

Kompresja danych multimedialnych

DR INŻ. MARIUSZ SZWOCH

3. Kompresja danych multimedialnych

- **Kompresja** - zmniejszenie rozmiaru danych poprzez eliminację 'nadmiarowej' informacji
- **Współczynnik kompresji** - stosunek rozmiaru danych nieskompresowanych do skompresowanych
- Problemy
 - **symetria czasowa** kompresji i dekompresji
 - **kompresja stratna** vs **bezstratna**
 - **kompresja sprzętowa** vs **programowa**
 - **skalowalność kompresji** - możliwość dekodowania informacji niezależnej od parametrów kodowania (a zależnych od platformy sprzętowej)

Kompresja bezstratna

- Nie traci żadnej informacji z danych = możliwość dokładnego odtworzenia danych oryginalnych
- Wykorzystuje właściwości danych
- Przykładowe metody
 - RLE (*run length encoding*)
 - metody słownikowe
 - LZ77, LZ78, LZW
 - kodowania entropijne
 - Huffmana, arytmetyczne
- Względnie niski poziom kompresji, np. ~3:1
- Granice kompresji wyznaczone przez entropię

Metoda RLE

- Zastosowanie

- obrazy o jednolitych powierzchniach, najczęściej syntetyczne: grafika wektorowa (tło i wypełnienie), filmy animowane, itp.

- Algorytm standardowy

- kodowanie na 2 pozycjach
 - 1. Kolor piksela
 - 2. Liczba powtórzeń

Przykład: 9 9 9 9 9 5 5 5 5 3 2 2 4 4 4 4 4 $\Rightarrow 18 \cdot 8 = 144b$

9₆ 5₄ 3₁ 2₂ 4₅ \Rightarrow 5 * 2 * 8 = 80b

Metoda Huffmana

- Metody statystyczne - kodowanie częściej występujących wartości na mniejszej liczbie bitów (słowa kodowe nierównej długości)
 - statyczne - tabela częstotliwości jest ustalona z góry i niezmienna
 - dynamiczne - tabela częstotliwości jest ustalana każdorazowo dla kompresowanych danych
 - adaptacyjne - wstępnie ustalona tabela częstotliwości jest korygowana w trakcie kompresji

○ Przykład: 9 9 9 9 9 5 5 5 5 3 2 2 4 4 4 4 $\Rightarrow 18 \cdot 8 = 144b$

Słownik: 9 - ₁₀, 4 - ₁₁, 5 - ₀₁, 2 - ₀₀₁, 3 - ₀₀₀ $\Rightarrow 15 \cdot 2 + 3 \cdot 3 = 39$

10 10 10 10 10 10 01 01 01 01 000 001 001 11 11 11 11 11

Ograniczenie kompresji bezstratnej

- Niski stopień kompresji danych multimedialnych
 - algorytmy ogólne
 - obraz $\sim 2:1$
 - dźwięk $\sim 10\%$
 - algorytmy specjalizowane
 - obraz $\sim 3:1$
 - dźwięk $\sim 2:1$
- Potrzebne stopnie kompresji
 - obraz $\sim 5 \div 50:1$
 - dźwięk $\sim 5 \div 20:1$
 - wideo $\sim 5 \div 300:1$

Kompresja stratna k.w.nst

- Utrata części informacji / danych
- Wykorzystanie właściwości
 - sygnału
 - przetwornika A/D
 - medium transmisyjnego
 - urządzenia wyjściowego
 - percepcji człowieka
- Wysokie stopnie kompresji, np. ~5-50-500:1

Fizjologiczne podstawy kompresji

○ Obraz

- większa wrażliwość na intensywność światła niż na kolory (l.pręcików $\approx 20 \cdot$ l.czopków)
- lepsze postrzeganie środka obrazu (większa gęstość kom. wzrokowych w ok. żółtej plamki)
 - ang. *foveated rendering* - hełmy wirtualnej rzeczywistości: Vive Pro Eye, Facebook Oculus, Sony VR2

○ Wideo

- uznanie za płynne obrazu o częstotliwości 15 fps (bezwładność wzroku)
- lepsza tolerancja niskich rozdzielczości niż przy obrazach statycznych

○ Dźwięk

- maskowanie dźwięków słabszych i o zbliżonych częstotliwościach

3.1. Kompresja stratna obrazów (1)

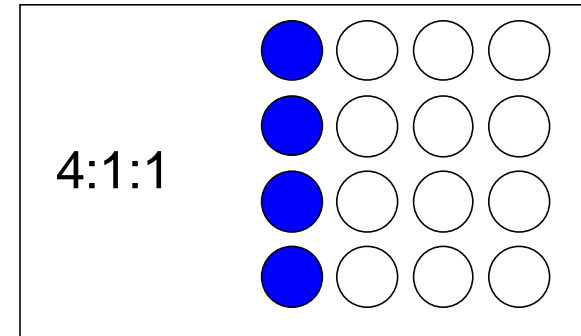
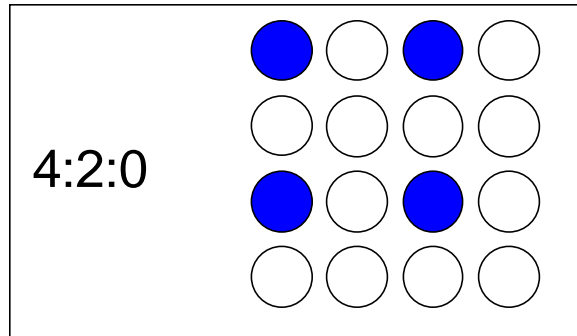
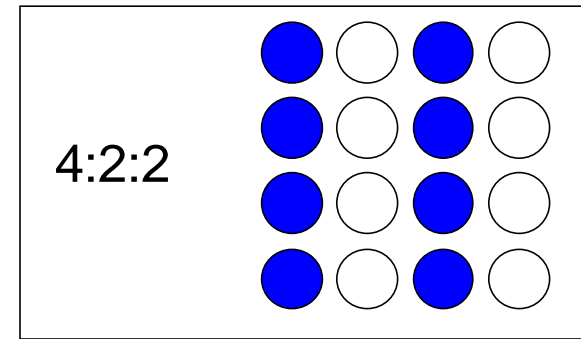
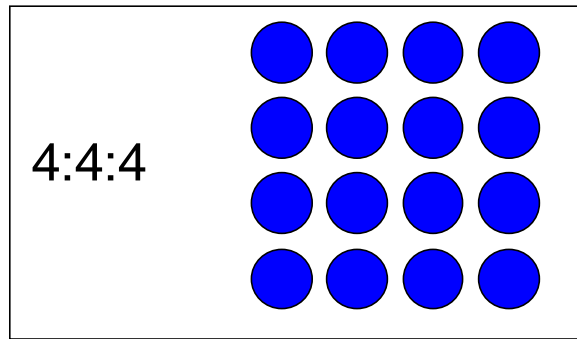
- **Redukcja dokładności kwantyzacji** (*Least Significant Bits Truncation*) - obcinanie najmniej znaczących bitów
 - RGB 555 (CD-I), YUV 6:5:5
- **Redukcja rozdzielczości** (*prunning*) - zastępowanie grupy pikseli jednym pikselem
- Wprowadzenie **palety barw** (*color-mapping*) - wybór 256 najczęściej występujących kolorów (średnia ważona):
 - problemy z odświeżaniem i wyświetlaniem wielu obrazów
- **Redukcja palety barw**, np. 256→16 kolorów

Kompresja stratna obrazów (2)

- **Redukcja chrominancji** (*color subsampling*) - konwersja RGB→YUV (4:2:2) - chrominancja o mniejszej rozdzielczości (kompresja 3:2)
- **Redukcja w dziedzinie widma**
 - transformacja danych do dziedziny widmowej z wykorzystaniem DCT (*Discrete Cosinus Transform*)
 - kwantyzacja wartości
 - zerowanie 'wysokich częstotliwości'
 - kodowanie bezstratne
- **Kompresja falkowa**
- **Kompresja fraktalna**

Próbkowanie chrominancji

- Konwersja $RGB \rightarrow YC_rC_b$
 - możliwa różna rozdzielczość próbkowania luminancji i chrominancji
- Schematy próbkowania



próbki luminancji



próbki chrominancji i luminancji

Kompresja JPEG obrazu

- JPEG - *Joint Photographic Experts Group* (1991)
 - algorytm symetryczny
 - RGB (4:4:4), YUV (4:2:2), (4:1:1), (4:4:4), CMYK
 - DCT - dyskretna transformacja cosinusowa, (dla bloków 8 x 8)

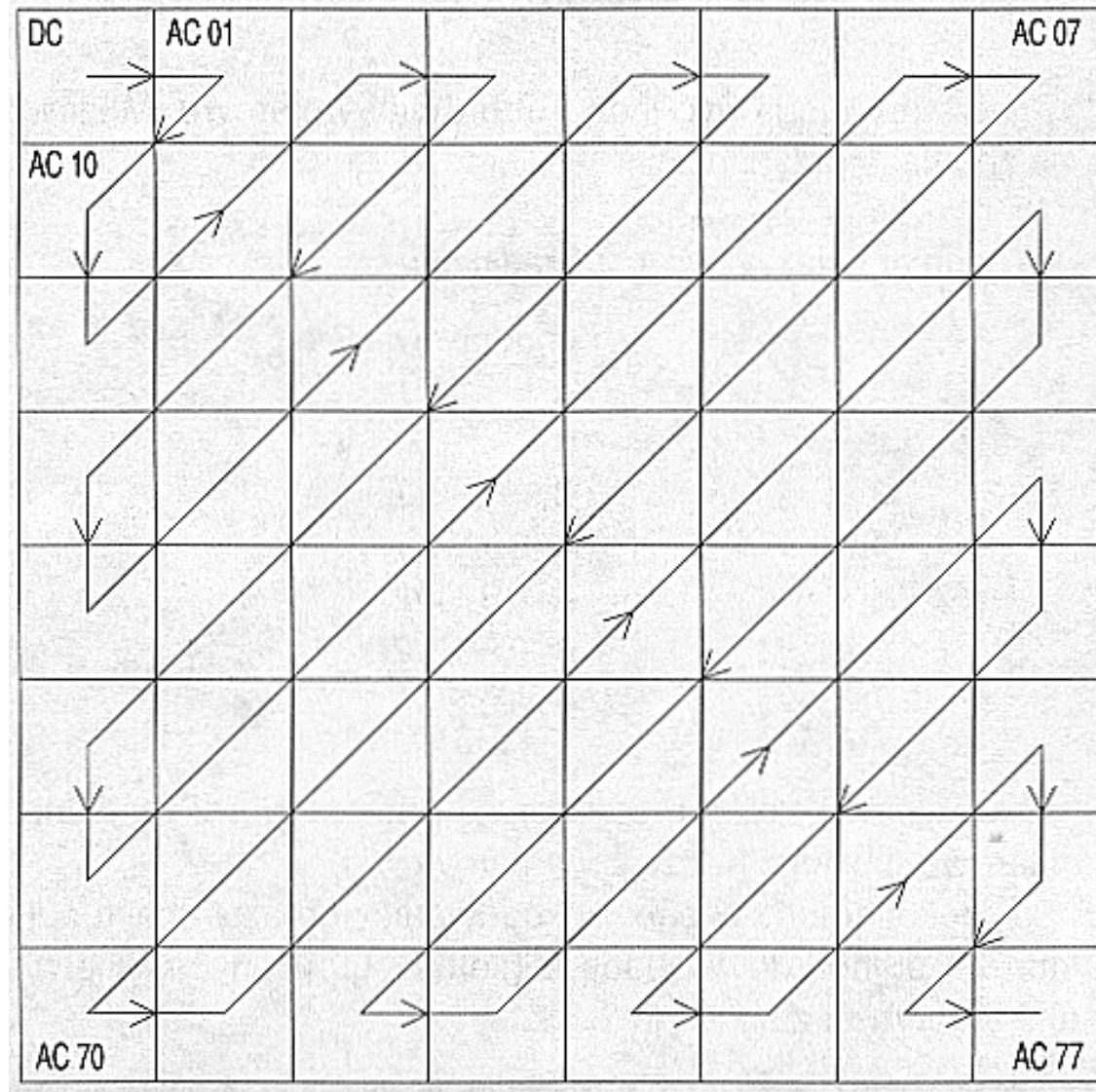
$$DCT(u, v) = \frac{2 \cdot c(u) \cdot c(v)}{n} \sum_{x=0}^{n-1} \sum_{y=0}^{n-1} f(x, y) \cdot \cos\left(\frac{(2x+1)u\pi}{2n}\right) \cdot \cos\left(\frac{(2y+1)v\pi}{2n}\right)$$

$$c(w) = \begin{cases} \frac{1}{\sqrt{2}} & \text{if } w = 0 \\ 1 & \text{if } w = 1, 2, 3, \dots, n-1 \end{cases}$$

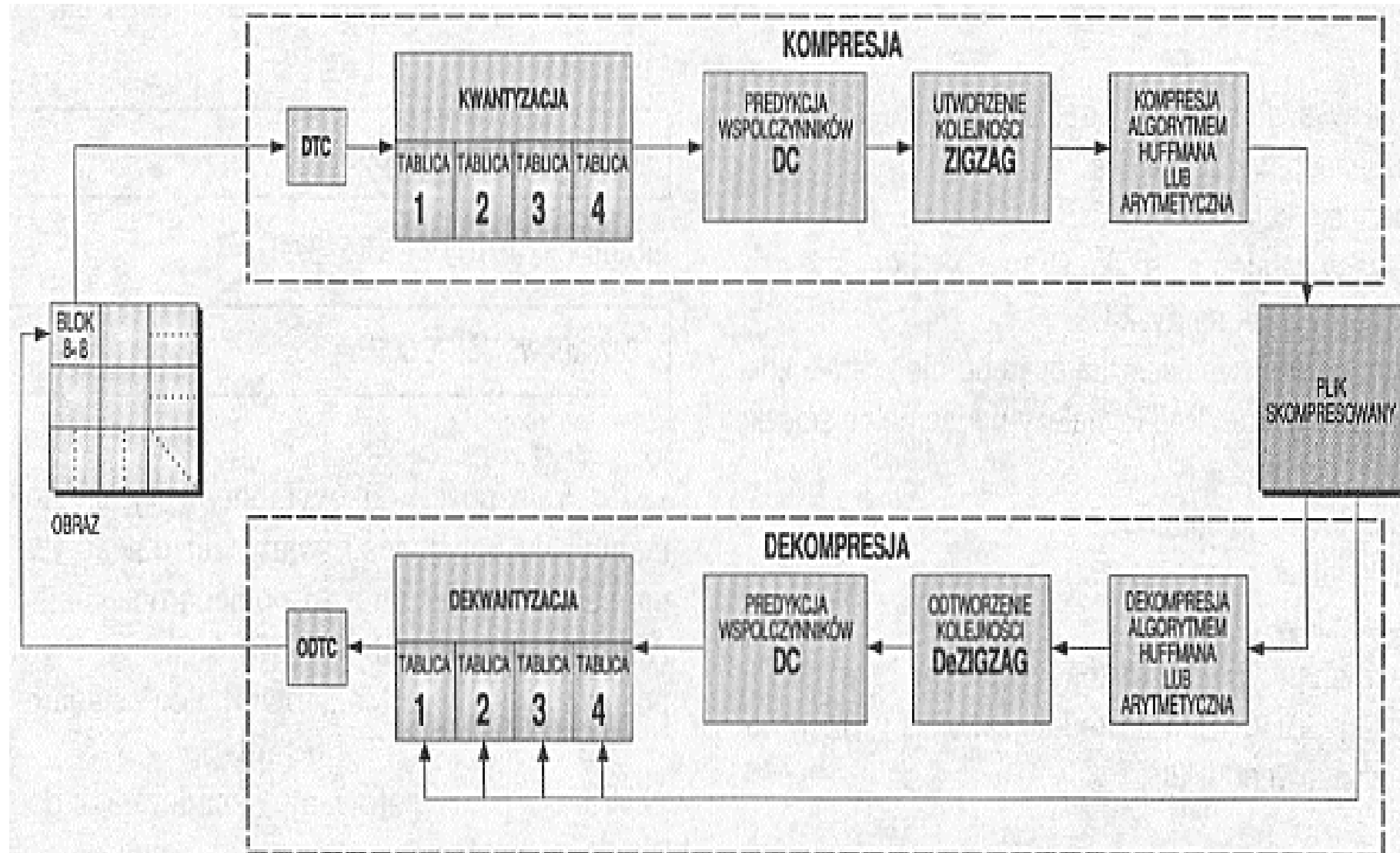
Kompresja JPEG cd.

- Kwantyzacja po stronie transformaty
- Czytanie typu „zigzag” i kodowanie RLE
- DPCM
- Kodowanie metodą Huffmana lub arytmetyczne
- 3 poziomy
 - bazowy (kodowanie metodą Huffmana)
 - rozszerzony (kodowanie progresywne, arytmetyczne)
 - specjalny (kompresja bezstratna)
- Zastosowanie: fotografia
- Współczynniki kompresji ~50:1

Odczyt w kolejności „zig-zag”



Kompresja JPEG - upr. schemat blokowy



Kompresja JPEG – przykład

- Obraz oryginalny



- Kompresja 1% (29827B)



- Kompresja 65% (4415B)



- Kompresja 99% (766B)



Fraktalna kompresja obrazów

- Grupy metod

- bazujące na **systemach funkcji iterowanych IFS** (M.F. Bransley i A.D. Sloan)
- wyszukiwanie **podobieństw pomiędzy fragmentami obrazu** (pomiędzy różnymi fragmentami obrazu lub podfragmentami danego fragmentu obrazu)

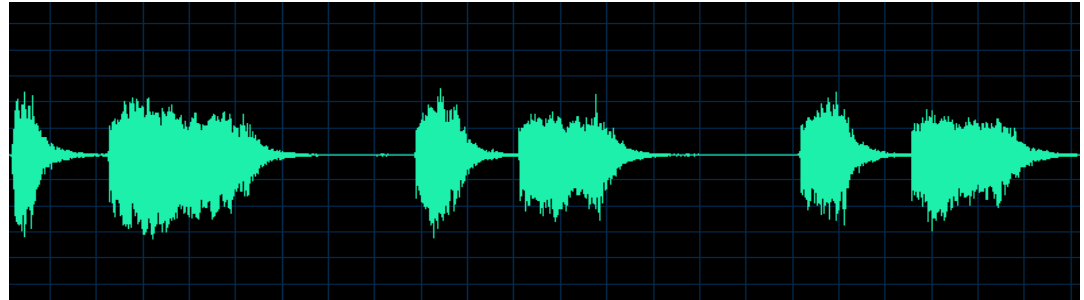
Stopień tolerancji przy wyszukiwaniu podobieństw decyduje o skuteczności oraz stratności metody (wpływa na stopień kompresji oraz na jakość skompresowanego obrazu)

3.2. Kompresja stratna dźwięku

- **Techniki różnicowe** - kodowanie przyrostu (różnicy) wartości sygnału w kolejnym próbkach
 - **DPCM** (*Differential Pulse Code Modulation*) - stała długość informacji różnicowej, np. 1/2 słowa
 - **ADPCM** (*Adaptive DPCM*) - predykcja sygnału i dynamiczne określanie liczby bitów do zapisu błędu
- **Wady metod różnicowych**
 - nie nadążają za szybkimi zmianami (DPCM)
 - nawarstwianie się błędów - konieczność okresowego wstawiania wartości pełnych

Kompresja mowy

- Usuwanie ciszy
- Obniżenie jakości
 - mono, 8 kHz, 8 bit (dynamika ok. 50 dB)
 - *joint stereo* - zapis mono dla niskich częstotliwości
- Skala logarytmiczna
 - **U-LAW** (CCITT G.711) - zapis dźwięku o rozdzielczości 12 bit w skali logarytmicznej na 8 bitach:
 - liczby zmiennoprzecinkowa
 - dynamika ok. 90 dB
 - **A-LAW** - rozszerzenie U-LAW w telefonii europejskiej



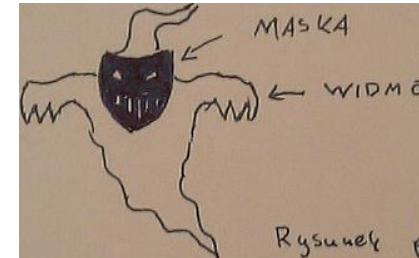
Kompresja psychoakustyczna

- **Modele psychoakustyczne** - zbiór założeń odnośnie percepcji dźwięków polifonicznych przez „przeciętnego” odbiorcę (np. efekt *maskowania*)
 - MPEG-Audio, Ogg-Vorbis, inne
- **Maskowanie** - efekt ograniczonej percepcji pewnych, nakładających się składowych dźwięku w pewnym okresie czasu, występujący w wyniku niedoskonałości słuchu człowieka:
 - maskowanie równoległe (w dziedzinie widma)
 - maskowanie nierównoległe (w dziedzinie czasu)
- **Ton maskujący** - silna składowa sygnału akustycznego powodująca ograniczenie percepcji innych składowych sygnału

Maskowanie widmowe

- Założenia **modelu psychoakustycznego**

- silna składowa S_0 sygnału akustycznego (dźwięku) o częstotliwości f_0 i natężeniu A_0 powoduje ograniczenie (uniemożliwienie) percepcji pozostałych, równocześnie występujących składowych S_i sygnału o częstotliwościach zbliżonych do f_0 ($f_i \rightarrow f_0$) i o niższych natężeniach ($A_i \ll A_0$)

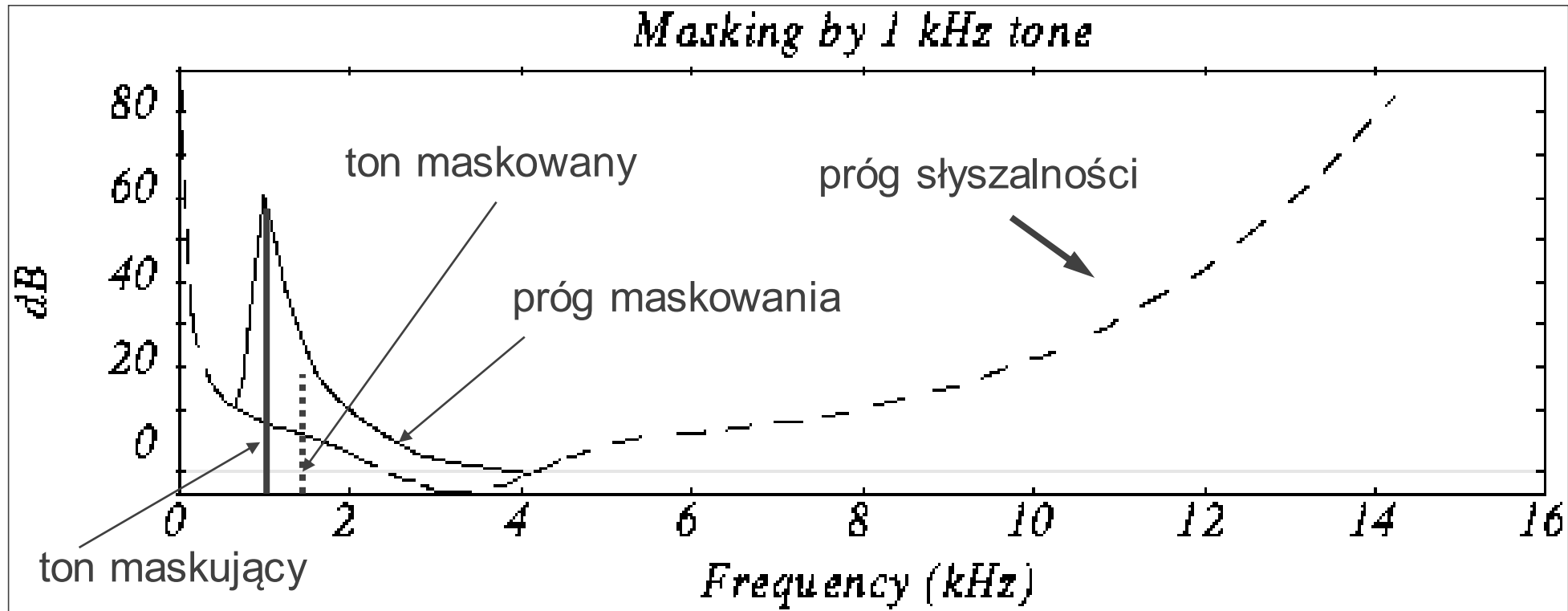


- **Maskowanie widmowe** (równoczesne)

- usunięcie z sygnału akustycznego wszystkich jego składowych S_i , których percepcja, zgodnie z przyjętym modelem psychoakustycznym, jest ograniczona przez równoczesne występowanie silniejszych składowych sygnału

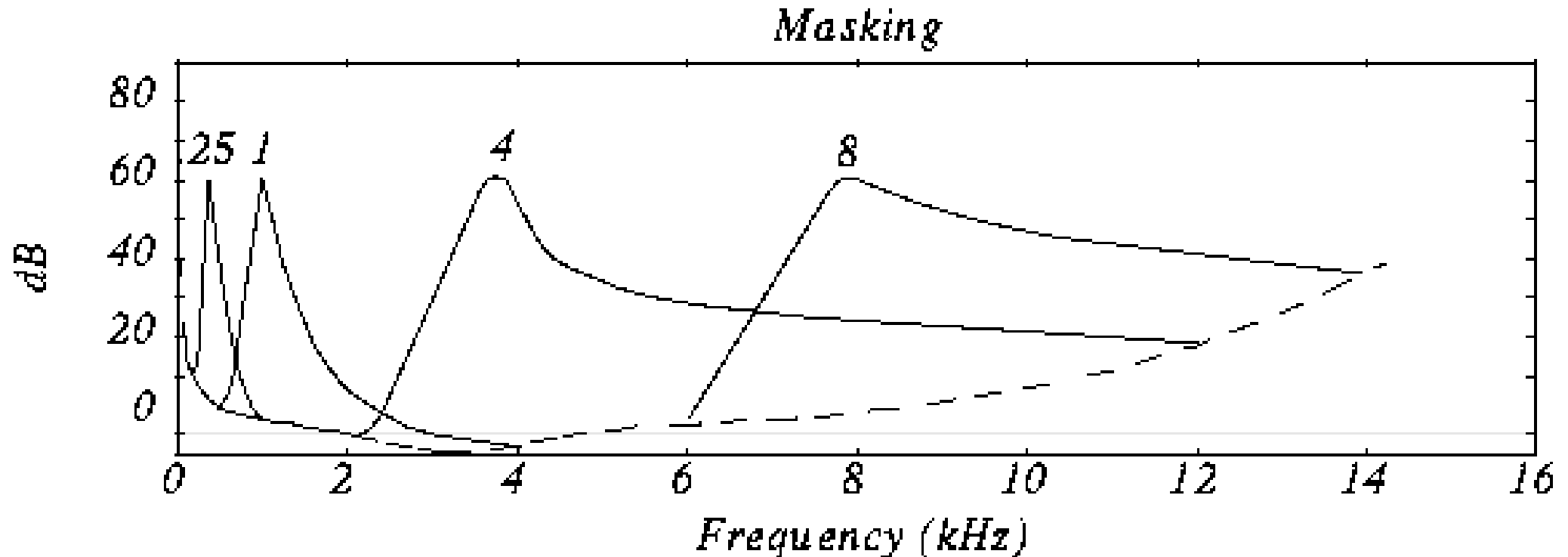
Maskowanie widmowe - przykład

- Ton maskujący o częstotliwości 1 kHz i natężeniu 60 dB powoduje ograniczenie percepcji słabszych dźwięków o zbliżonej częstotliwości mimo, iż dźwięki te byłyby słyszalne, gdyby różnica ich częstotliwości była większa!



Maskowanie widmowe - zmienność

- Maskowanie widmowe przy różnych częstotliwościach tonu maskującego:
 - 250 Hz, 1 kHz, 4 kHz, 8 kHz
- Siła efektu maskowania zależy od częstotliwości tonu maskującego i maskowanego



Maskowanie czasowe

- Założenia ***modelu psychoakustycznego***

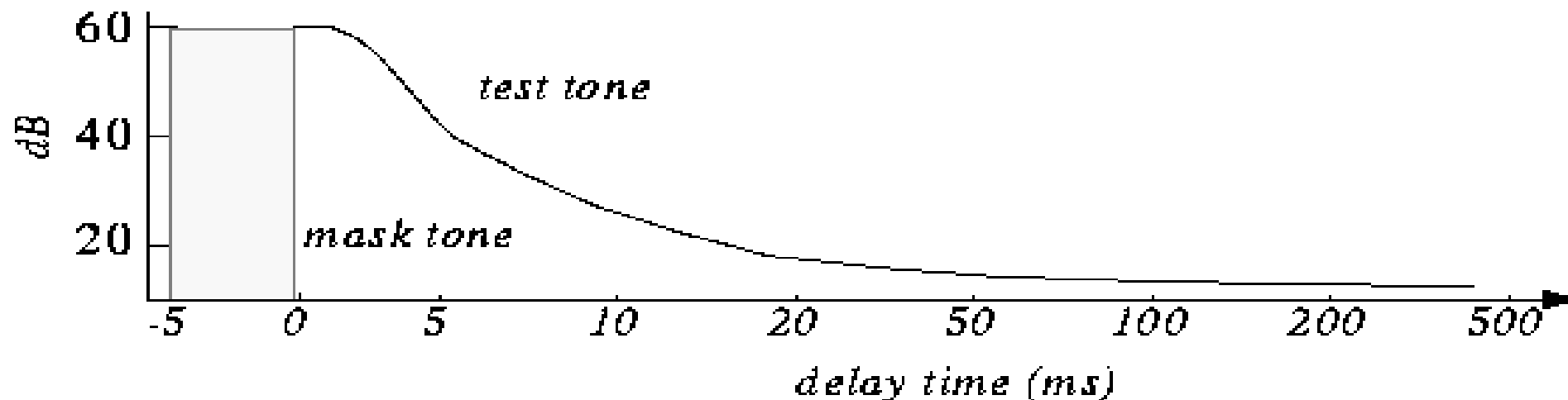
- silna składowa S_0 sygnału akustycznego (dźwięku) o częstotliwości f_0 i natężeniu A_0 przez pewien okres czasu Δt po swoim zakończeniu powoduje ograniczenie (uniemożliwienie) percepcji pozostałych, występujących w tym okresie składowych S_i sygnału o częstotliwościach zbliżonych do f_0 ($f_i \rightarrow f_0$) i o niższych natężeniach ($A_i \ll A_0$)

- ***Maskowanie czasowe*** (nierównoczesne)

- usunięcie z sygnału akustycznego wszystkich jego składowych S_i , których percepcja, zgodnie z przyjętym modelem psychoakustycznym, jest ograniczona przez wcześniejsze wystąpienie silniejszych składowych sygnału

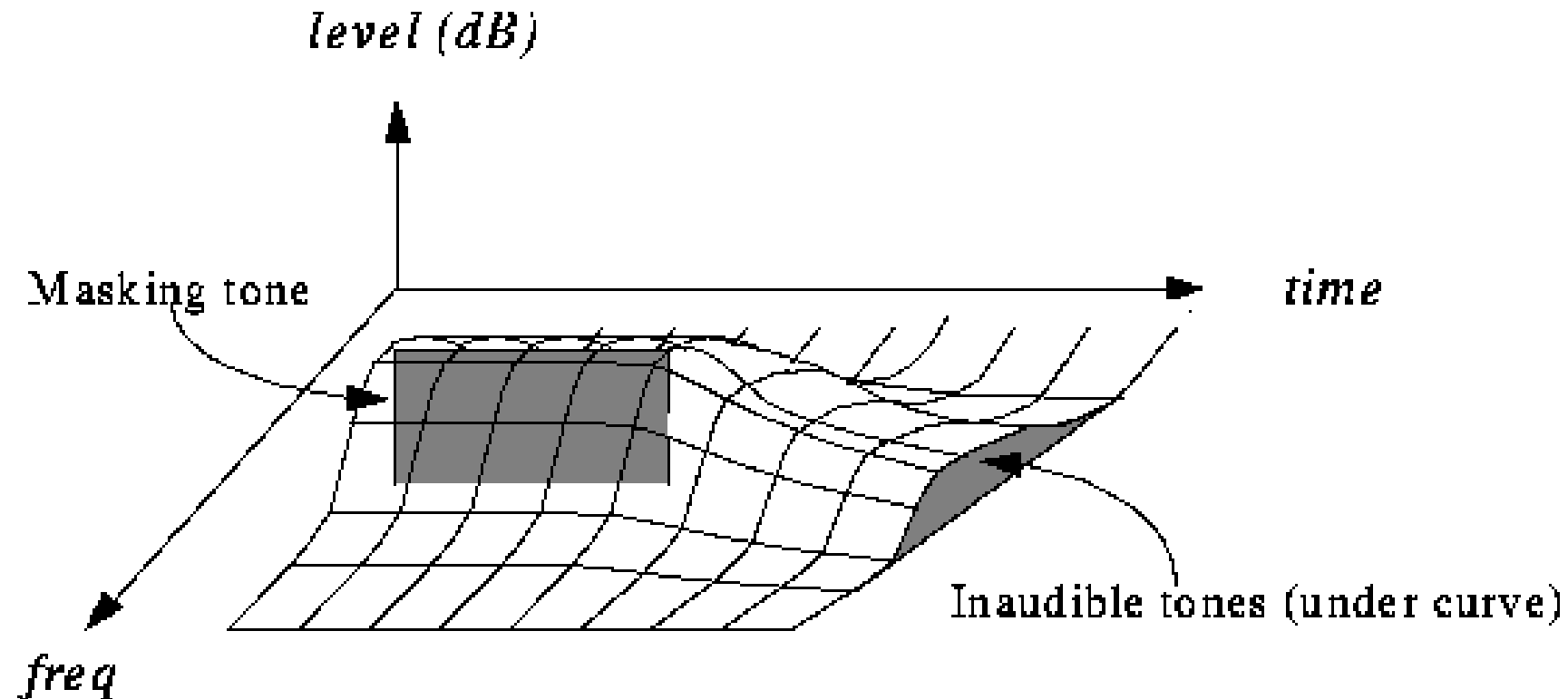
Maskowania czasowe - przykład

- Ton maskujący o częstotliwości 1 kHz i o natężeniu 60dB zanikający w chwili t_0 maskuje przez pewien okres czasu ton o częstotliwości 1.1 kHz o niższym natężeniu.
- Siła efektu maskowania czasowego zależy od częstotliwości tonu maskującego i czasu, jaki upłynął od zaniku tonu maskującego



Maskowania widmowe i czasowe

- Maskowanie widmowe związane jest z ograniczoną zdolnością percepcji polifonicznej.
- Maskowanie czasowe związane jest z bezwładnością słuchu.



MPEG-1 - audio

- Model psychoakustyczny percepcji człowieka
- Layer I i II: 260÷390 kbit/s (4-8:1)
- Layer III \equiv MP3: 64÷448 kbit/s stereo
- 256 kbit/s jakość zbliżona do CD
- kompresja $\sim 4\div 96$ (12:1 przy 128 kbit/s)
- Tryb CBR (*constant bitrate*) - stała wartość bitrate (prostszy algorytm i mniejsza złożoność obliczeniowa)
- Tryb VBR (*variable bitrate*) - uśredniona wartość *bitrate* (chwilowa wartość BR zależy od stopnia kompresowalności ramki)

Etapy kompresji MP3

- Podział sygnału na ramki ~25ms
- DCT ramki
- Podział na 32 pasma częstotliwości i usunięcie skrajnych
- Usuwanie (maskowanie równ.) szumów i słabych sygnałów
- Przydział bitów dla poszczególnych pasm
- Ustalenie wymaganej liczby bitów dla ramki
- Ustalenie maksymalnej liczby bitów przypadających na ramkę na podstawie zadanego parametru (CBR – constant bitrate)
- Wyznaczenie ilości informacji, która musi zostać odrzucona
- Kompresja stratna sygnału na podstawie przyjętych parametrów (w tym maskowanie czasowe i joint stereo) - zaoszczędzone miejsce można wykorzystać na nadmiarowe dane z następnej ramki
- Kompresja strumienia danych metodą Huffmana
- Poprzedzenie ramek nagłówkami i scalenie

3.3. Kompresja wideo

- Celem kompresji wideo jest zwarta reprezentacja sekwencji obrazów cyfrowych
- Możliwość wykorzystania znacznego wzajemnego podobieństwa kolejnych obrazów sekwencji wideo
- Rodzaje technik kompresji
 - technika typu *intra*
 - bazuje na informacji tylko z bieżącej ramki
 - technika typu *inter*
 - wykorzystuje informację z innych obrazów sekwencji
 - technika mieszana (*inter+intra*)
 - stosowanie techniki *intra* dla ramek kluczowych oraz techniki *inter* dla pozostałych

Wyzwania kompresji filmów

- Swobodny dostęp do każdej klatki
- Synchronizacja obrazu i dźwięku
- Szybkie przeszukiwanie w przód i wstecz
- Odtwarzanie wstecz
- Edytowanie strumienia bitów
- Elastyczność formatu obrazu
- Skuteczna korekcja błędów
- Kodowanie w czasie rzeczywistym
- Małe opóźnienia kodowania i dekodowania

edycja i
odtworzenie
video

wideokonferencje,
wideotelefony

Zniekształcenia w kompresji wideo

- Przyczyna:
 - odrzucenie zbyt dużej ilości danych, w pliku wideo pojawiają się zniekształcenia – tzw. artefakty
- Efekt - pojawienie się artefaktów
- Objawy:
 - rozmycie
 - schodki
 - obwódki wokół krawędzi
 - posteryzacja
- Sposoby zapobiegania:
 - zwiększenie prędkości bitowej kompresji
 - zwiększenie liczby klatek kluczowych

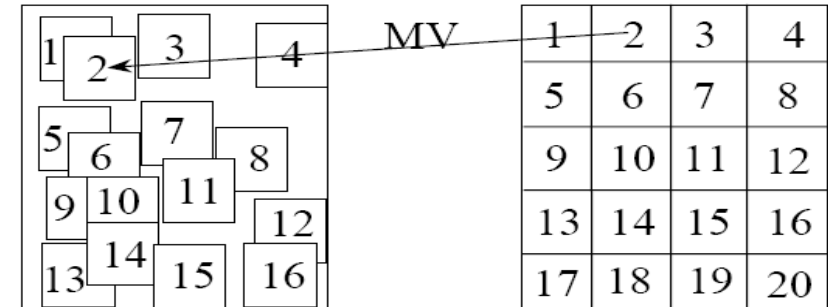
Predykcja, estymacja i kompensacja ruchu

- **Predykcja** (ang. *prediction*) - technika wyznaczania opisu ramki obrazu za pomocą pewnej liczby innych ramek, zwanych referencyjnymi
- **Estymacja ruchu ME** (ang. *Motion Estimation*) - wyznaczenie obszaru R^0 w referencyjnej ramce obrazu I_{ref} , który stanowi najlepsze odniesienie (w sensie wybranego kryterium podobieństwa) dla danego obszaru R^{kod} w bieżącej (kodowanej) ramce obrazu
- **Wektor ruchu MV** (ang. *Movement Vector*) definiuje wzajemne położenie obszarów R^0 oraz R
- **Kompensacja ruchu MC** (ang. *Motion Compensation*) - technika predykcji wykorzystująca estymację ruchu

Blokowa kompensacja ruchu

- Blokowa kompensacja ruchu BMC (ang. *Block Motion Compensation*)

- obraz bieżący dzielony jest na sąsiadujące bloki (np. 16x16)
- dla każdego bloku znajdowany jest najlepiej do niego pasujący blok¹ w ramce referencyjnej
- wyznaczany jest wektor ruchu bieżącego bloku
- jeżeli bloki są do siebie wystarczająco podobne, zamiast bieżącego bloku kodowane są (DCT) jedynie różnice między nim, a blokiem referencyjnym



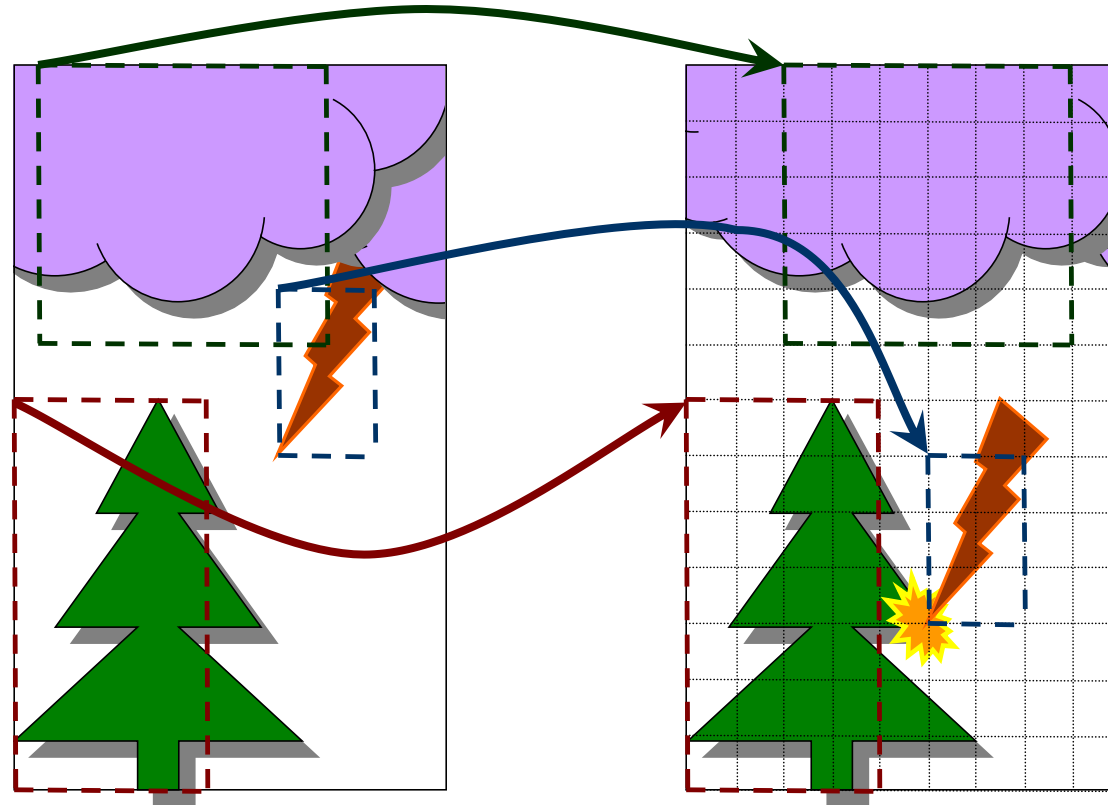
- Wady

- nieciągłości obrazu na granicy bloków (ang. *blocking artifacts*)
- próby usunięcia poprzez stosowanie filtrów wygładzających (ang. *deblocking filters*)

- Uwaga

- bloki referencyjne mogą się nakładać

Blokowa kompensacja ruchu - ilustracja



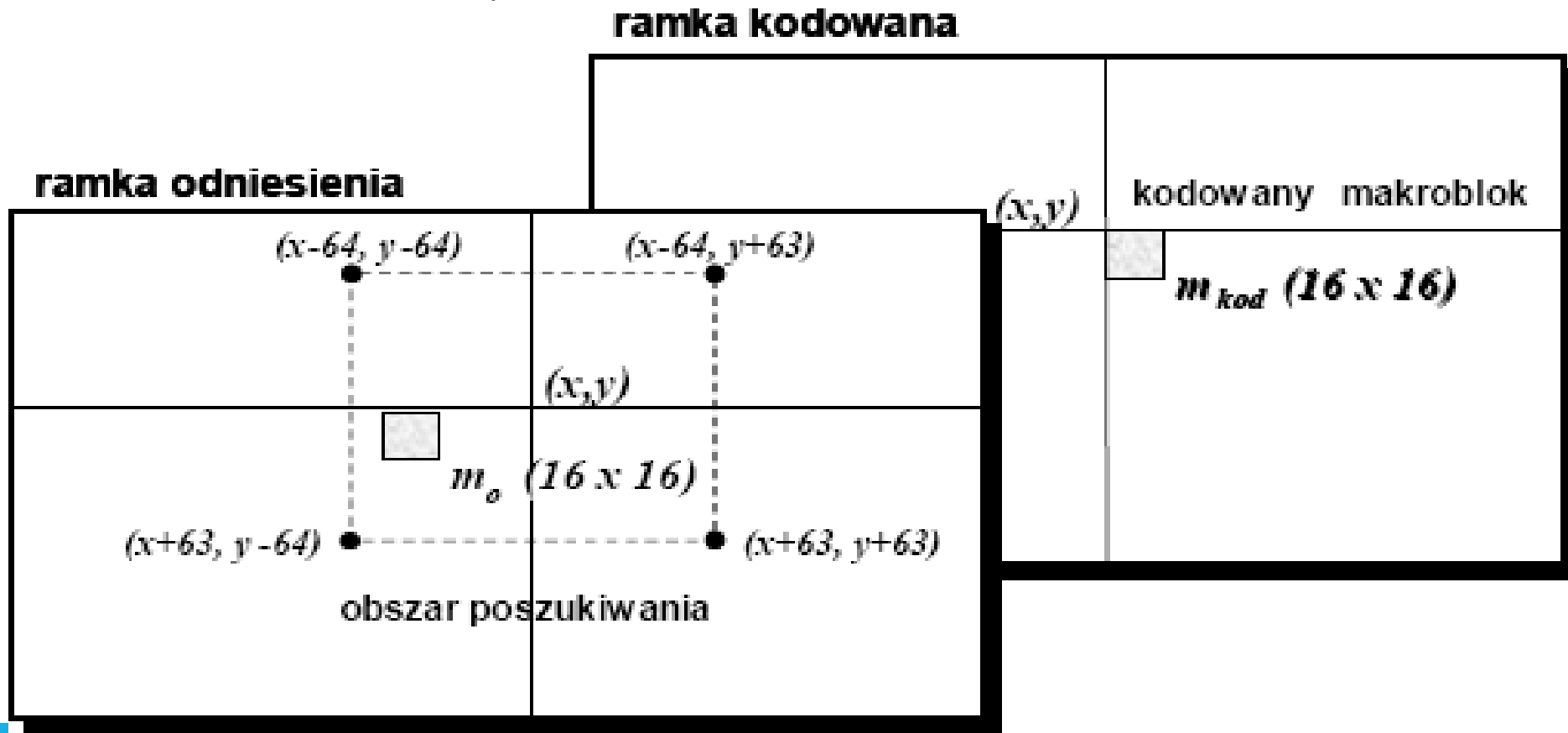
Blokowa kompensacja ruchu - problemy

- Podstawowe problemy
 - rozmiar bloku
 - wielkość
 - możliwość dynamicznej zmiany
 - wybór ramki odniesienia (referencyjnej)
 - liczba ramek: 1, 2, .. 32
 - ramka poprzednia, następna
 - metoda poszukiwania bloku referencyjnego
 - zakres poszukiwania bloku referencyjnego
 - miara podobieństwa bloków
 - uwzględnienie wyłącznie składowej Y lub wszystkich składowych
 - wartość progu podobieństwa
 - metryka podobieństwa
 - „plan B”, gdy brak bloków podobnych
 - rozwiązania których problemów definiować w standardzie?

BMC - przykład rozwiązania

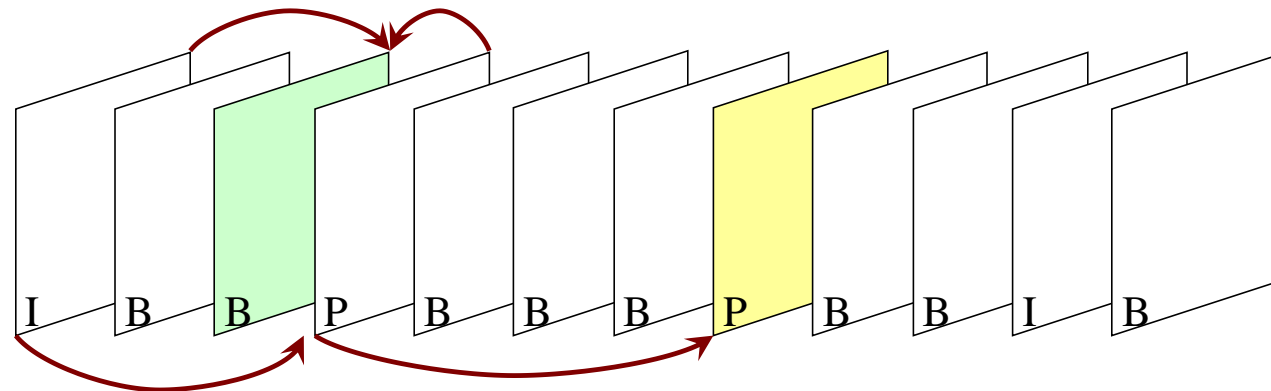
- Miara podobieństwa bloków dla $MV=[u,v]$

$$d(u, v) = \frac{1}{256} \sum_{i=0}^{255} \sum_{j=0}^{255} |m_{kod}(x+i, y+j) - m_o(x+i+u, y+j+v)|$$



Rodzaje ramek

- Ramki w kompresji wideo:
 - I (ang. *Intra*) – kodowane niezależnie od pozostałych ramek
 - najmniejsza kompresja
 - punkty swobodnego dostępu do nagrania wideo (klatki kluczowe)
 - P (ang. *Predicted*) - ramki przewidywane na podstawie poprzedniej ramki I lub P
 - B (ang. *Bidirectional*) – interpolowane (z wagami) z 2 najbliższych (sąsiednich) ramek typu I lub P
 - najwyższa kompresja, dopuszczalna niska jakość kompresji (błędy nie propagują się)

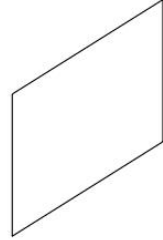


Wzorce ramek IPB

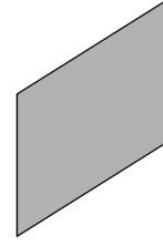
- Różnorodne wzorce zawartości ramek I, P, B
 - zwiększenie liczby ramek I (P) zwiększa jakość wideo oraz możliwości swobodnego dostępu do wideo
 - zwiększenie liczby ramek B obniża jakość wideo
 - ramka I przynajmniej 1 raz na 132 ramki (zalecane min. co 0.4s)
- Przykładowe sekwencje
 - **I I I I I I I I I I I I I I I I** (bez predykcji \approx M-JPEG)
 - **I P I P I P I P I P I P I**
 - **I B P B I B P I B P B I**
 - **I B B P B B P B B P B B I**
 - **I B B B I B B B I B B B I**

Rodzaje predykcji

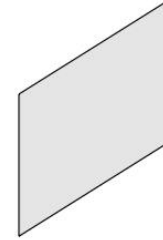
Ramka I



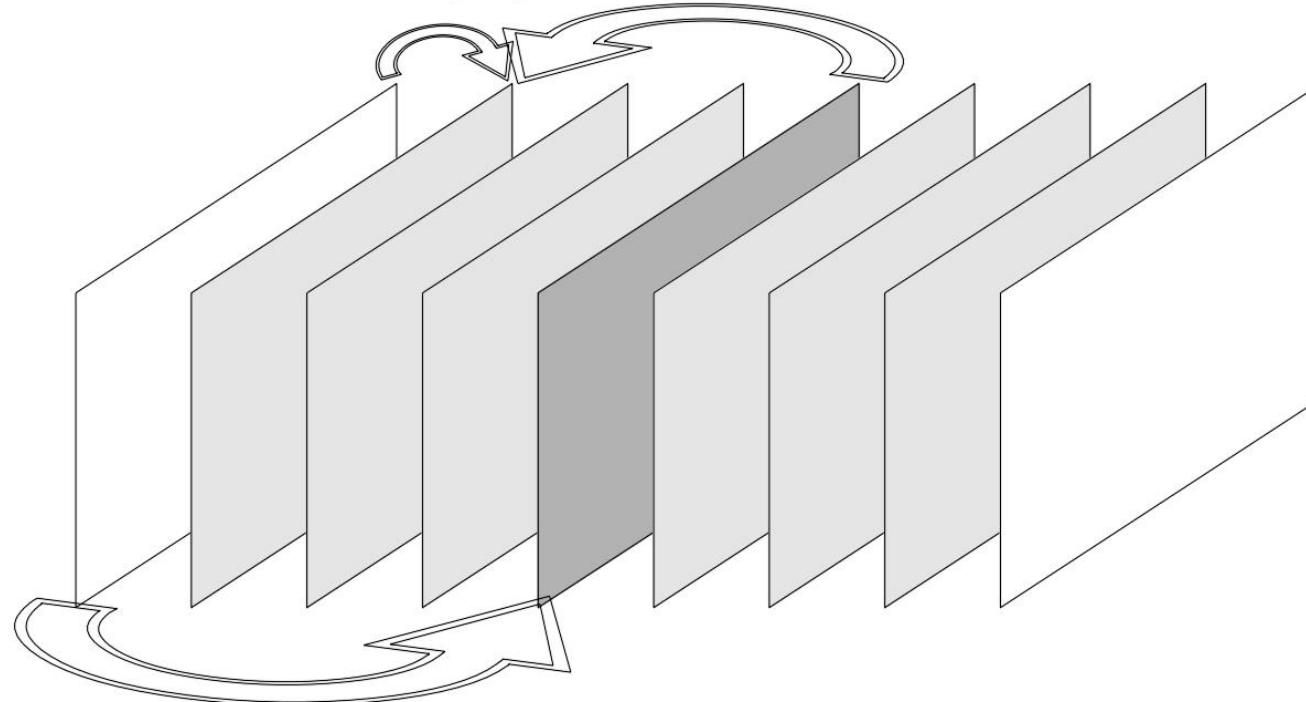
Ramka P



Ramka B



Predykcja dwukierunkowa

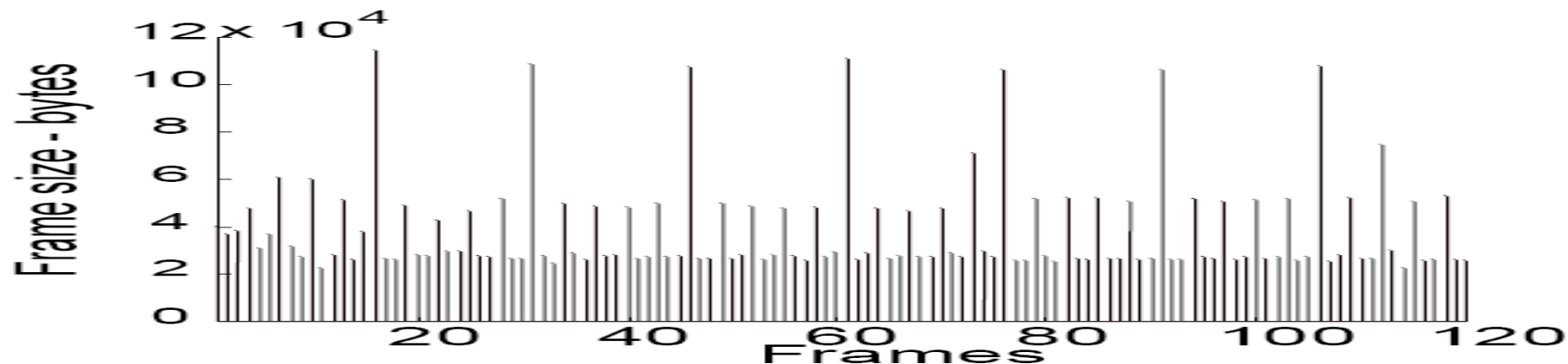


Predykcja w przód

Rozmiar ramek

- Przykładowe statystyczne dane pomiarowe (kodowanie VCD) dla sekwencji I B B P B B P B B

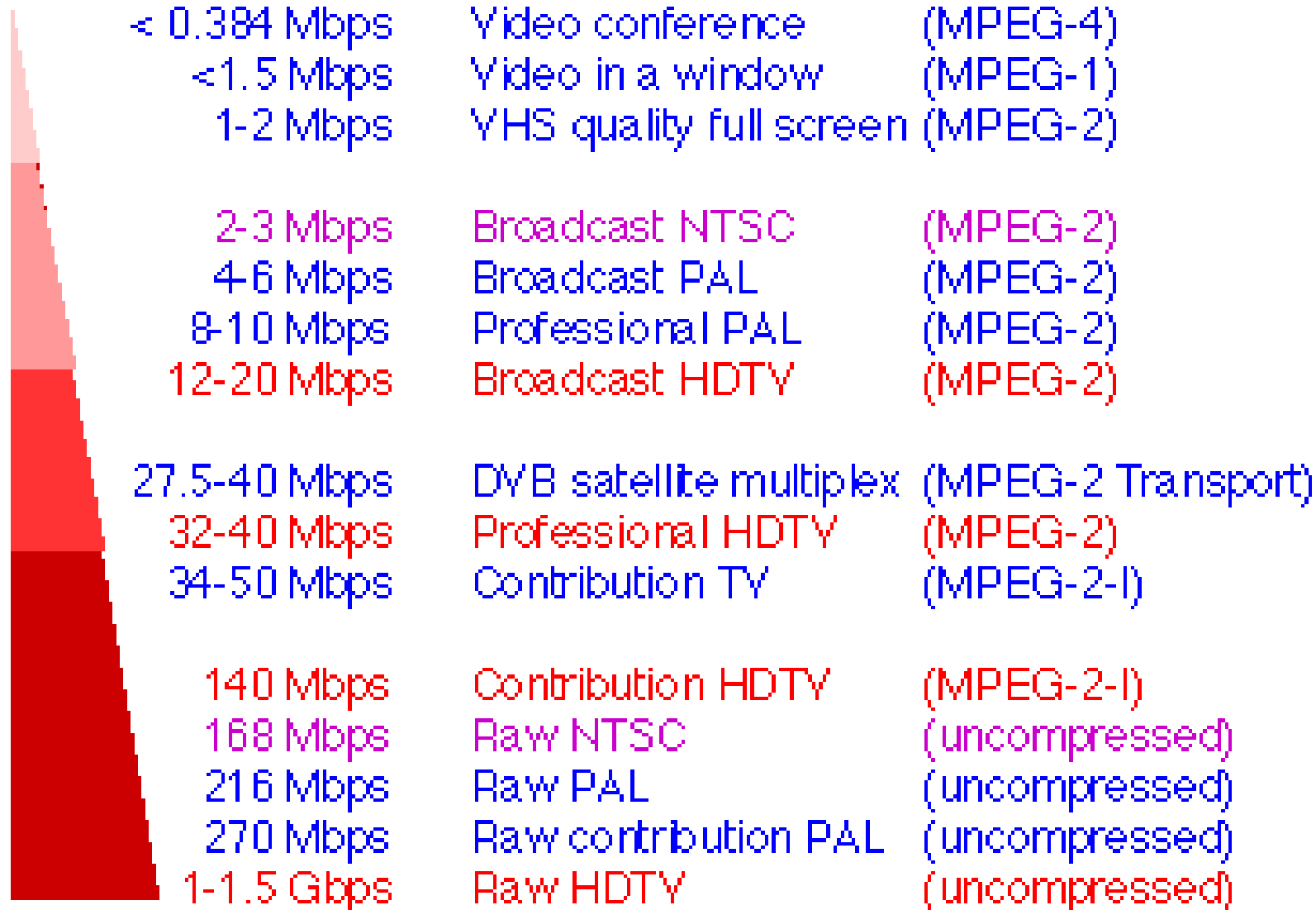
Pomiar zajętości pasma	bez rozróżniania [B/frame]	I [B/frame]	P [B/frame]	B [B/frame]	łącznie sekwencja [B/sequence]	M-JPEG [B/frame]
średni	15 598	60 381	23 069	7 199	187 185	222 330
minimalny	476	11 754	2 192	476	77 754	68 976
maksymalny	185 267	185 267	174 294	64 785	932 710	627 672



3.5.2. Generacje algorytmów kompresji

Generacja	Algorytmy	Uwagi
wewnątrz-obrazowe	M-JPEG, M-JPEG 2000, DV, DVPRO	niezależne kodowanie każdej klatki
1. martwe	H.120, CCIR 723	
2. przestarzałe	H.261, MPEG-1	kodowanie hybrydowe
3. powszechne	MPEG-2/H.262, H.263/H.263+/H.263++ MPEG-4 (DivX, Xvid)	kodowanie hybrydowe, zmienna wielkość bloków, zwiększona dokładność, skalowalność
4. nowe	MPEG-4 AVC/H.264, WMP/VC1, H.265	j.w. + zaawansowane techniki

Przepływność danych



Grupy robocze ISO

- Komisja normalizacyjna ISO (ang. *International Organisation for Standardisation*)
 - grupy robocze komitetu technik komunikacyjnych JTC1, zajmujące się metodami kompresji danych multimedialnych
 - **JPEG** (ang. *Joint Photographic Experts Group*) - cyfrowa kompresja i kodowanie obrazów: JPEG (JFIF), JPEG-2000
 - **MPEG** (ang. *Moving Picture Experts Group*) - cyfrowa kompresja i kodowanie wideo: MPEG-1, MPEG-2, MPEG-4
 - **MHEG** (ang. *Multimedia and Hypermedia Information Coding Experts*) - standard kodowania i prezentacji danych multimedialnych i hipermedialnych: MPEG-7, MPEG-21
 - **JVT** (ang. *Joint Video Team*) połączona grupa MPEG oraz VCEG (ang. *Video Coding Experts Group*) z ITU-T (ang. *International Telecommunications Union - Telecommunication Standardisation Sector*) - standard kodowania MPEG-4 AVC (Part 10) \equiv H.264

Zastosowania standardów kompresji

Standard	Przeznaczenie	Format danych wej.
JBIG	Statyczne obrazy binarne	dowolny
JPEG	Statyczne obrazy wielopoziomowe	dowolny
H.261	Wideokonferencje, wideotelefonía	CIF, QCIF
H.263	Wideotelefonía	QCIF
MPEG-1	Archiwizacja na nośnikach DSM	SIF
MPEG-2	Emisja telewizyjna, HDTV, VoD	BT.601
MPEG-4	Komunikacja interaktywna	CIF, SIF, BT.601

Kompresja MPEG

- MPEG – *Moving Pictures Expert Group* - standaryzacja w dziedzinie audio i wideo
 - strona domowa: <http://www.chiariglione.org/mpeg/>
 - odpłatna dokumentacja standardów : <http://www.iso.org/>
- Standardy kompresji zawartości multimedialnej:
 - MPEG-1 (1991 r.): VideoCD, monitoring TV
 - MPEG-2 (1995 r.): DVD, HDTV, MP3
 - MPEG-4 (1998 r.): VoD, GSM 3G, studia cyfrowe
 - MPEG-4 p.10 AVC (Advanced Video Coding) = H.264 (2004)
 - MPEG-H HEVC (High Efficiency Video Coding) = H.265 (2013)
 - MPEG-I p.3 VVC (Versatile Video Coding) = H.266 (2020)
- Standardy opisu obiektów multimedialnych:
 - MPEG-7 (2003+ r.): opis zawartości audiowizualnej
 - MPEG-21 (2004 r.): struktury rozproszonych systemów multimedialnych

3.6. MPEG-1

- MPEG-1 – standard kompresji wideo
 - norma ISO 11172 – 1991 r. (rozwijana do 1999 r.)
- Zastosowanie – zapis wideo na płytach CD
 - zapis filmów na płytach CD - standard Video-CD (*White Book*)
- Podstawowe parametry profilu CPB (ang. *Constrained Parameters Bitstream*)
 - obraz bez przeplotu
 - rozdzielczość obrazu SIF (ang. *Source Input Format*)
 - systemy PAL, SECAM: 352x288 (25fps)
 - system NTSC: 352x240 (30 fps)
 - częstotliwość ramek: 24-30 fps
 - dokładność kwantyzacji wideo 8-bit
 - przepływność wideo (razem z audio) do 1.5 Mb/s
 - przepływność audio do 192 kb/s (224kb/s)

Dokumenty MPEG-1

- Dokumenty opisujące standard (odpłatne)
 - 11172-1: Systems
 - struktura pakietowa do łączenia (multipleksowania) kodowanych danych audio i video w jeden, zsynchronizowany w czasie strumień bitowy
 - 11172-2: Video
 - kompresja wideo
 - 11172-3: Audio
 - kompresja audio
 - 11172-4: Compliance testing
 - procedury i opis testów zgodności
 - 11172-5: Software simulation
 - oprogramowanie referencyjne, pełna implementacja pierwszych trzech części standardu MPEG-1

MPEG-1 - system

- Norma ISO 11172-1: Systems (Part 1)
- Podział video i audio na pakiety
- Znaczniki czasu - synchronizacja video i dźwięku
- Parametry zapisu:
 - długość pakietu
 - prędkość transmisji danych
 - rozdzielczość obrazu
 - rodzaj nośnika informacji ...

3.6.1. Charakterystyka

- Prędkość transmisji 1.5 Mb/s
 - DS1 (ang. *Digital Service 1*)
 - CD-ROM, taśmy DAT
- Rozdzielczość SIF odpowiadająca VHS ($\frac{1}{4}$ PAL)
 - Y: 720x576 \rightarrow 352x288
 - $C_b C_r$: 360x576 \rightarrow 176x144
- Stopień kompresji 100-200:1
- Wymagana moc obliczeniowa (algorytm asymetryczny)
 - kodowanie 1041 MIPS
 - dekodowanie 58 MIPS

Struktura sekwencji wideo

- Sekwencja wideo – ciąg kolejnych ramek (obrazów)
- Grupa obrazów GOP (ang. *Group of Pictures*)
 - rodzaje ramek
 - I (*Intra*) – ramka kodowana niezależnie od pozostałych
 - P (*Predictive*) – kodowanie na podstawie poprzedniej ramki I lub P
 - B (*Bidirectional predictive*) – kodowanie ważone na podstawie dwóch najbliższych ramek I lub P
 - D (*Direct*) - każdy blok zawiera wyłącznie informację o średnim poziomie jasności (składowa DC)
 - struktura
 - pojedyncza ramka I
 - elastyczny dobór proporcji ramek P i B w GOP
 - maksymalna długość - 132 ramki

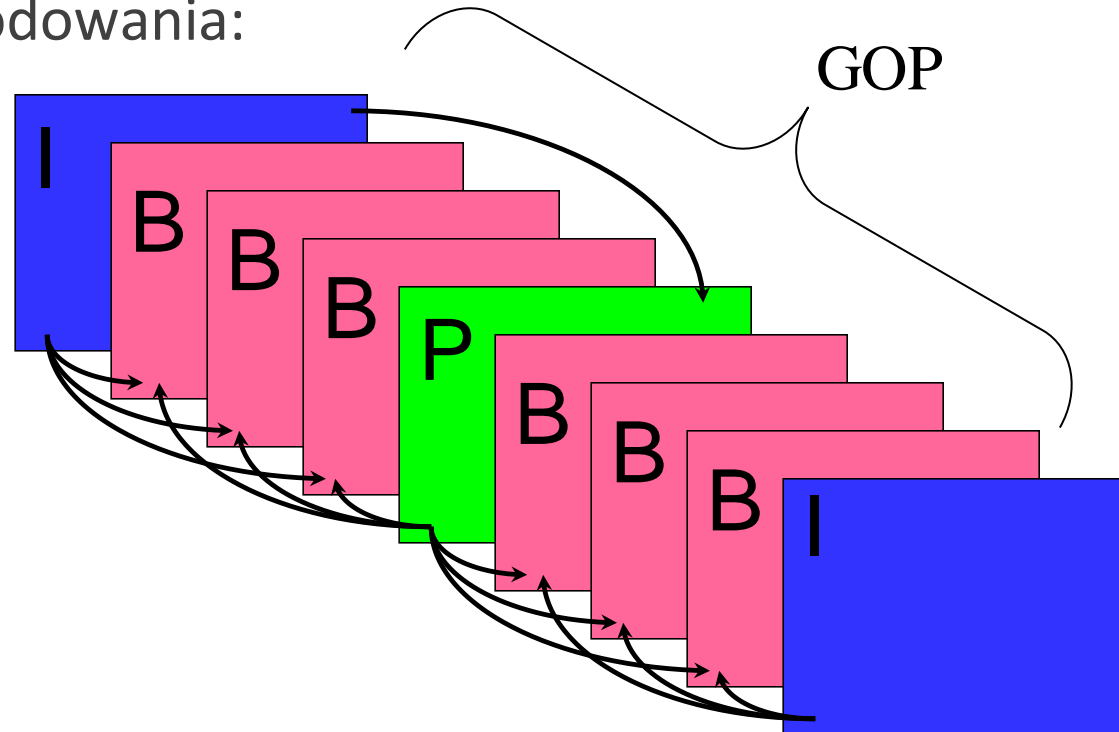
Kolejność ramek w GOP

- Przykładowa sekwencja ramek:

I B B B P B B B I

Kolejność na wyjściu kodowania:

I P B B B I B B B



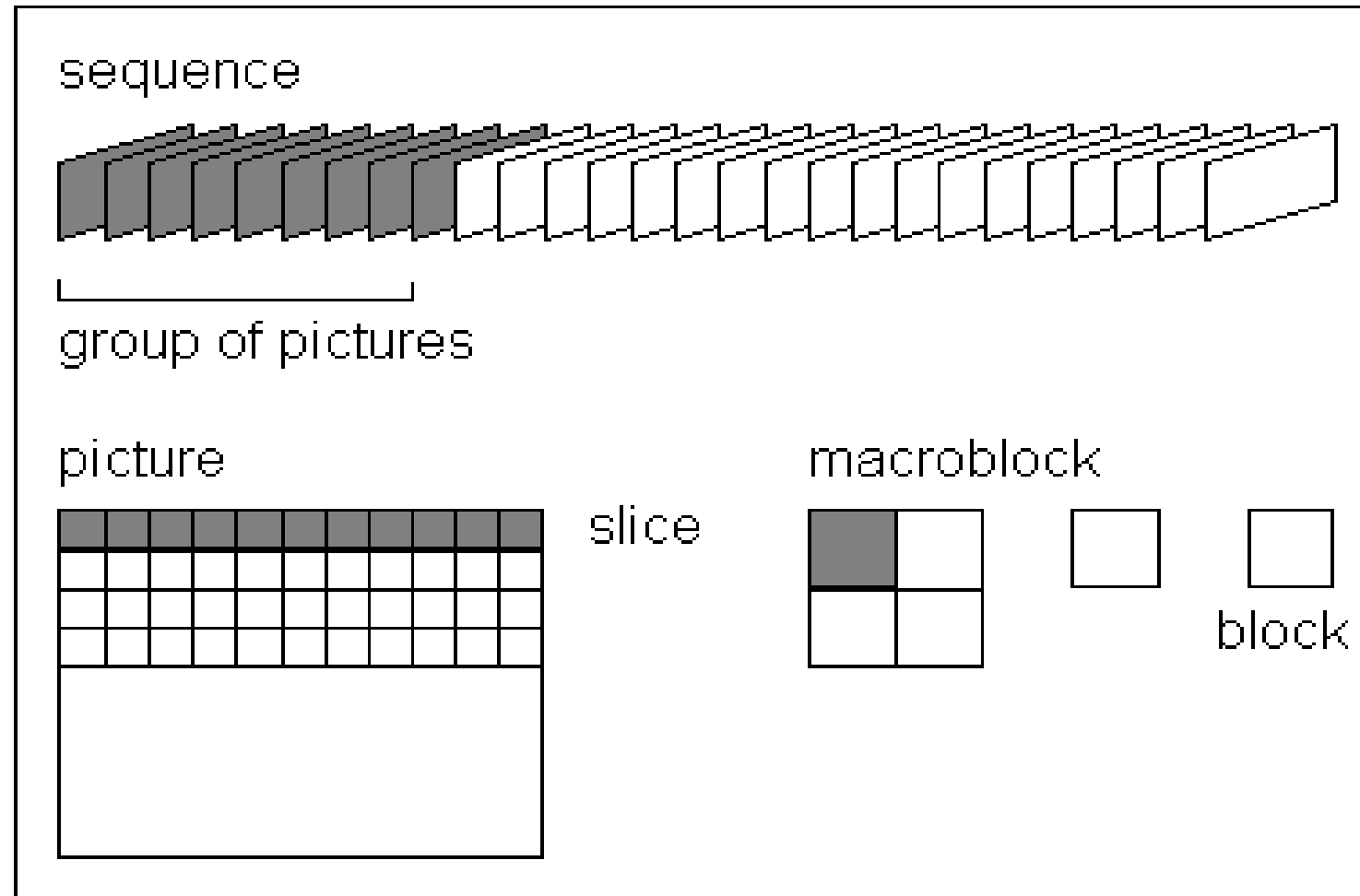
Konsekwencje stosowania ramek typu B

- Zmiana kolejności ramek w zakodowanym sygnale
- Wprowadzenie znaczników czasowych
 - DTS (ang. *Decoding Time Stamp*)
 - PTS (ang. *Presentation Time Stamp*)
- Wprowadzenie buforów dekodera na jeszcze nie wyświetlone ramki I lub P

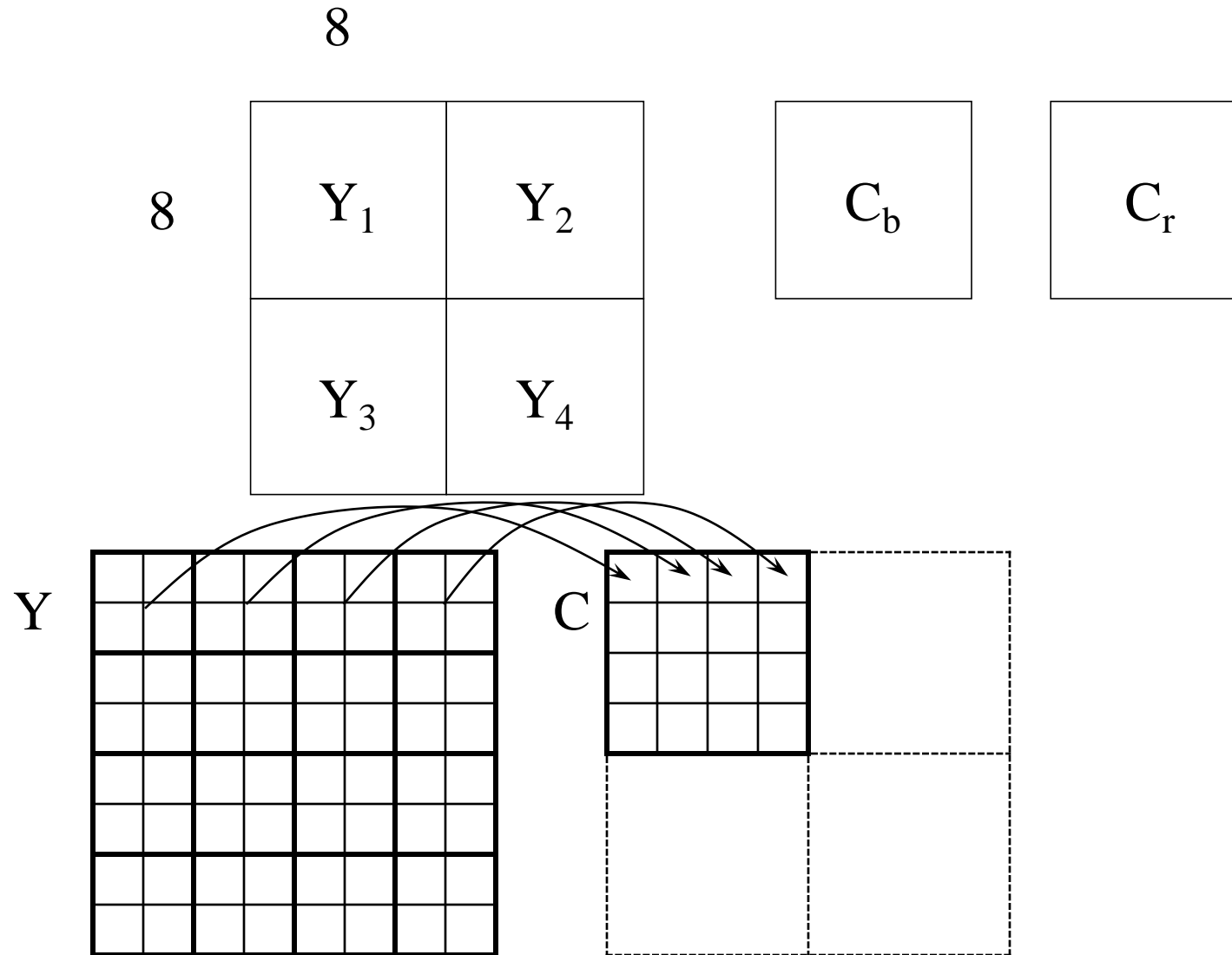
Struktura ramki obrazu

- Rozmiar ramki SIF - 352×288 (PAL)
 - model barw YC_bC_r
 - rzeczywista rozdzielczość kanałów chrominancji - 176×144
- Przekroje/warstwy (ang. *slices*) - sekwencja makrobloków o tych samych parametrach kompresji
- Makrobloki o rozmiarze 16×16
 - liczba makrobloków w każdej ramce - 396 (22×18)
 - budowa
 - 4 bloki 8×8 dla kanału luminancji Y
 - 1 blok dla kanału chrominancji C_b
 - 1 blok dla kanału chrominancji C_r

Struktura ramki



Struktura makrobloku 4:2:0



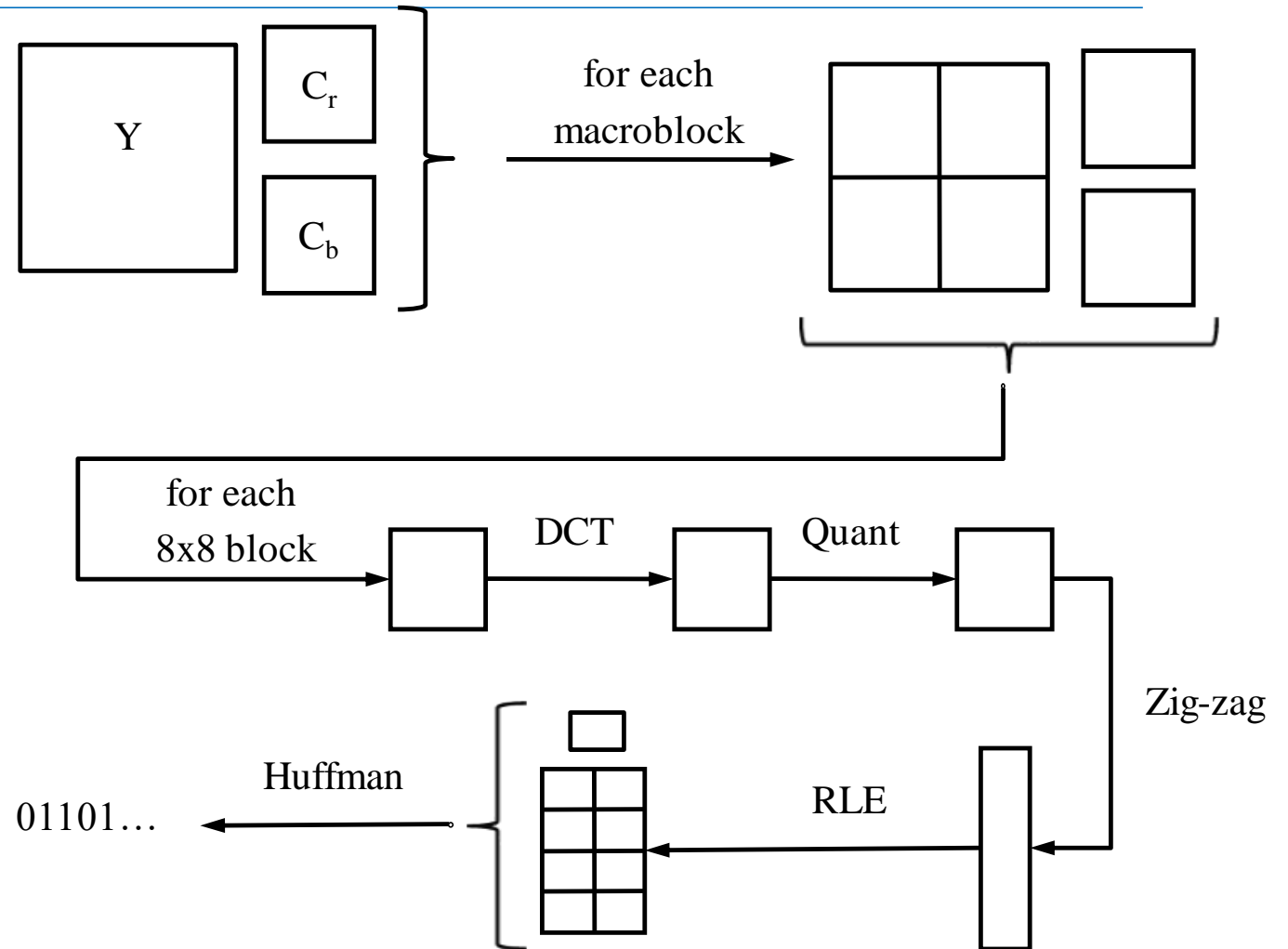
3.6.2. Algorytmy kodowania i dekodowania

- Kodowanie oparte o kompensację ruchu oraz techniki wykorzystywane w JFIF-JPEG
 - ramki I - operacje na oryginalnych wartościach pikseli
 - ramki P i B - operacje na różnicach wartości przewidywanych i rzeczywistych (możliwe również na oryginalnych wartościach)
 - różne tablice kwantyzacji dla ramek I, P i B
 - mnożnik kwantyzacji umożliwia zmianę wielkości strumienia (zmniejszenie kosztem jakości)
 - oznaczanie bloków z zerowymi współczynnikami
 - predefiniowane kody Huffmana (koder entropijny)

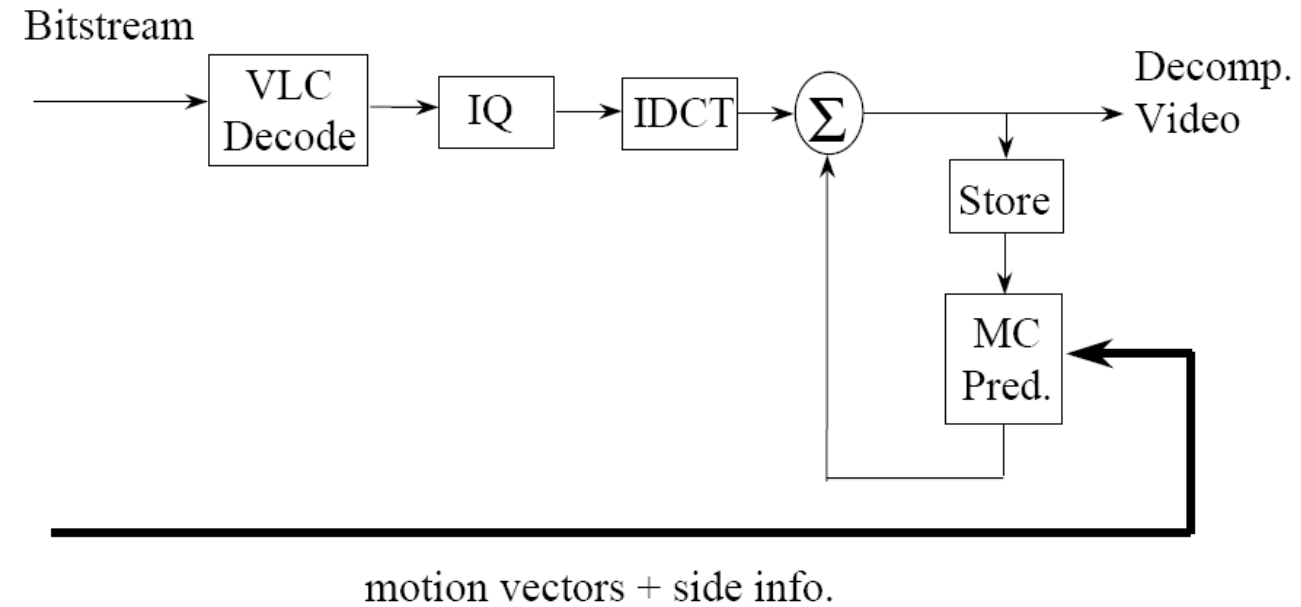
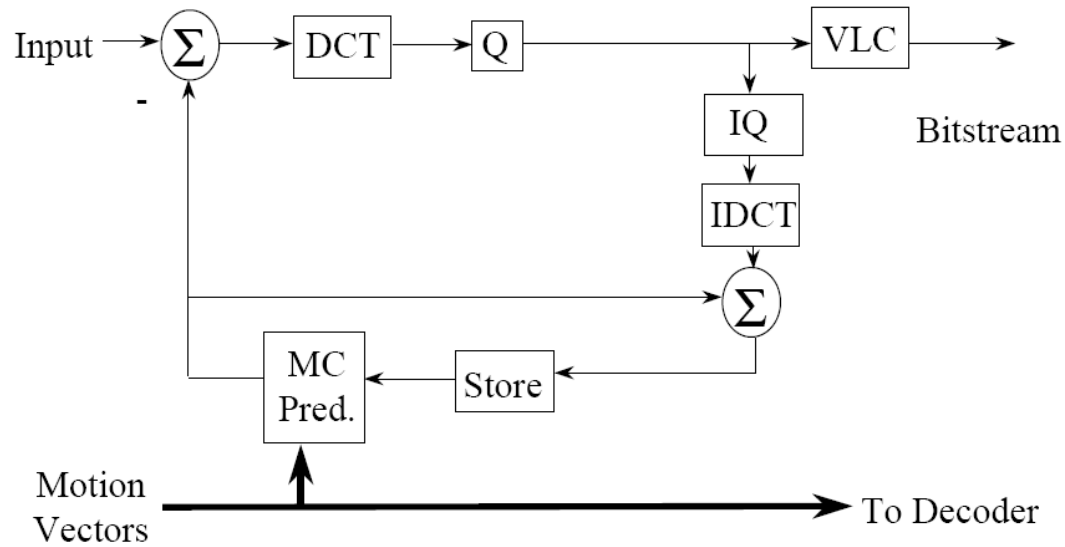


Kompresja ramek I

- Podział na makrobloki
- Transformacja DCT dla bloków
- Kwantyzacja
- Skanowanie
- Kodowanie



Koder i dekodery MPEG-1 (ramki P i B)



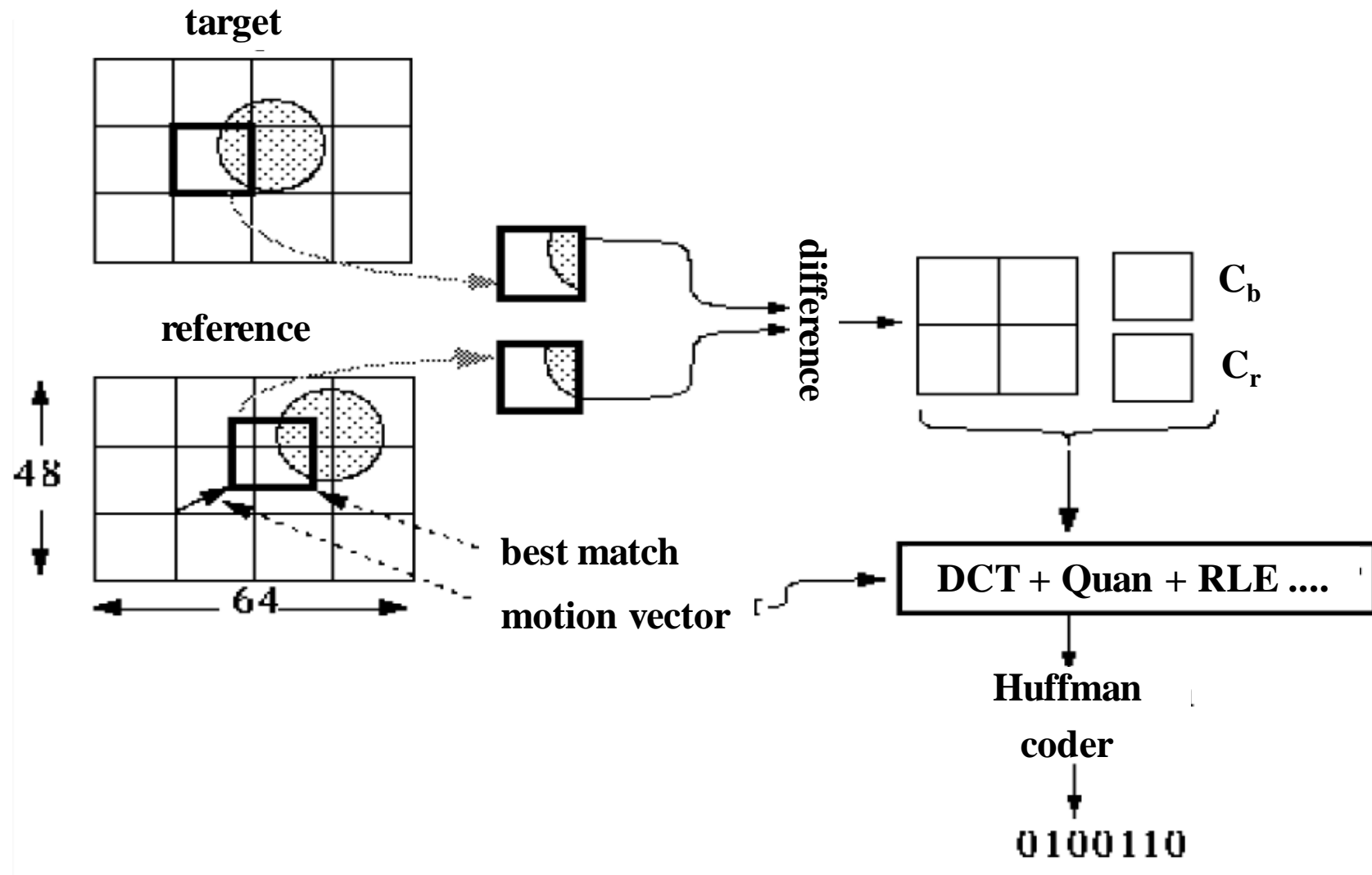
Predykcja

- **Predykcja** - ekstrapolacja (przewidywanie) wyglądu obrazu na podstawie obrazów referencyjnych (wcześniejszych lub późniejszych)
- MPEG-1 - predykcja na podstawie estymacji ruchu
 - jednostka estymacji ruchu - makroblok
 - dla danego makrobloku poszukiwany jest w obrazie odniesienia blok najbardziej do niego zbliżony
 - standard nie definiuje sposobu wyszukiwania podobnych makrobloków!
 - metoda *brute force* ma dużą złożoność obliczeniową
 - stosowane są różne metody suboptymalne najczęściej oparte wyłącznie o kanał luminancji
 - wektor ruchu - informacja o względnym położeniu obszaru referencyjnego w ramce referencyjnej
 - możliwość wykrywania wyłącznie translacji (nie jest wykrywany obrót ani skalowanie)

Kompresja ramek P

- Oparta na idei kompensacji ruchu bloków bieżącego obrazu względem bloków obrazu referencyjnego (poprzedniej ramki typu I lub P)
 - obliczana jest różnica między blokiem 8x8 z bieżącego obrazu a najlepiej dopasowanym (wg pewnej miary podobieństwa) blokiem w obrazie referencyjnym
 - do obliczonej różnicy stosuje się transformację DCT
 - oprócz kodu różnicy zapamiętywany jest wektor ruchu – przesunięcie między blokami
 - wektor ruchu wspólny dla wszystkich bloków makrobloku
 - najczęściej wyznaczamy go dopasowując bloki 16x16 leżące tylko w składowej luminancji
 - skanowanie, kwantyzacja i kodowanie
- Możliwość rezygnacji z kompensacji ruchu

Kompresja ramek P



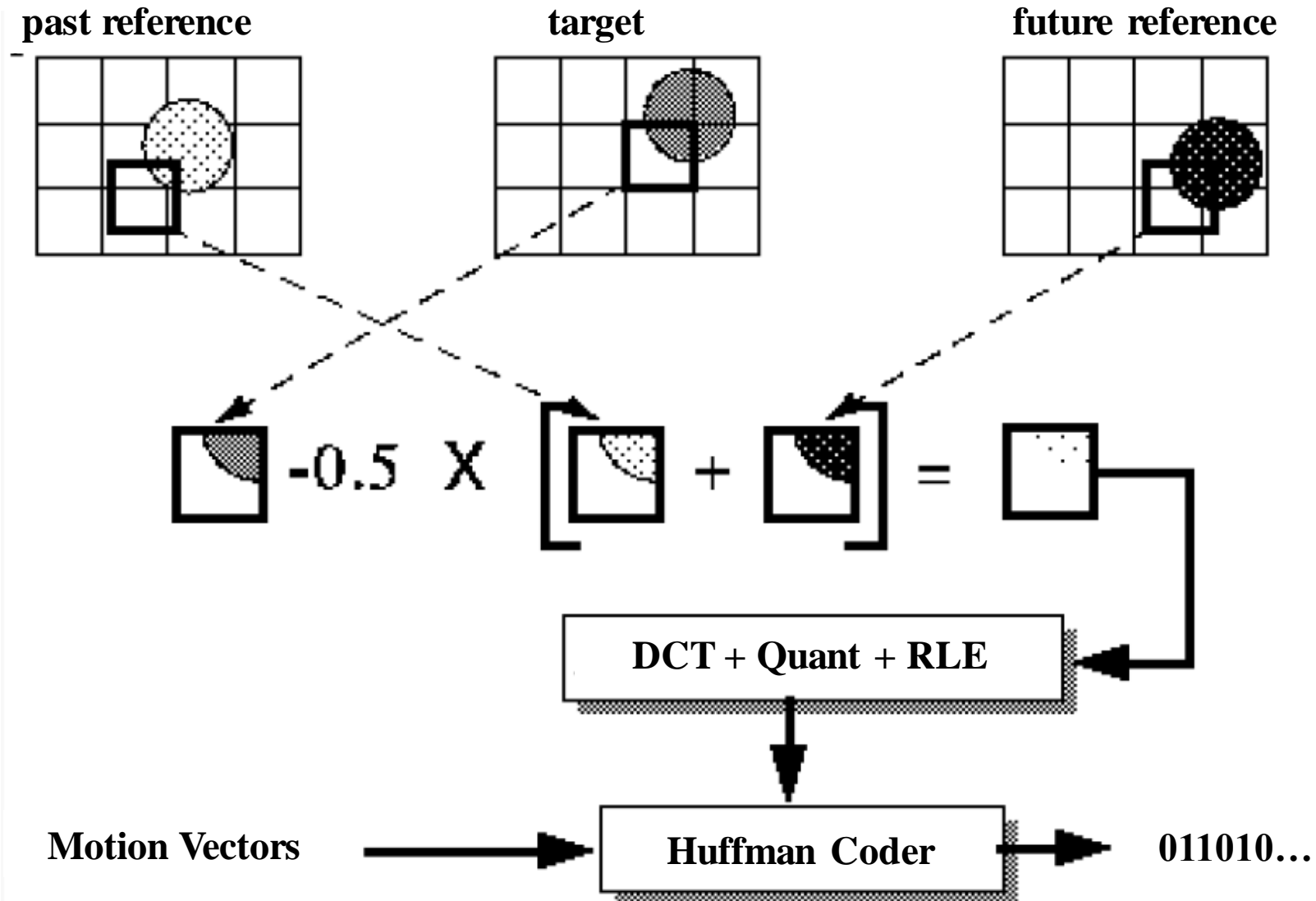
Kompresja ramek B

- Dwie ramki referencyjne typu I lub P
- Rodzaje kompensacji ruchu bloków
 - ważona względem obydwu ramek referencyjnych (ang. *bidirectional prediction* lub *motion-compensated interpolation*)
 - wykorzystuje **dwa**, oddzielne wektory ruchu dla makrobloków
 - DCT stosuje się do różnicy między blokiem bieżącym a średnią bloków referencyjnych
 - w przypadku dłuższej sekwencji ramek typu B, wagi nie muszą być równe
 - bez kompensacji ruchu
 - w przypadku, gdy estymacja ruchu daje zbyt duży błąd
 - DCT stosuje się do oryginalnych wartości pikseli w bloku bieżącym
 - względem wcześniejszej ramki referencyjnej
 - względem późniejszej ramki referencyjnej

Predykcja ramek B

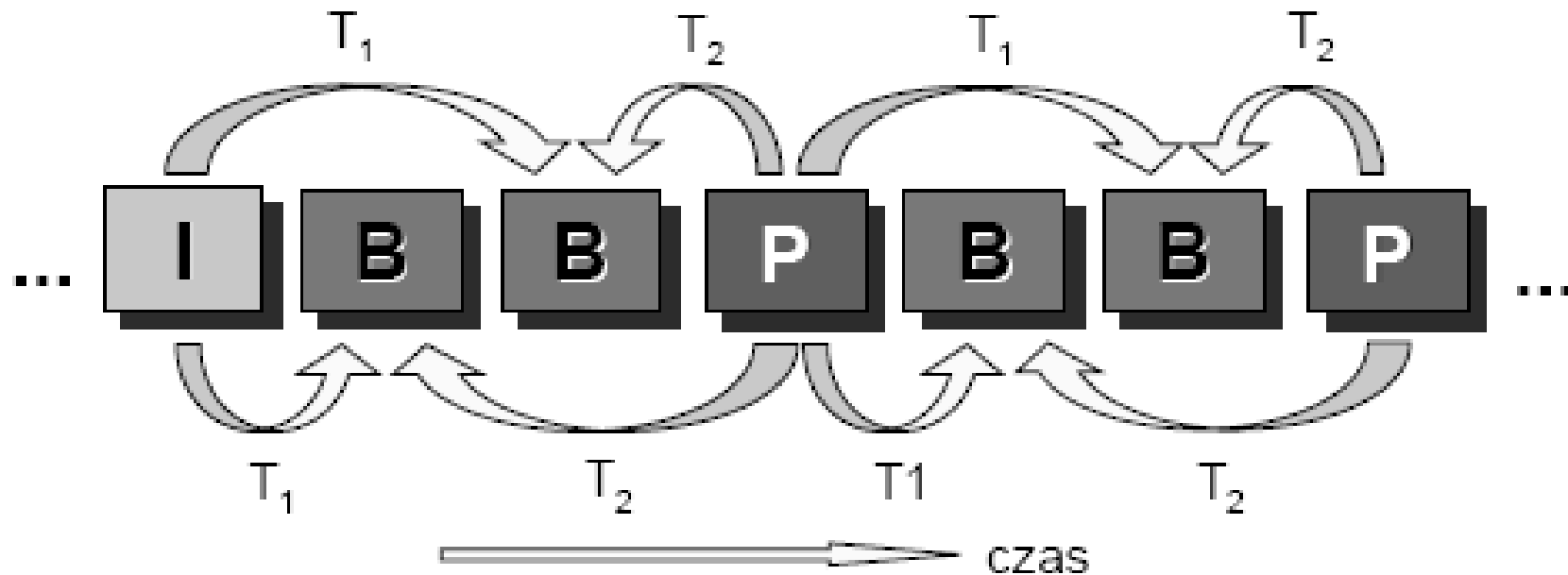
Typ makrobloku	Predyktor (ang. <i>Predictor</i>)	Błąd predykcji
Intra	$\hat{I}(\vec{p}) = 128$	$\hat{I}(\vec{p}) - I(\vec{p})$
predykcja w przód (ang. <i>Forward P.</i>)	$\hat{I}(\vec{p}) = I_{o1}(\vec{p} + \vec{mv}_{o1})$	
predykcja w tył (ang. <i>Backward P.</i>)	$\hat{I}(\vec{p}) = I_{o2}(\vec{p} + \vec{mv}_{o2})$	
Średnia	$\hat{I}(\vec{p}) = \frac{1}{2} \cdot [I_{o1}(\vec{p} + \vec{mv}_{o1}) + I_{o2}(\vec{p} + \vec{mv}_{o2})]$	

Kompresja ramek B



Ważona predykcja ramek B

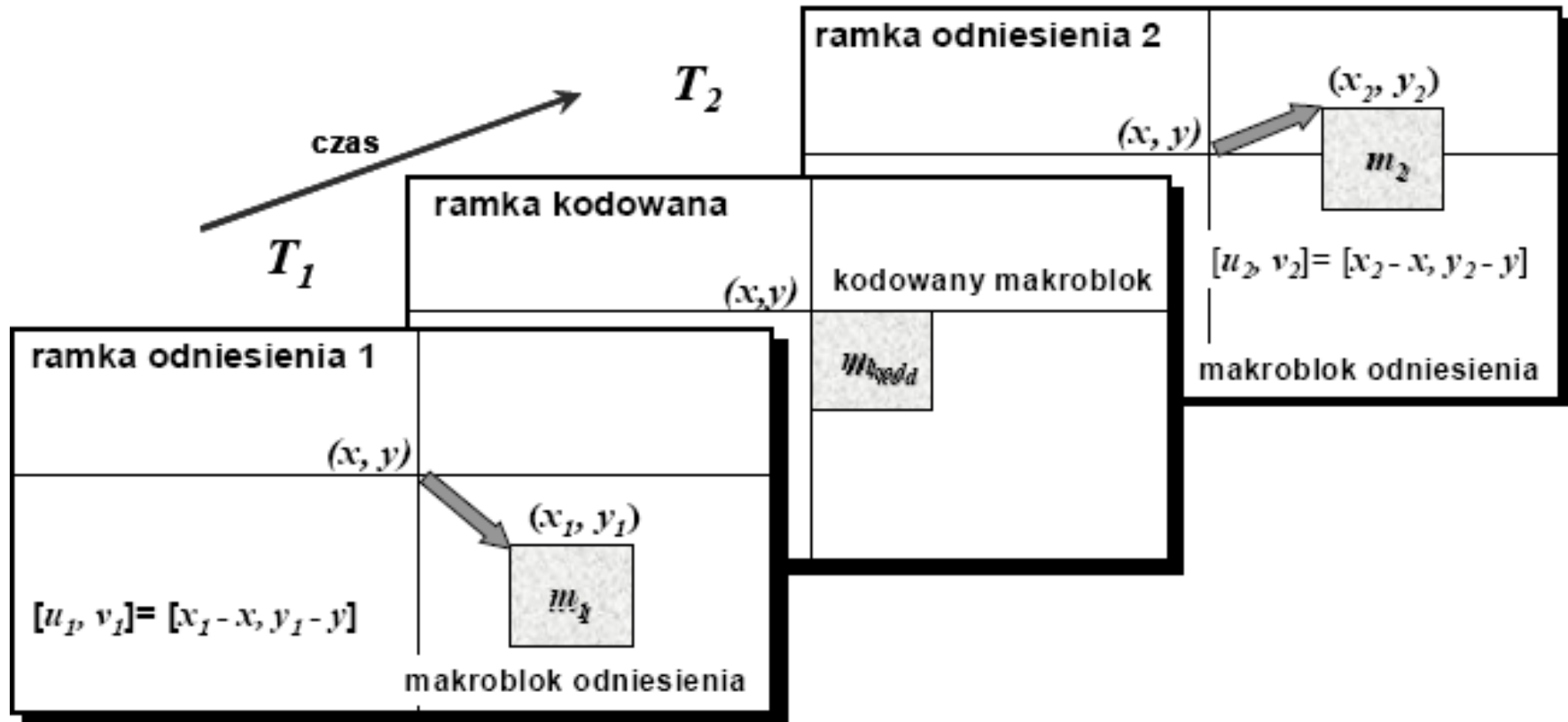
- Wyliczanie różnic względem bloków odniesienia uwzględnia odległość czasową obydwu ramek referencyjnych
- Dwa wektory ruchu MV_1 i MV_2



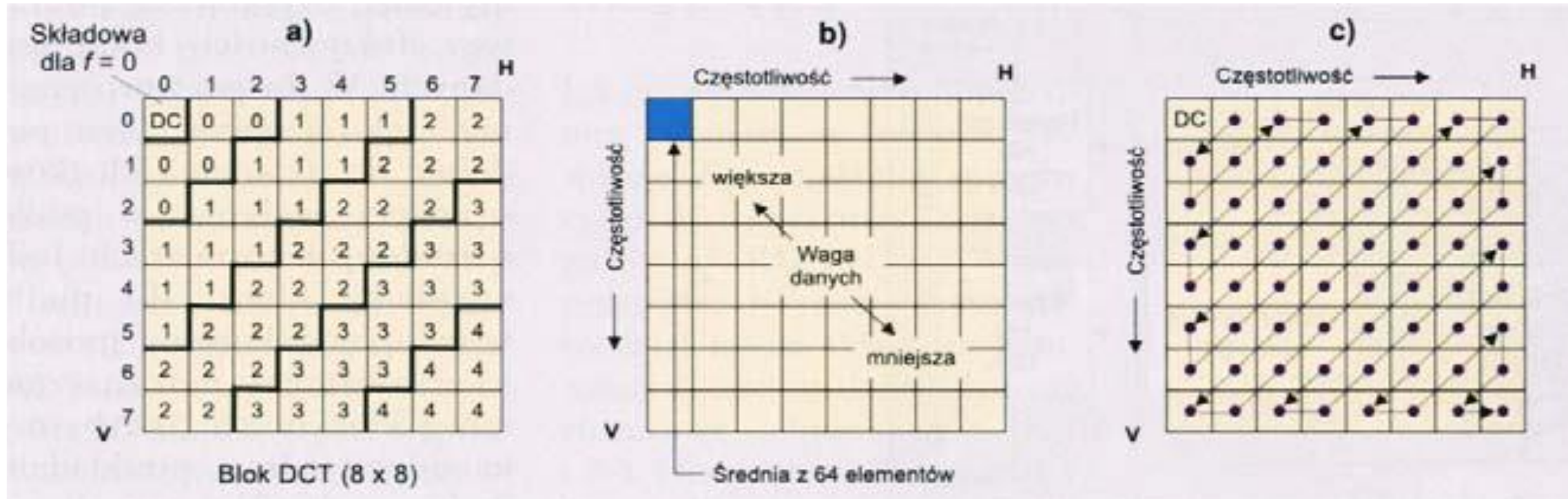
Ważona predykcja ramek B

- Ważone sumowanie bloków odniesienia

$$m_o(i, j) = \frac{T_2}{T_1 + T_2} m_{o1}(i + u_1, j + v_1) + \frac{T_1}{T_1 + T_2} m_{o2}(i + u_2, j + v_2)$$



DCT i kodowanie



Kwantyzacja współczynników DCT

- Dyskretnej transformacji cosinusowej podlegają oryginalne bloki lub ich błędy predykcji
- oddzielne progi kwantyzacji dla każdego kanału
- kanał DC w ramach I ma stały próg 8
- możliwość skalowania tablicy kwantyzacji Q dla każdego makrobloku
 - współczynnik skalowania (ang. *QuantizerScale*) w zakresie od 1 do 31
- zaokrąglanie wyniku dzielenia przez próg kwantyzacji
 - ramki typu I - do najbliższej liczby całkowitej

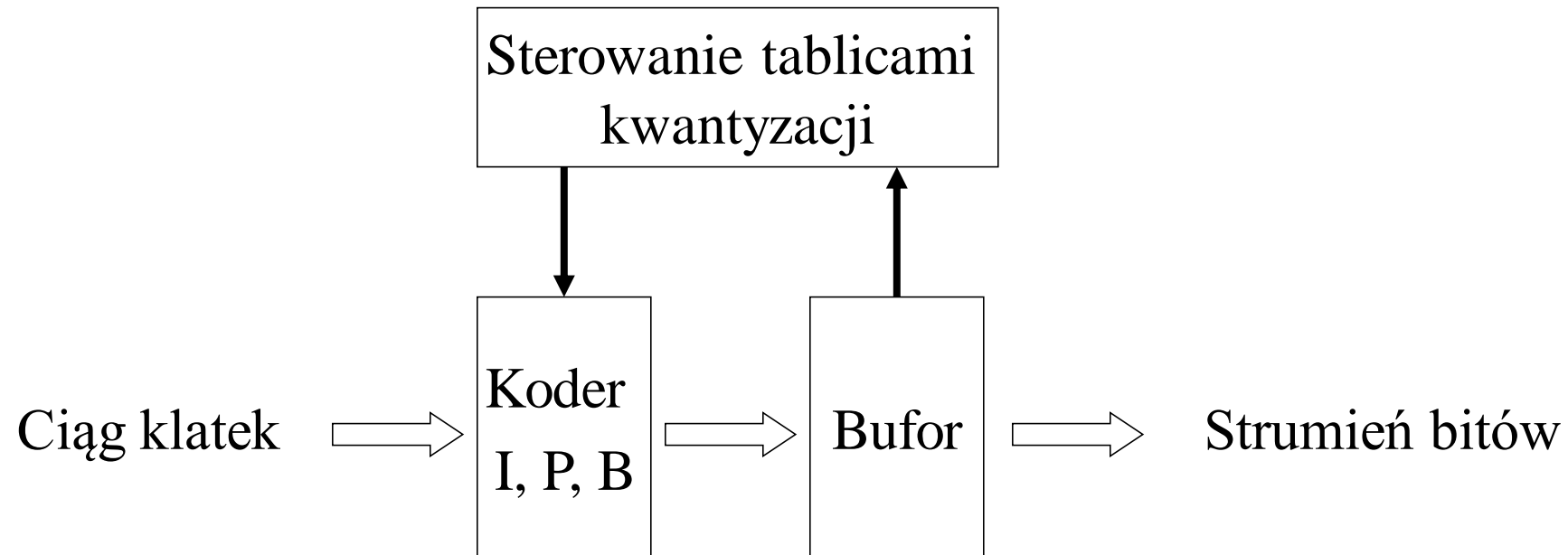
$$QDCT = \frac{(16 \cdot DCT) + (Sign(DCT) \cdot QuantizerScale \cdot Q)}{2 \cdot QuantizerScale \cdot Q}$$

- ramki typu P i B - w dół do liczby całkowitej mniejszej (co do modułu)

$$QDCT = \frac{16 \cdot DCT}{2 \cdot QuantizerScale \cdot Q}$$

Kontrola przepływności bitowej

- Dynamiczna zmiana tablic kwantyzacji w celu kontroli przepływności



Przykład kodowania

Ramka referencyjna z zaznaczonymi wektorami ruchu

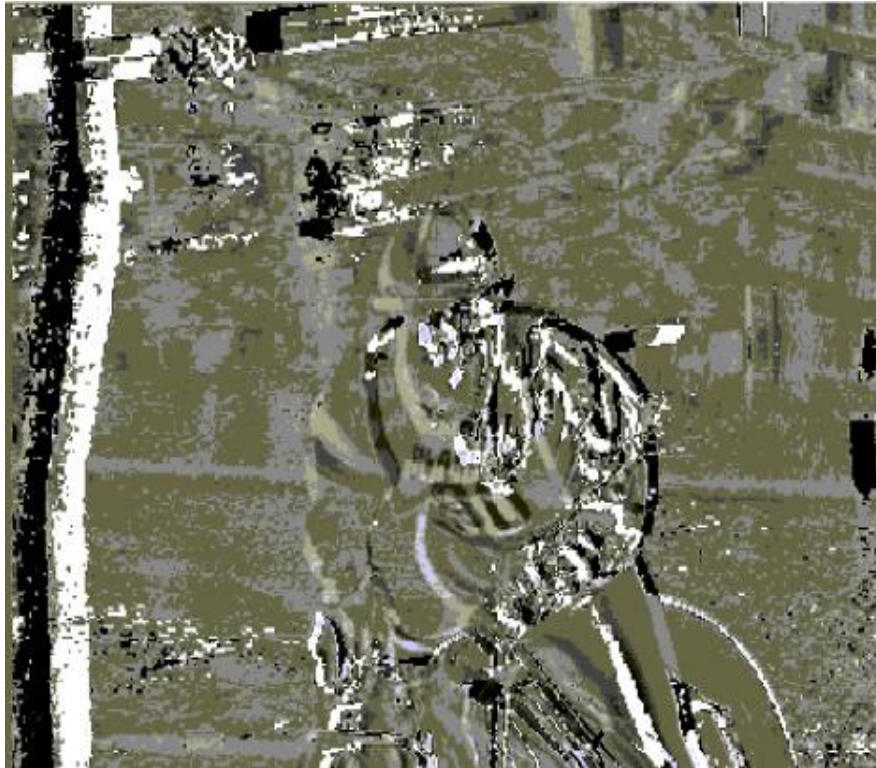


Ramka bieżąca (kodowana)



Błąd predykcji

bez kompensacji ruchu (bezpośrednie
porównanie ramek) ($MV=[0,0]$)



z kompensacją ruchu

