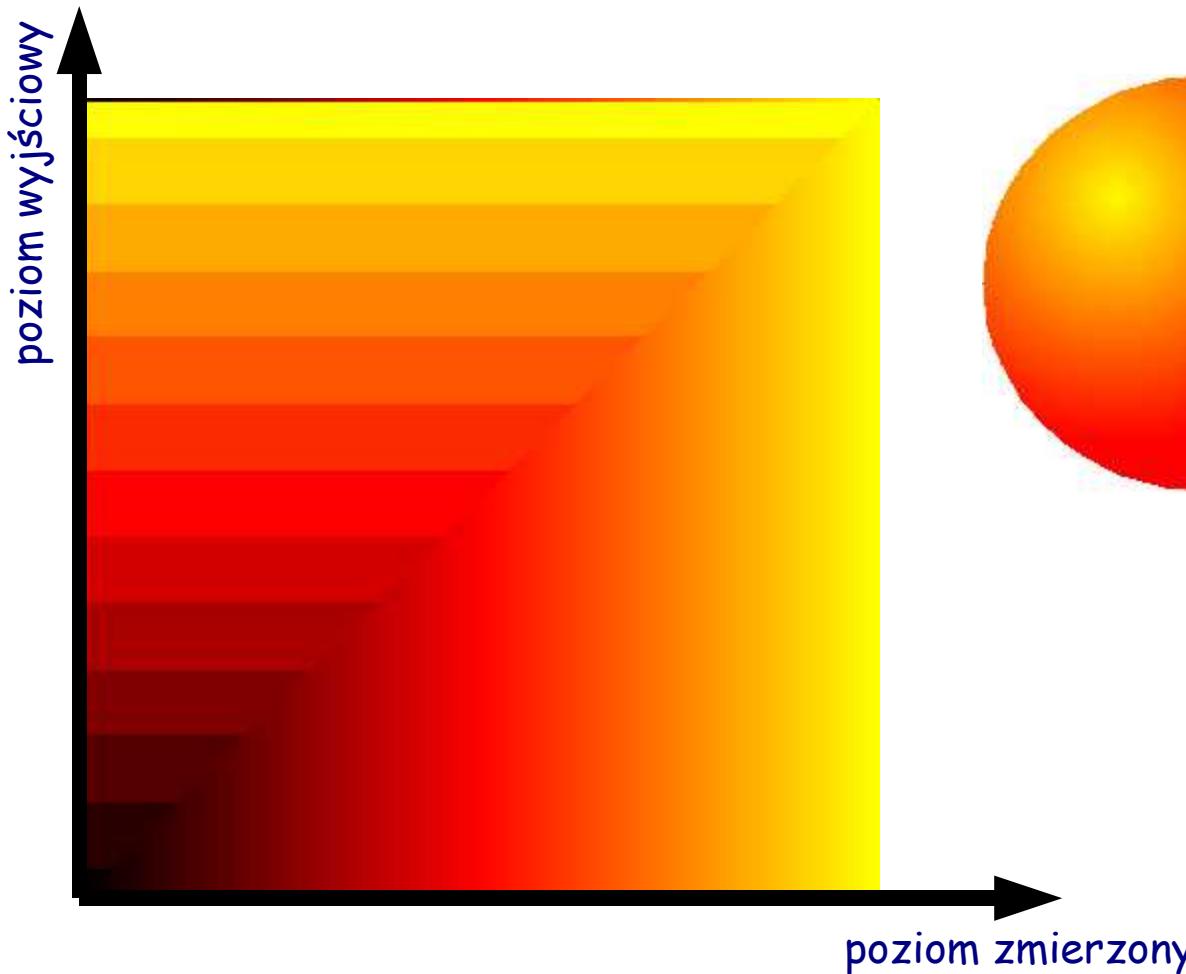


Loisel, Wikipedia



## Kwantyzacja kolorów

- urządzenia cyfrowe nie umożliwiają zapisu ciągłych wartości
- zmierzone wartości kanałów R, G, B, luminancji, chrominancji, ... muszą być zmapowane na skończony zbiór wartości
- najprostsze mapowanie proporcjonalne w przypadku pojedynczego kanału





## Kwantyzacja jednokanałowa

obraz oryginalny  
256 poziomów



128 poziomów



64 poziomy



32 poziomy



16 poziomów



8 poziomów



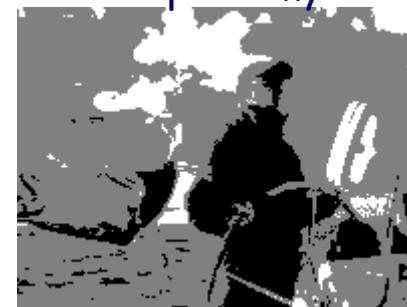
6 poziomów



4 poziomy



3 poziomy



2 poziomy

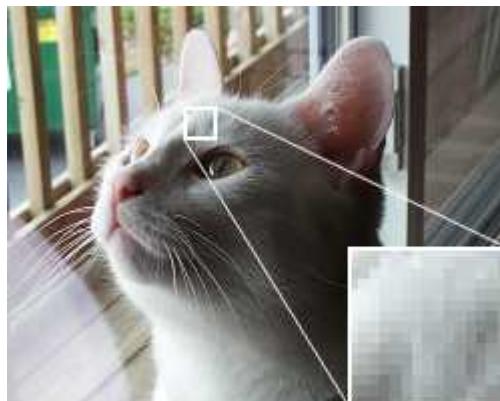




## Dithering (rozsiewanie, rozpraszanie)

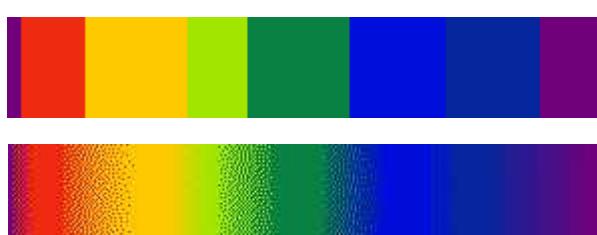
- symulacja większej liczby kolorów niż ich rzeczywista liczba (paleta) przez złożenie wzorca z istniejących kolorów
- również „zlagodzenie” krawędzi (przejścia kolorów)

Wapcaplet, Wikipedia



oryginalny obraz  
pełna skala poziomów

Wikipedia



zlagodzenie przejścia  
8 kolorów



ograniczona skala poziomów  
bez ditheringu

Wapcaplet,  
Wikipedia

symulacja koloru  
różowego  
z 2 kolorów:  
czerwonego  
i niebieskiego



ograniczona skala poziomów  
z ditheringiem



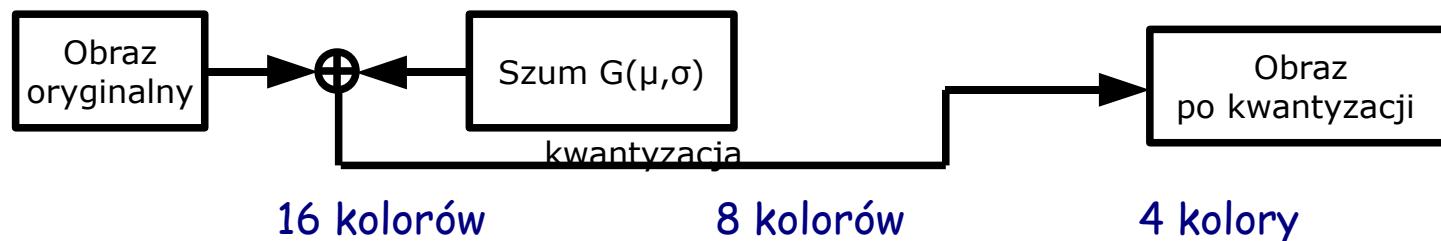
Gerbrant,  
Wikipedia

obraz binarny  
udający obraz  
w skali  
szarości



## Rozsiewanie losowe

- czasami wystarczy przed kwantyzacją dodać do obrazu losowy szum o rozkładzie Gaussa ( $\mu=0$ ,  $\sigma=M/(4q)$ ) gdzie: M – maksymalna wartość intensywności (zwykle 255) a q – liczba bitów kolorów w skwantyzowanym obrazie



kwantyzacja  
bez szumu



kwantyzacja  
po dodaniu szumu



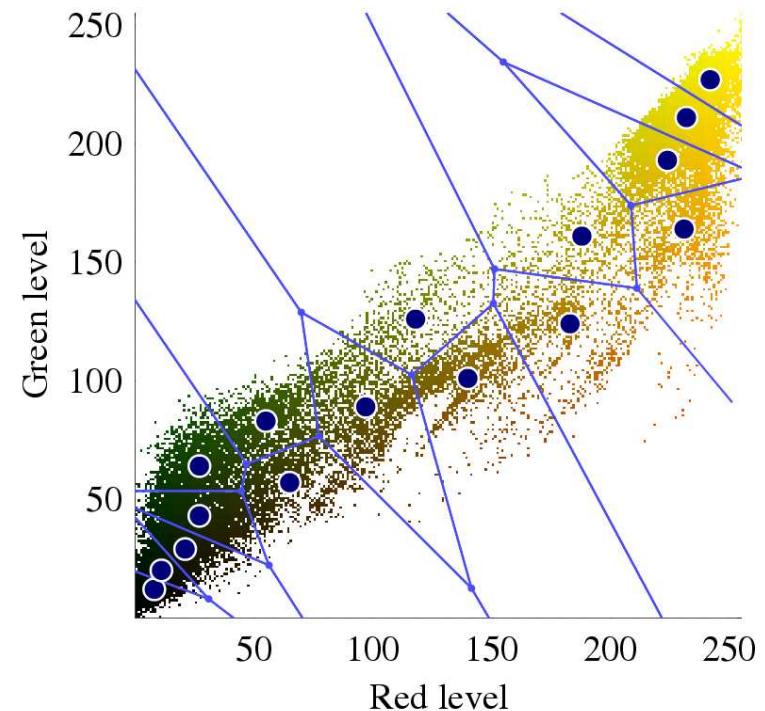


## Kwantyzacja wielokanałowa

- obrazy barwne są zlokalizowane w trójwymiarowej przestrzeni kolorów
- każdy kanał można kwantyzować oddzielnie, ale np. w modelu RGB nie prowadzi to do zbyt dobrych wyników (nienaturalne kolory)
- statyczna kwantyzacja stosuje stałe poziomy
- dynamiczna kwantyzacja dopasowuje poziomy wyjściowe do występujących najczęściej kolorów (analiza skupień w przestrzeni 3D, ang. *clustering*)



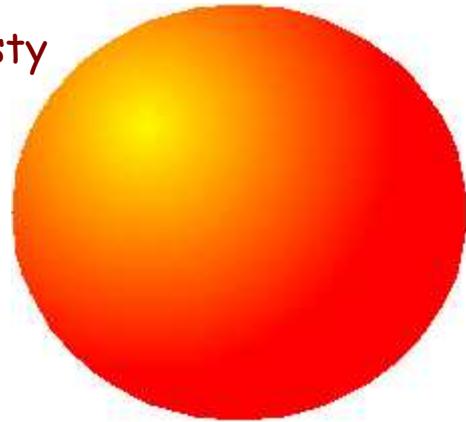
Dcoetze, Wikipedia



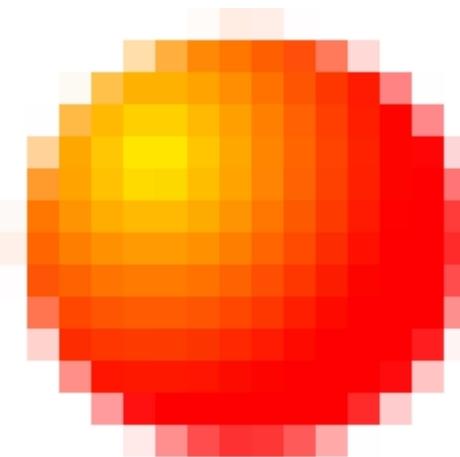
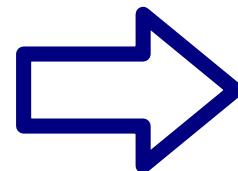


## Obraz cyfrowy

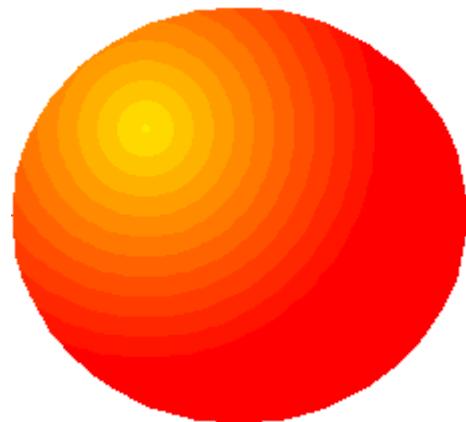
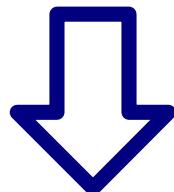
Obraz  
rzeczywisty



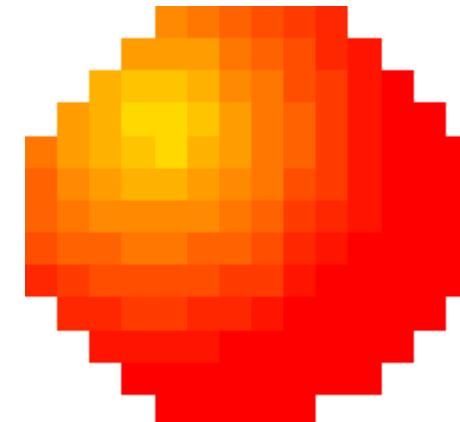
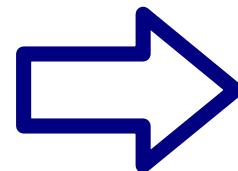
Pikselizacja



Kwantyzacja  
kolorów



Pikselizacja



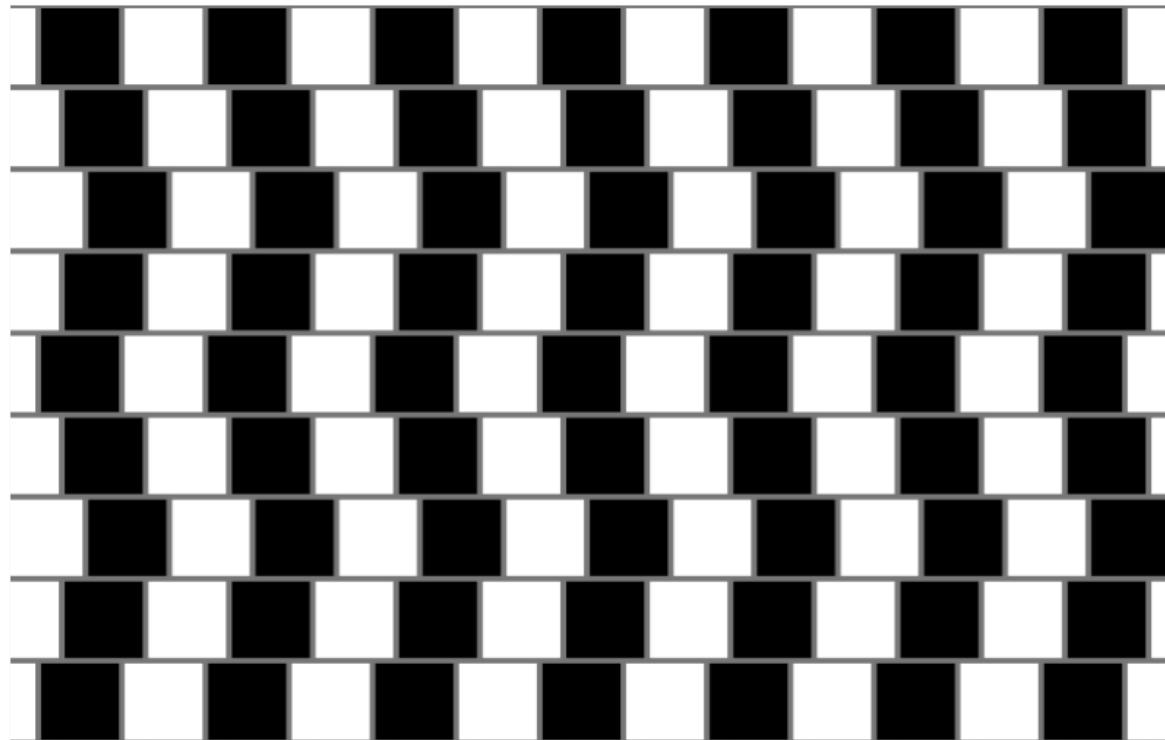
Kwantyzacja  
kolorów

Obraz  
cyfrowy



## Pułapki w percepcji kształtów (1)

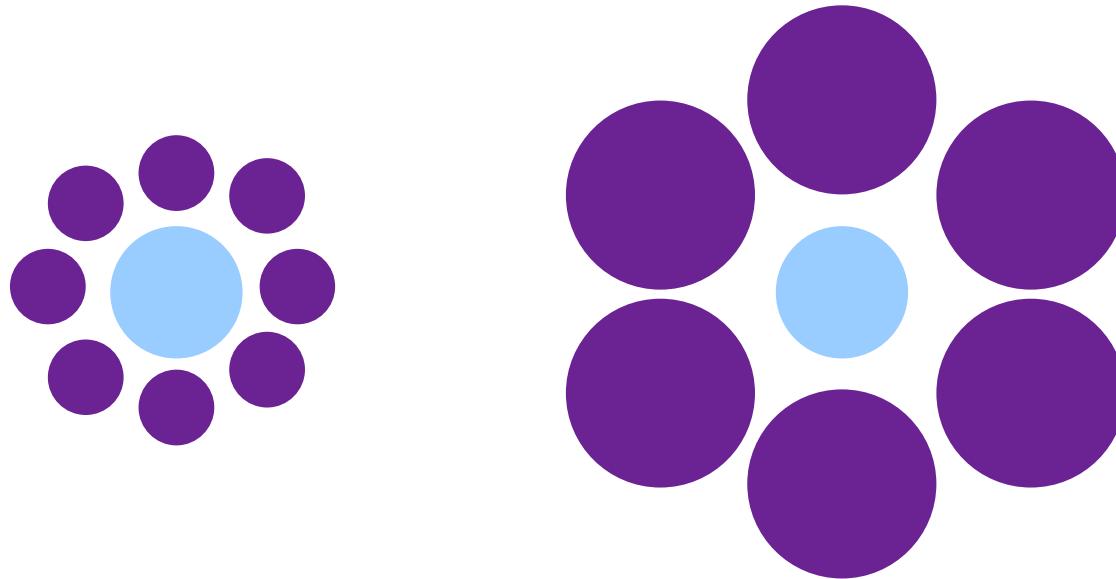
Fibonacci, Wikipedia



Iluzja 'Café Wall' – linia są równoległe czy nie ?



## Pułapki w percepcji kształtów (2)

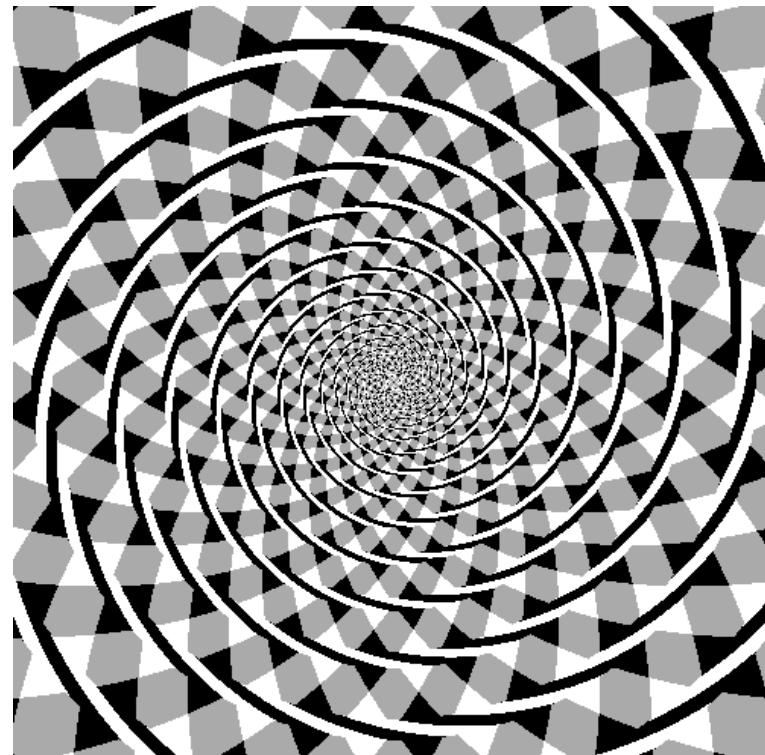


Iluzja Ebbinghausa – które z niebieskich kół jest większe ?



## Pułapki w percepcji kształtów (3)

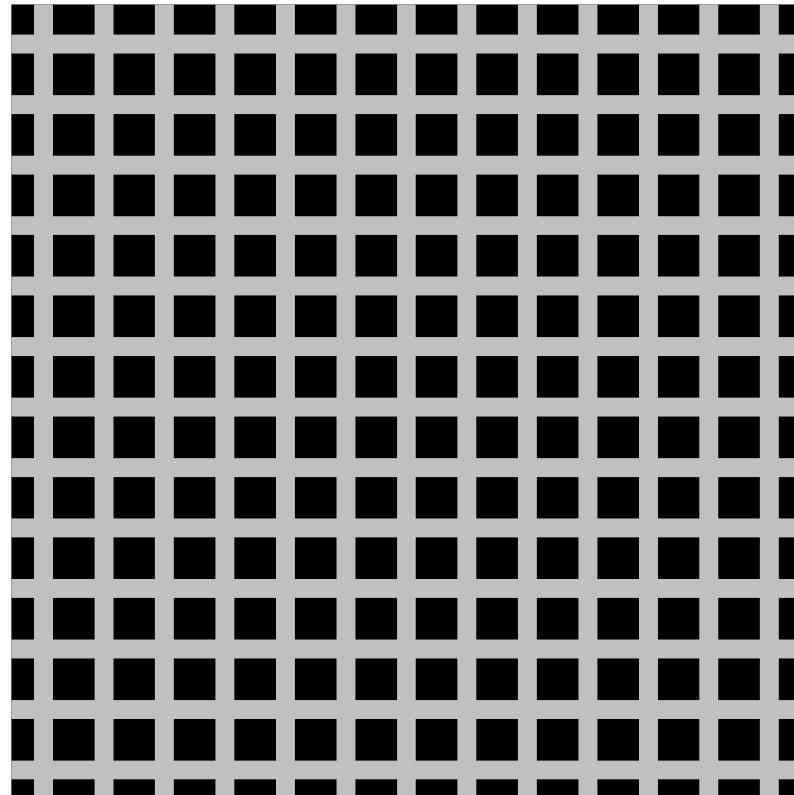
Wikipedia



Spirala Frasera – czy aby na pewno spirala ?



## Pułapki w percepcji kształtów (4)

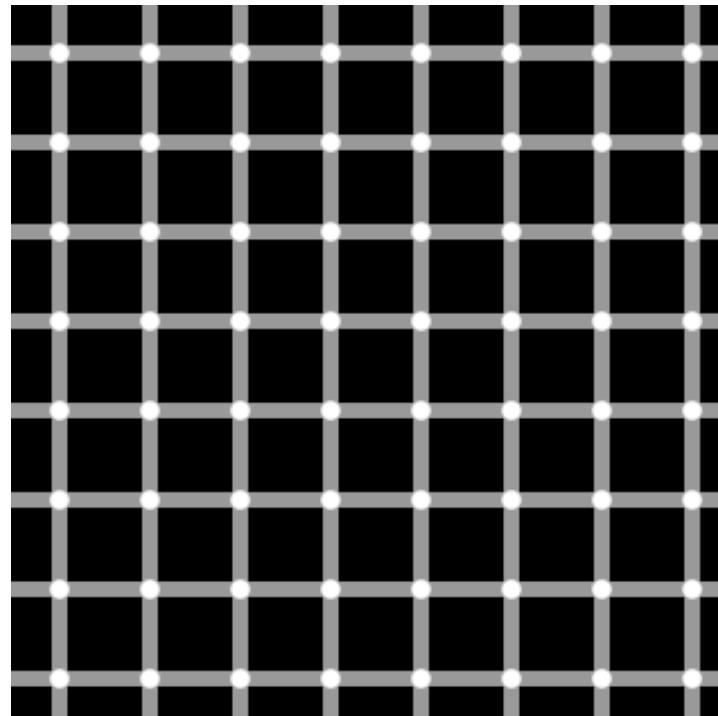


Siatka Hermanna – czy widać czarne kropki ?



## Pułapki w percepcji kształtów (4)

Wikipedia

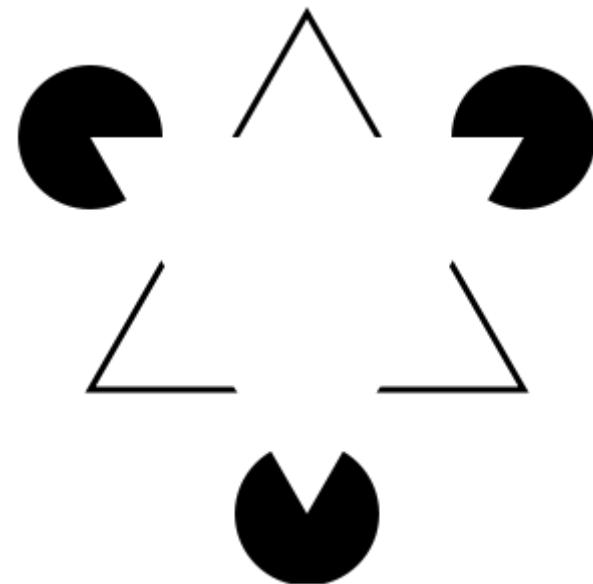


A teraz ?



## Pułapki w percepcji kształtów (5)

Fibonacci. Wikipedia

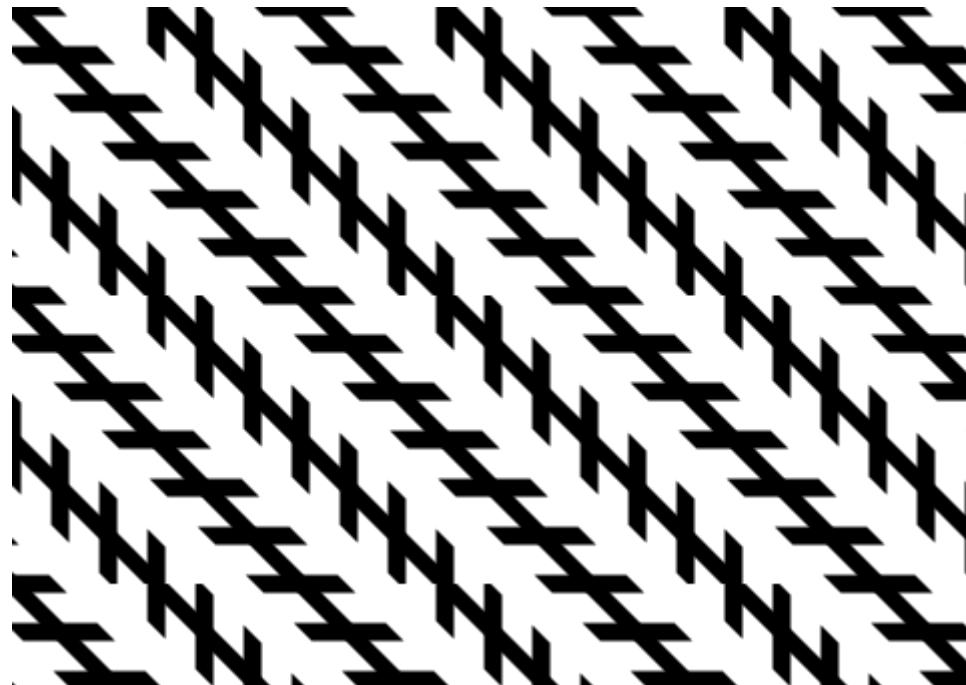


Trójkąt Kanizsa - istnieje czy nie ?



## Pułapki w percepcji kształtów (6)

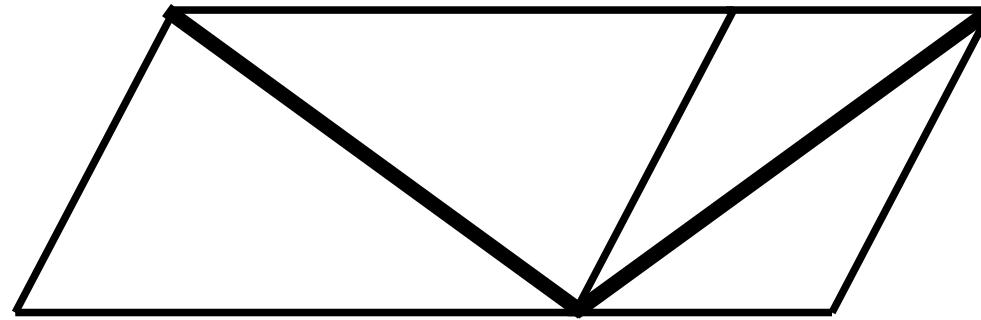
Fibonacci. Wikipedia



Iluzja Zollnera – wszystkie linie są równoległe !



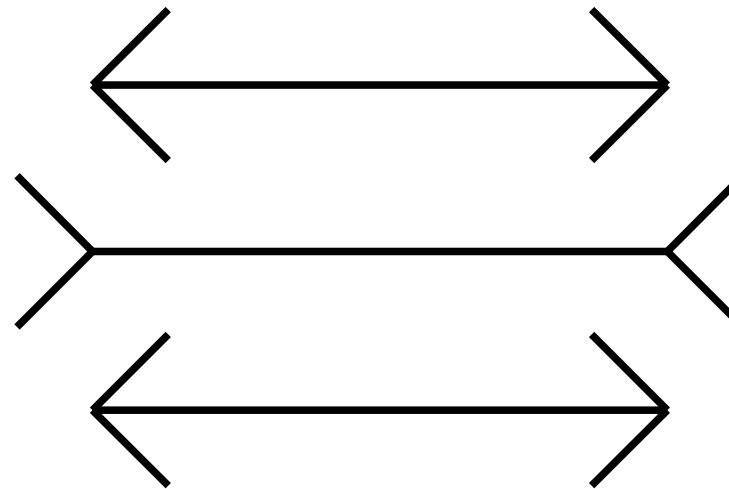
## Pułapki w percepcji kształtów (7)



Iluzja Sandera – która z przekątnych jest dłuższa ?



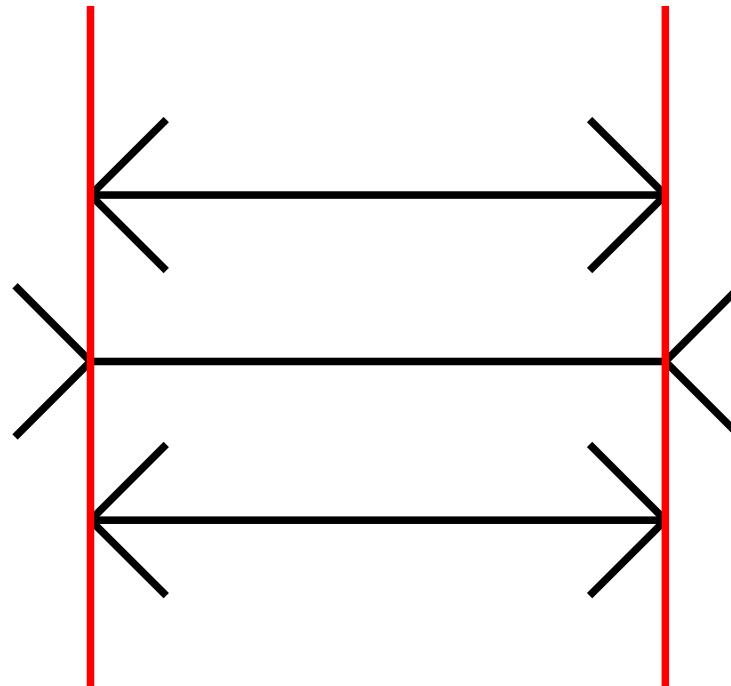
## Pułapki w percepcji kształtów (8)



Iluzja Müllera-Lyera – znowu problem z długością !

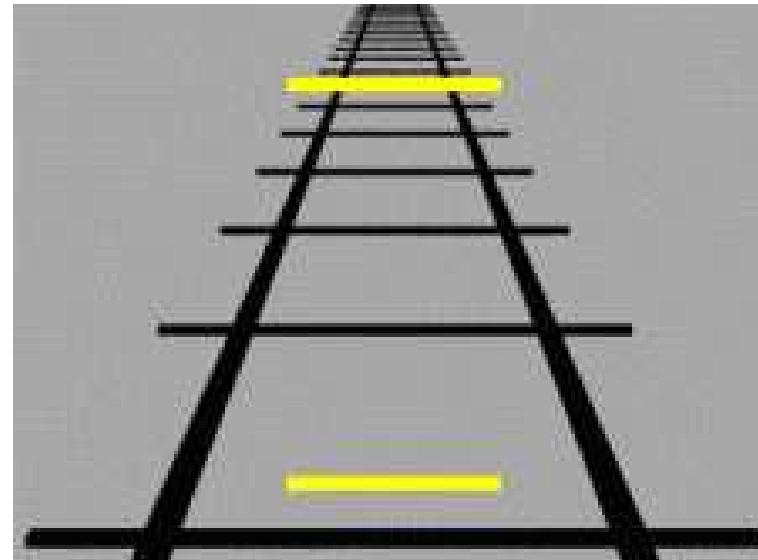


## Pułapki w percepcji kształtów (8)





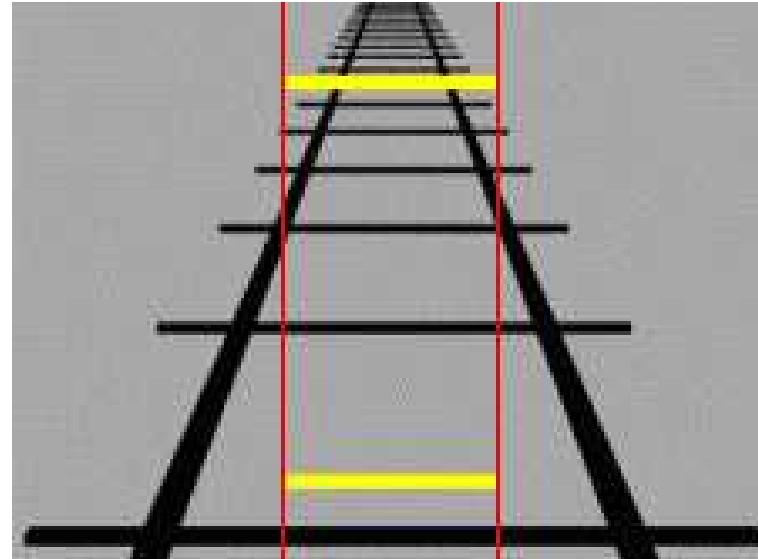
## Pułapki w percepcji kształtów (9)



Iluzja Ponza – długość znowu zależy od kontekstu !



## Pułapki w percepcji kształtów (9)



Iluzja Ponza – długość znowu zależy od kontekstu !

# GNU Free Documentation License



Version 1.2, November 2002

Copyright (C) 2000,2001,2002 Free Software Foundation, Inc.  
51 Franklin St, Fifth Floor, Boston, MA 02110-1301 USA  
Everyone is permitted to copy and distribute verbatim copies  
of this license document, but changing it is not allowed.

## 0. PREAMBLE

The purpose of this License is to make a manual, textbook, or other functional and useful document "free" in the sense of freedom: to assure everyone the effective freedom to copy and redistribute it, with or without modifying it, either commercially or noncommercially. Secondly, this License preserves for the author and publisher a way to get credit for their work, while not being considered responsible for modifications made by others.

This License is a kind of "copyleft", which means that derivative works of the document must themselves be free in the same sense. It complements the GNU General Public License, which is a copyleft license designed for free software.

We have designed this License in order to use it for manuals for free software, because free software needs free documentation: a free program should come with manuals providing the same freedoms that the software does. But this License is not limited to software manuals; it can be used for any textual work, regardless of subject matter or whether it is published as a printed book. We recommend this License principally for works whose purpose is instruction or reference.

## 1. APPLICABILITY AND DEFINITIONS

This License applies to any manual or other work, in any medium, that contains a notice placed by the copyright holder saying it can be distributed under the terms of this License. Such a notice grants a world-wide, royalty-free license, unlimited in duration, to use that work under the conditions stated herein. The "Document", below, refers to any such manual or work. Any member of the public is a licensee, and is addressed as "you". You accept the license if you copy, modify or distribute the work in a way requiring permission under copyright law.

A "Modified Version" of the Document means any work containing the Document or a portion of it, either copied verbatim, or with modifications and/or translated into another language.

A "Secondary Section" is a named appendix or a front-matter section of the Document that deals exclusively with the relationship of the publishers or authors of the Document to the Document's overall subject (or to related matters) and contains nothing that could fall directly within that overall subject. (Thus, if the Document is in part a textbook of mathematics, a Secondary Section may not explain any mathematics.) The relationship could be a matter of historical connection with the subject or with related matters, or of legal, commercial, philosophical, ethical or political position regarding them.

The "Invariant Sections" are certain Secondary Sections whose titles are designated, as being those of Invariant Sections, in the notice that says that the Document is released under this License. If a section does not fit the above definition of Secondary then it is not allowed to be designated as Invariant. The Document may contain zero Invariant Sections. If the Document does not identify any Invariant Sections then there are none.

The "Cover Texts" are certain short passages of text that are listed, as Front-Cover Texts or Back-Cover Texts, in the notice that says that the Document is released under this License. A Front-Cover Text may be at most 5 words, and a Back-Cover Text may be at most 25 words.

A "Transparent" copy of the Document means a machine-readable copy, represented in a format whose specification is available to the general public, that is suitable for revising the document straightforwardly with generic text editors or (for images composed of pixels) generic paint programs or (for drawings) some widely available drawing editor, and that is suitable for input to text formatters or for automatic translation to a variety of formats suitable for input to text formatters. A copy made in an otherwise Transparent file format whose markup, or absence of markup, has been arranged to thwart or discourage subsequent modification by readers is not Transparent. An image format is not Transparent if used for any substantial amount of text. A copy that is not "Transparent" is called "Opaque".

Examples of suitable formats for Transparent copies include plain ASCII without markup, Texinfo input format, LaTeX input format, SGML or XML using a publicly available DTD, and standard-conforming simple HTML, PostScript or PDF designed for human modification. Examples of transparent image formats include PNG, XCF and JPG. Opaque formats include proprietary formats that can be read and edited only by proprietary word processors, SGML or XML for which the DTD and/or processing tools are not generally available, and the machine-generated HTML, PostScript or PDF produced by some word processors for output purposes only.

The "Title Page" means, for a printed book, the title page itself, plus such following pages as are needed to hold, legibly, the material this License requires to appear in the title page. For works in formats which do not have any title page as such, "Title Page" means the text near the most prominent appearance of the work's title, preceding the beginning of the body of the text.

A section "Entitled XYZ" means a named subunit of the Document whose title either is precisely XYZ or contains XYZ in parentheses following text that translates XYZ in another language. (Here XYZ stands for a specific section name mentioned below, such as "Acknowledgements", "Dedications", "Endorsements", or "History".) To "Preserve the Title" of such a section when you modify the Document means that it remains a section "Entitled XYZ" according to this definition.

The Document may include Warranty Disclaimers next to the notice which states that this License applies to the Document. These Warranty Disclaimers are considered to be included by reference in this License, but only as regards disclaiming warranties: any other implication that these Warranty Disclaimers may have is void and has no effect on the meaning of this License.

## 2. VERBATIM COPYING

You may copy and distribute the Document in any medium, either commercially or noncommercially, provided that this License, the copyright notices, and the license notice saying this License applies to the Document are reproduced in all copies, and that you add no other conditions whatsoever to those of this License. You may not use technical measures to obstruct or control the reading or further copying of the copies you make or distribute. However, you may accept compensation in exchange for copies. If you distribute a large enough number of copies you must also follow the conditions in section 3.

You may also lend copies, under the same conditions stated above, and you may publicly display copies.



### 3. COPYING IN QUANTITY

If you publish printed copies (or copies in media that commonly have printed covers) of the Document, numbering more than 100, and the Document's license notice requires Cover Texts, you must enclose the copies in covers that carry, clearly and legibly, all these Cover Texts: Front-Cover Texts on the front cover, and Back-Cover Texts on the back cover. Both covers must also clearly and legibly identify you as the publisher of these copies. The front cover must present the full title with all words of the title equally prominent and visible. You may add other material on the covers in addition. Copying with changes limited to the covers, as long as they preserve the title of the Document and satisfy these conditions, can be treated as verbatim copying in other respects.

If the required texts for either cover are too voluminous to fit legibly, you should put the first ones listed (as many as fit reasonably) on the actual cover, and continue the rest onto adjacent pages.

If you publish or distribute Opaque copies of the Document numbering more than 100, you must either include a machine-readable Transparent copy along with each Opaque copy, or state in or with each Opaque copy a computer-network location from which the general network-using public has access to download using public-standard network protocols a complete Transparent copy of the Document, free of added material. If you use the latter option, you must take reasonably prudent steps, when you begin distribution of Opaque copies in quantity, to ensure that this Transparent copy will remain thus accessible at the stated location until at least one year after the last time you distribute an Opaque copy (directly or through your agents or retailers) of that edition to the public.

It is requested, but not required, that you contact the authors of the Document well before redistributing any large number of copies, to give them a chance to provide you with an updated version of the Document.

### 4. MODIFICATIONS

You may copy and distribute a Modified Version of the Document under the conditions of sections 2 and 3 above, provided that you release the Modified Version under precisely this License, with the Modified Version filling the role of the Document, thus licensing distribution and modification of the Modified Version to whoever possesses a copy of it. In addition, you must do these things in the Modified Version:

- \* A. Use in the Title Page (and on the covers, if any) a title distinct from that of the Document, and from those of previous versions (which should, if there were any, be listed in the History section of the Document). You may use the same title as a previous version if the original publisher of that version gives permission.
- \* B. List on the Title Page, as authors, one or more persons or entities responsible for authorship of the modifications in the Modified Version, together with at least five of the principal authors of the Document (all of its principal authors, if it has fewer than five), unless they release you from this requirement.
- \* C. State on the Title page the name of the publisher of the Modified Version, as the publisher.
- \* D. Preserve all the copyright notices of the Document.
- \* E. Add an appropriate copyright notice for your modifications adjacent to the other copyright notices.
- \* F. Include, immediately after the copyright notices, a license notice giving the public permission to use the Modified Version under the terms of this License, in the form shown in the Addendum below.
- \* G. Preserve in that license notice the full lists of Invariant Sections and required Cover Texts given in the Document's license notice.
- \* H. Include an unaltered copy of this License.
- \* I. Preserve the section Entitled "History", Preserve its Title, and add to it an item stating at least the title, year, new authors, and publisher of the Modified Version as given on the Title Page. If there is no section Entitled "History" in the Document, create one stating the title, year, authors, and publisher of the Document as given on its Title Page, then add an item describing the Modified Version as stated in the previous sentence.
- \* J. Preserve the network location, if any, given in the Document for public access to a Transparent copy of the Document, and likewise the network locations given in the Document for previous versions it was based on. These may be placed in the "History" section. You may omit a network location for a work that was published at least four years before the Document itself, or if the original publisher of the version it refers to gives permission.
- \* K. For any section Entitled "Acknowledgements" or "Dedications", Preserve the Title of the section, and preserve in the section all the substance and tone of each of the contributor acknowledgements and/or dedications given therein.
- \* L. Preserve all the Invariant Sections of the Document, unaltered in their text and in their titles. Section numbers or the equivalent are not considered part of the section titles.
- \* M. Delete any section Entitled "Endorsements". Such a section may not be included in the Modified Version.
- \* N. Do not retitle any existing section to be Entitled "Endorsements" or to conflict in title with any Invariant Section.
- \* O. Preserve any Warranty Disclaimers.

If the Modified Version includes new front-matter sections or appendices that qualify as Secondary Sections and contain no material copied from the Document, you may at your option designate some or all of these sections as invariant. To do this, add their titles to the list of Invariant Sections in the Modified Version's license notice. These titles must be distinct from any other section titles.

You may add a section Entitled "Endorsements", provided it contains nothing but endorsements of your Modified Version by various parties--for example, statements of peer review or that the text has been approved by an organization as the authoritative definition of a standard.

You may add a passage of up to five words as a Front-Cover Text, and a passage of up to 25 words as a Back-Cover Text, to the end of the list of Cover Texts in the Modified Version. Only one passage of Front-Cover Text and one of Back-Cover Text may be added by (or through arrangements made by) any one entity. If the Document already includes a cover text for the same cover, previously added by you or by arrangement made by the same entity you are acting on behalf of, you may not add another; but you may replace the old one, on explicit permission from the previous publisher that added the old one.

The author(s) and publisher(s) of the Document do not by this License give permission to use their names for publicity for or to assert or imply endorsement of any Modified Version.



## 6. COLLECTIONS OF DOCUMENTS

You may make a collection consisting of the Document and other documents released under this License, and replace the individual copies of this License in the various documents with a single copy that is included in the collection, provided that you follow the rules of this License for verbatim copying of each of the documents in all other respects.

You may extract a single document from such a collection, and distribute it individually under this License, provided you insert a copy of this License into the extracted document, and follow this License in all other respects regarding verbatim copying of that document.

## 7. AGGREGATION WITH INDEPENDENT WORKS

A compilation of the Document or its derivatives with other separate and independent documents or works, in or on a volume of a storage or distribution medium, is called an "aggregate" if the copyright resulting from the compilation is not used to limit the legal rights of the compilation's users beyond what the individual works permit. When the Document is included in an aggregate, this License does not apply to the other works in the aggregate which are not themselves derivative works of the Document.

If the Cover Text requirement of section 3 is applicable to these copies of the Document, then if the Document is less than one half of the entire aggregate, the Document's Cover Texts may be placed on covers that bracket the Document within the aggregate, or the electronic equivalent of covers if the Document is in electronic form. Otherwise they must appear on printed covers that bracket the whole aggregate.

## 8. TRANSLATION

Translation is considered a kind of modification, so you may distribute translations of the Document under the terms of section 4. Replacing Invariant Sections with translations requires special permission from their copyright holders, but you may include translations of some or all Invariant Sections in addition to the original versions of these Invariant Sections. You may include a translation of this License, and all the license notices in the Document, and any Warranty Disclaimers, provided that you also include the original English version of this License and the original versions of those notices and disclaimers. In case of a disagreement between the translation and the original version of this License or a notice or disclaimer, the original version will prevail.

If a section in the Document is Entitled "Acknowledgements", "Dedications", or "History", the requirement (section 4) to Preserve its Title (section 1) will typically require changing the actual title.

## 9. TERMINATION

You may not copy, modify, sublicense, or distribute the Document except as expressly provided for under this License. Any other attempt to copy, modify, sublicense or distribute the Document is void, and will automatically terminate your rights under this License. However, parties who have received copies, or rights, from you under this License will not have their licenses terminated so long as such parties remain in full compliance.

## 10. FUTURE REVISIONS OF THIS LICENSE

The Free Software Foundation may publish new, revised versions of the GNU Free Documentation License from time to time. Such new versions will be similar in spirit to the present version, but may differ in detail to address new problems or concerns. See <http://www.gnu.org/copyleft/>.

Each version of the License is given a distinguishing version number. If the Document specifies that a particular numbered version of this License "or any later version" applies to it, you have the option of following the terms and conditions either of that specified version or of any later version that has been published (not as a draft) by the Free Software Foundation. If the Document does not specify a version number of this License, you may choose any version ever published (not as a draft) by the Free Software Foundation.

## 5. COMBINING DOCUMENTS

You may combine the Document with other documents released under this License, under the terms defined in section 4 above for modified versions, provided that you include in the combination all of the Invariant Sections of all of the original documents, unmodified, and list them all as Invariant Sections of your combined work in its license notice, and that you preserve all their Warranty Disclaimers.

The combined work need only contain one copy of this License, and multiple identical Invariant Sections may be replaced with a single copy. If there are multiple Invariant Sections with the same name but different contents, make the title of each such section unique by adding at the end of it, in parentheses, the name of the original author or publisher of that section if known, or else a unique number. Make the same adjustment to the section titles in the list of Invariant Sections in the license notice of the combined work.

In the combination, you must combine any sections Entitled "History" in the various original documents, forming one section Entitled "History"; likewise combine any sections Entitled "Acknowledgements", and any sections Entitled "Dedications". You must delete all sections Entitled "Endorsements."

(not as a draft) by the Free Software Foundation.