

```
/* *****  
* "THE BEER-WARE LICENSE":  
* Autors (alphabetical): Perwel Krzysztof<perkristo@gmail.com>,  
* Reut Michał<michalreut@gmail.com>, Rogowski Piotr<piotrekrogowski@gmail.com>  
* As long as you retain this notice you can do whatever you want  
* with this stuff. If we meet some day, and you think  
* this stuff is worth it, you can buy us a beer.  
* *****
```

MTI Egzamin:

Opracowanie wszystkiego!!! (Prawie ;P)

Wykład 1 - Podstawowe problemy i zagadnienia Standardy ITU H.32X; Projekt OpenH323

Podstawowe problemy związane z przesyłaniem danych audio/video w sieci:

Bandwidth - przepustowość

- Konferencja ISDN potrzebuje 128-384 Kbps
- W standardzie H.323 bazująca na standardzie IP około 384-768 Kbps
- Pakietyzacja powoduje nadłożenia informacji (20% więcej przez nagłówki)
- Wyższe jakości H.323 nawet 1.5 - 4 Mbps
- 6-20Mbps NTSC/PAL (MPEG1/2/4)
- 20-50Mbps dla HDTV
- Szerokość pasma powinna być symetryczna przy odbiorze i wysłaniu,
- MCU/bridge odbiera wszystkie strumienie jednocześnie (bardzo obciążone)
- Najmniej wymagająca jakość dla 8 osób wymaga $8 * 384\text{Kbps} = 3\text{Mbps}$ (na te czasy to drobne, ale znajdźcie mi takie łącze symetryczne (ściąganie i wysyłanie z taką samą prędkością))

Packet loss - utrata pakietów

- Mała szerokość pasma, przez co routery odrzucają niektóre pakiety - oglądasz film online, a tu brakuje części informacji, zazwyczaj włącza się buforowanie.
- Zdarzają się na łączach satelitarnych lub bezprzewodowych (wiadomo, burza jest i są zakłócenia)

- W konferencjach nie ma buforowania, więc ucina fragmenty filmów, albo są zkałócenia dźwięku.

Latency - opóźnienia

- Czas od zajścia zdarzenia (ktoś mrugnął) do momentu aż urządzenie końcowe się o tym dowie (Twój komputer zacznie przetwarzać to, że ktoś mrugnął)
- Opóźnienia wynikają z:
 - kodowanie/dekodowanie (można zainwestować w lepszy sprzęt)
 - przesył pakietów
- Czasami problemem jest brak synchronizacji dźwięku i obrazu

Jitter - wariacja opóźnień

- pakiety przychodzą w niewłaściwej kolejności
- powoduje zakłócenia dźwięku ("metaliczny" dźwięk)
- można zminimalizować zwiększając bufor, ale powoduje to wzrost opóźnień

Policies - zabezpieczenia

- Firewall/NAT mogą blokować porty, co uniemożliwi ustanowienie połączenia (zewnętrzne IP i te sprawy)

Do lania wody:

- Elementy:
 - Bandwith - potrzebna duża prędkość neta,
 - Packet loss- pakiety lubią się gubić, więc tracimy informacje
 - Jitter - opóźnienia się zmieniają mamy mieszaną kolejność pakietów,
 - Policies - Firewallle i inne zabezpieczenia blokują porty
- Potrzebujemy symetrycznego łącza, które ma przepustowość minimum liczba_osób*384kbs. MCU bardzo obciążone, ciężko o łącze z dużą prędkością udostępniania.
- Znikają nam pakiety, przez wszelkie problemy w przesyłaniu, mamy buforowanie które to rozwiązuje, ale nie w konferencjach,
- Opóźnienia spowodowane są kodowania/dekodowania i samego przesyłu.

ITU H.323

Komponenty:

Terminale

Komputery PC - klienci

Gatekeepers

- gatekeepery + terminale + gateways + MCU = strefa lokalna
- terminale rejestrują się gatekeeperze
- mapuje aliasy na adresy,
- akceptuje, albo odrzuca połączenia,
- może odrzucić połączenie, gdy jest już zbyt dużo połączeń
- kontroluje jakie urządzenia mają dostęp do local zone
- przetwarzanie sygnałów kontrolnych (?)
- logowanie połączeń
- przekierowanie połączeń

Gateways

- Może być mostem między H.323 i innymi technologiami,
- tłumaczy usługi, oraz transkoduje audio i wideo

MCU (multipoint control units)

- “Virtual meeting room” - łączy wiele rozmów w jedną konferencję,
- Dwa komponenty wg H.323 - scentralizowany i niescentralizowany
- MC (multipoint controller):
 - określa wspólne możliwości uczestników,
- MP (multipoint processor)
 - miksuje audio i przesyła spowrotem

Protokoły używane w H.323

RTP/RTCP (Real-Time Transport Protocol / Real-Time Transport Control Protocol)

- Opisuje transport danych w czasie rzeczywistym

H.225 (RAS - Registration, Admission and Status)

- wykorzystywany do połączenia między urządzeniem końcowym, a gatekeeperem
- najważniejsze komunikaty:
 - żąda nawiązania połączenia
 - odpowiedź odmowna
 - odpowiedź zatwierdzająca
- korzysta z UDP

H.255.0

- protokół używany m.in. do opisanie sygnalizacji, mediów, pakietyzacji, synchronizacji.

H.245

- komunikaty do otwierania i zamykania kanałów audio video

Q.931

- Sygnalizacja połączenia

RSVP (*Resource ReSerVation Protocol*)

- rezerwacja zasobów sieciowych

T.120

- przesyłanie danych i kontrola konferencji podczas interaktywnej komunikacji

KODEKI:

G.711 i inne – audio

H.261, H.263 – video

Dołania wody:

- Elementy:
 - Terminale to zwykłe komputery,
 - Gatekeepers - (jak bramkarze na dyskotece), mogą odrzucać połączenia, gdy jest ich już za dużo, i kontrolują kto ma dostęp,
 - Gateways - pozwalają się łączyć między standardami
 - MCU - Sprawdza kto ma jakiego kompa, odbiera wszystkie strumienie, miksuje je i odsyła.

Wykład 2 - SIP, kompresja mowy

SIP (Session Initiation Protocol)

Opis:

- Zaprojektowany do nawiązywania, utrzymywania i rozłączania sesji,
- Zapewnia podstawowe funkcje (wybieranie numeru, sygnał zjawności itp.)

Adres:

- Adres SIP - SIP URL - podobne do adresu e-mail 3847362378@89.76.81.12
- pierwsza część numer, lub nazwa użytkownika,
- druga część - domena, lub adres sieciowy

Komponenty

- Agenci Użytkownika
 - User-Agent Client - inicjuje zapytania i odpowiada od strony użytkownika
 - User-Agent Server - odbiera zapytania i zwraca odpowiedzi od strony użytkownika
- Serwery sieciowe
 - Proxy server - pracuje w imieniu innych klientów. Interpretuje, oraz przypisuje nagłówki zapytań zanim prześle je do innych serwerów, zapewnia otrzymanie odpowiedzi tą samą trasą,
 - Redirect server - pomaga znaleźć użytkownika w sieci. Akceptuje zapytania SIP i odsyła adres serwera obsługującego poszukiwanego użytkownika

Komunikaty

- Żądania
 - INVITE - zaproszenie do sesji, sprawdzanie możliwości sprzętowej
 - ACK - potwierdzenie połączenia
 - OPTIONS - zbiera informacje o agentach i możliwościach serwerów
 - BYE - rozłącza połączenia
 - CANCEL - odwołuje zapytanie
 - REGISTER - żąda rejestracji w serwerze SIP
- Odpowiedzi:
 - SUCCESS
 - CLIENT-ERROR
 - SERVER-ERROR

Kompresja dźwięku

Opis:

- Człowiek słyszy 20Hz do 20kHz
- Dźwięk jest ciągły w czasie
- Kompresja to, to kompromis między jakością, a przepływnością strumienia danych,
- Koder koduje, dekodek rekonstruuje (dekoduje)
- PCM (Puls Code Modulation) - format przekazywany do kodera
- Schemat: Przetwornik - PCM -> KODER - SIEĆ -> DEKODER - PCM -> Przetwornik
- Dzielimy metody na kompresję mowy i szerokopasmowego sygnału.
- VO(ices)CODERS - modelują źródło dźwięku,
- Układy kompresji szerokopasmowego dźwięku - nie robią założeń odnośnie dźwięku.
- Często kompresuje się następujące po sobie ramki/segmenty/bloki (trwają 20ms)

Próbkowanie

Pomiar wartości amplitudy co stały odcinek czasu

- Im większa częstotliwość próbkowania tym dokładniejsze przybliżenie dźwięku,
- Teoria Nyquista - próbkowanie dźwięku musi być co najmniej dwa razy większa od

największej częstotliwości sygnału odwzorowywanego.

- Za duże częstotliwości mogą powodować zakłócenia o niższych częstotliwościach,

Niektóre częstotliwości próbkowania:

- 44.10 (Cd-Audio) - Pełen zakres pasma (muzyczka w końcu), uciną o tyle częstotliwość, żeby filtry poprawiające sobie radziły,
- 16kHz (komunikacja multimedialna) - Odwzorowuje szerokopasmowy sygnał mowy o zakresie pasma 50Hz - 7kHz.
- 8kHz (GSM) - pozwala odwzorować wąskopasmowy sygnał mowy o zakresie częstotliwości 200Hz - 3.2 kHz

Kwantyzacja

- Konwersja ciągłego zakresu amplitudy na skończony zbiór wartości (dyskretyzacja)
- **Szum kwantyzacji** - różnica pomiędzy dźwiękiem analogowym, a cyfrowym
- Czym więcej bitów na których reprezentujemy amplitudę, tym lepsza jakość
- 16 bitów wystarczy do dźwięku wysokiej jakości, 12-14 dla mowy w telekomunikacji.

Miary jakości

- SNR (Signal to Noise Ratio) - wyrażany w dB. Stosunek sygnałów wejściowych, do różnicy sygnałów wejściowych i wyjściowych.
- Subiektywna percepcja - 40 słuchaczy wystawia oceny.

Wymagana przepustowość

- próbki an sekundę * bitów na próbkę * kanały
- Sterefonia $44100 * 16 * 2 = 1.41\text{Mbps}$
- Wąskopasmowa $8000 * 12 * 1 = 96\text{kbps}$

Wykład 3 - Kodowanie mowy

Metody kompresji mowy

VOicesCODERS

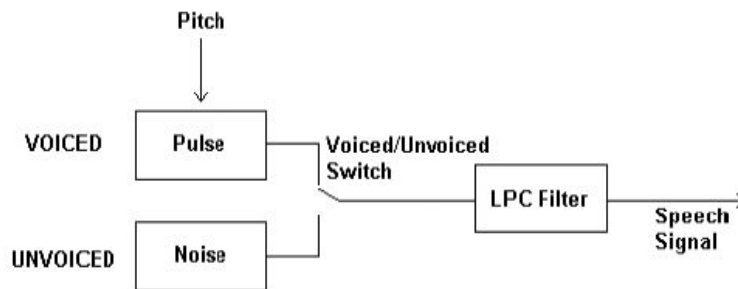
Metoda kompresji dynamiki amplitudy,

- Przy kwantyzacji równomiernej potrzebujemy 12 bitów
- Używając logarytmicznej możemy ograniczyć rozdzielczość do 8 bitów,
- Wady:
 - Mały stopień kompresji,
- Zalety
 - Prawie nie do usłyszenia,

- Prosta obliczeniowo,
- Małe opóźnienia.

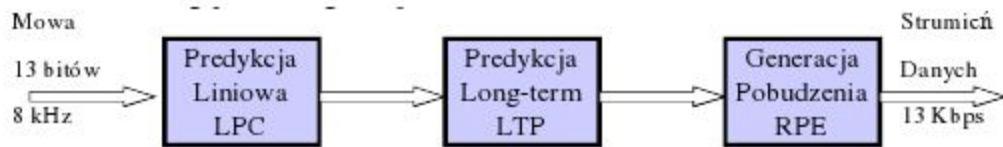
Metoda predykcji liniowej LPC,

- Dzielenie ramek na dźwięczne (ton krtaniowy, struny) i bezdźwięczne (szum),
- Zamodelowanie kanału głosowego człowieka,
- Klasyfikacja ramek:
 - Etap 1: Badanie energii sygnału (duży -> dźwięczna)
 - Etap 2: Rozstrzyganie niepewnych (dźwięczne częściej przecinają oś X)
 - Etap 3: Analiza otoczenia: Głoski bezdźwięcznej nie otaczają dźwięczne
- Określenie okresu drgań struny głosowej podczas mówienia, do dekodowania dźwięcznych ramek,
- Skorzystanie z AMDF (jakiejś dziwnej funkcji ;)) która ma minimum, gdy argument jest równy okresowi drgań (częstotliwości), oraz pozwala klasyfikować na dźwięczne / bezdźwięczne
- Filtr dekodujący bazuje na współczynnikach wyznaczanych podczas kodowania (trzeba przesłać razem z dźwiękiem)
 - 10 parametrów dla dźwięcznych, 4 dla bezdźwięcznych
- Sygnał wejściowy nazywany jest sygnałem pobudzenia.
- Współczynniki wyznaczamy minimalizując średni błąd kwantowy,
- Każdą próbkę można przedstawić jako kombinację M poprzednich próbek.
- Model LPC jest uproszczony, stąd strata jakości.



Metoda pobudzenia równomiernego RPE,

- Z grupy analiza - synteza (AbS)
- Syntezuje się taki sygnał pobudzenia, który po przejściu przez LPC będzie najbliższy oryginałowi,
- Zamiast podziału na dźwięczne/bezdźwięczne sygnał pobudzenia jest zbiorem stałej liczby impulsów pobudzających, określanych przez koder.
- Koder określa położenie pierwszego impulsu i amplitudy kolejnych.
- Czym więcej impulsów na ramkę, tym lepsza jakość,
- Używane w GSM



Metoda pobudzenia kodowego CELP.

- Rozwinięcie CELP
- Użycie książek kodowych, zawierający “sygnały resztkowe”.
- Z książki są dobierane najlepsze sygnały pobudzenia (minimalizacja błędów)
- Wady:
 - Długa książka kodowa -> Długi czas przeszukiwania, Długie kody

waveform codecs

- Nie wykorzystują informacji o sposobie powstania dźwięku,
- Produkują sygnał, łatwy do zrekonstruowania.

Metoda predykcji adaptacyjnej ADPCM

- szacowanie amplitudy próbki na podstawie amplitudy próbki poprzedzającej,
- Zastępuje wartość amplitudy, różnicą amplitud sąsiednich,
- Korzystając z kodowania entropijnego (Hufmana) można uzyskać dużą wierność reprezentacji,
- Adaptacja polega na dostosowywaniu się do lokalnych właściwości sygnału,
- Feedforward - parametry adaptacji - przesyłane w strumieniu danych,
- Feedbackward - adaptacja na zakodowanej wersji sygnału (więcej błędów, ale większa kompresja)

SB-ADPCM

- Dzielenie sygnału na podpasma częstotliwościowe,
- Bardziej znaczące częstotliwości, kodowane dokładniej - z większą ilością bitów.

Kodery mowy – podsumowanie

| Protocol | Bit Rate / Kbps | End to End Delay in Ms | Coding |
|----------|-----------------|------------------------|---------------|
| G.711 | 48, 56, 64 | 0.125 | PCM / A and m |
| G.726 | 16/24/32/40 | 0.250 0.375 | ADPCM |
| G.721 | 32 | - | ADPCM |
| GSM | 13 | 20 | RPE LTP |
| G.722 | 48 | - | ADPCM |
| G.727 | 16, 24, 32, 40 | 60 | ADPCM |
| G.728 | 16 | 2.5 | LD Celp |
| G.729 | 8 | 15 | CS ACELP |
| G.723.1 | 5.3 and 6.3 | - | MPMLQ |

Do łania wody:

- Vocodery próbują odzworować mowę człowieka,
- LPC:
 - Dzieli na dźwięczne/bezdźwięczne, przez badanie energii,
 - Wyznaczanie okresu maksymalnej wysokości dźwięku
 - Budujemy filtr budujący dźwięk na podstawie zbioru współczynników
 - Łączymy parametry

Wykład 4 - Kompresja dźwięku szerokopasmowego

Standard kompresji MPEG-1

- MPEG - Moving Pictures Experts Group - Ruszające się obrazki grupa ekspertów kurwa!
- Standard kompresji zsynchronizowanych sekwencji obrazów i dźwięku
- MPEG-1 składa się z 5-ciu części. Trzecia poświęcona jest kompresji dźwięku.
- Opisuje: Syntakę strumienia danych, dekompresję i testy zgodności
- 3 warstwy kompresji
- Dekoder warstwy wyższej dekoduje strumień danych warstwy niższej.

Model psychoakustyczny

- Założenie kompresji dźwięku szerokopasmowego: Usuwanie tego czego nie słyszy człowiek (informacje percepcyjnie nieistotne)
- Człowiek słyszy od 20Hz do 20kHz (najwięcej 2 - 4kHz) - dlatego kodujemy z różną dokładnością,
- Wyznacza funkcję częstotliwości i czasu poniżej których dźwięk nie jest słyszalny,
- Nieliniowe skalowanie dziedzinie częstotliwości,
- Kwantujemy sygnał w każdym podpaśmie tak, aby nie tworzyć szumów,
- Trzeba uwzględnić maskowanie czasowe.

Efekt maskowania

- próg słyszenia (najniższy słyszalny dźwięk) jest zależny od odbieranych dźwięków
- Efekt maskowania następuje w dziedzinie częstotliwości,
- Zakres zależy od amplitudy (skoku) dźwięku,

Maskowanie wyprzedzające (forward masking)

Dźwięk maskuje dźwięki następujące po nim, sięga do 200ms,

Maskowanie wsteczne (backward masking)

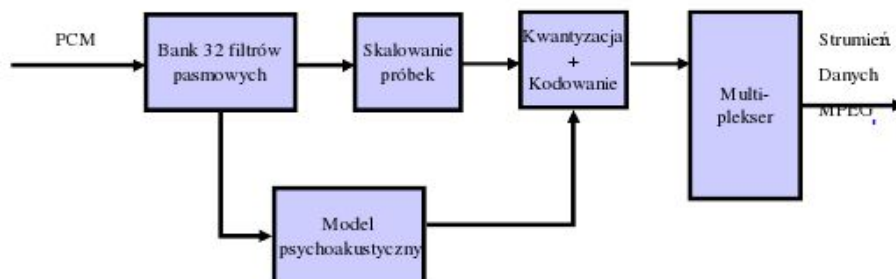
Dźwięk maskuje inny, występujący wcześniej, 2 do 10 ms

Maskowanie tymczasowe (temporal masking)

Maskowanie występujące zaraz po głośnym dźwięku, 5ms

Koder MPEG-1

- Wykorzystuje model wierności percepcyjnej dźwięku,



Schemat blokowy układu kodera MPEG-1

Algorytm kompresji MPEG-1

- Jest używany niezależnie dla każdego kanału
- Aby zmniejszyć przepływność strumienia korzysta się z trybów
 - Intensity stereo - wysyła się zsumowany sygnał i czynniki skalujące pozwalające je rozdzielić,
 - M/S stereo (Middle/Side) - lewy kanał transmituje sumę, prawy różnicę. Może być stosowany w Layer 3
- Składowe ramki: bity synchronizacji (12) + bity informacji systemowej (20), bity CRC (16) do detekcji błędów, Opis alokacji bitów (4), Czynniki skalujące (6), Zakodowane próbki każdej częstotliwości
- W dekodерze nie trzeba robić analizy psychoakustycznej, dlatego jest mniejsza złożoność.

Layer1

1. Filtracja sygnału wejściowego przez 32 filtry na 32 podpasma po 12 próbek
2. Analizujemy wyznaczone podpasma (wyznaczamy alokację bitów i progi maskowania dla każdego podpasma)
3. Skalowanie sygnału dla 12 próbek (współczynnik skalowania równy maksymalnej z próbek)
4. Kodowanie próbek i tworzenie wyjściowego strumienia łącznie z informacją organizacyjną i synchronizacyjną,
5. Obliczanie transformacji FFT dla 512 próbek sygnału oraz składowych tonalnych i nietonalnych,
6. W każdym podpaśmie z 32 wyznaczonych, progi porównywane są z maksymalną wartością w tym podpaśmie (ten stosunek to inaczej SMR - Signal to Mask ratio), jest to wejście układu kwantyzacji.
7. Iteracyjna minimalizacja MNR (Mask to Noise Ratio), równa SNR-SMR (różnica sygnał mask i signal noise ratio).
8. Podobno zapewnia minimalną słyszalność szumu, dzięki maskowaniu szumów z sąsiadujących podpasem.

Layer2

- Wykorzystuje się podobieństwa między współczynnikami skalowania, i dla pasm o wyższej częstotliwości,
- lepsza kompresja, większa złożoność

Layer3 - MP3

- Bloki długie (12 próbek) dla gwałtownych zmian sygnału
- Bloki krótkie (36 próbek) do reprezentacji stacjonarnych fragmentów
- Zbieżność podpasem częstotliwościowych i krytycznych,

- Stosowane kwantyzacja nierównomierna (np logarytmiczna) i kodowanie entropijne (np. Huffman)
- Algorytm kompresji iteracyjnie dokonuje alokacji tak, aby uzyskać największą korzyść w percepcji dźwięku.

MPEG-1, warstwy

Layer1

- Filtry DCT tylko z jedną ramką,
- takie samo rozłożenie częstotliwości dla każdego pasma,
- Model psychoakustyczny wykorzystuje tylko maskowanie częstotliwości

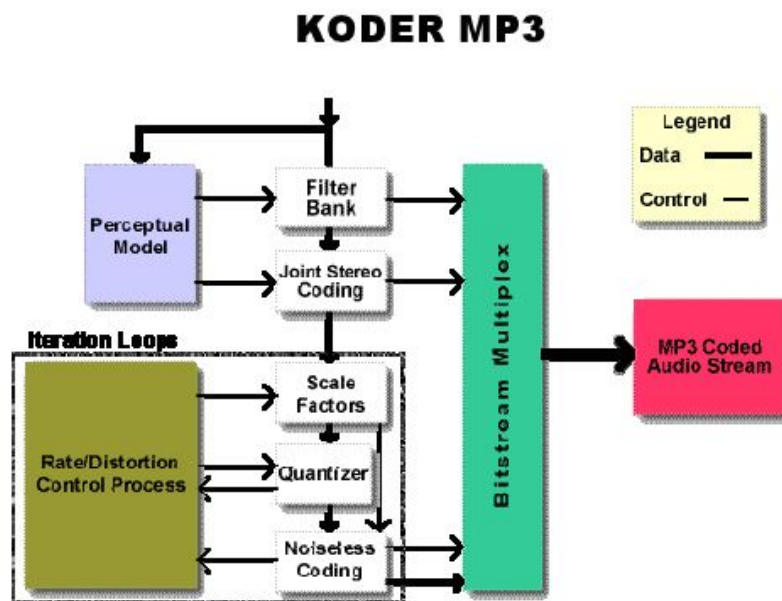
Layer2

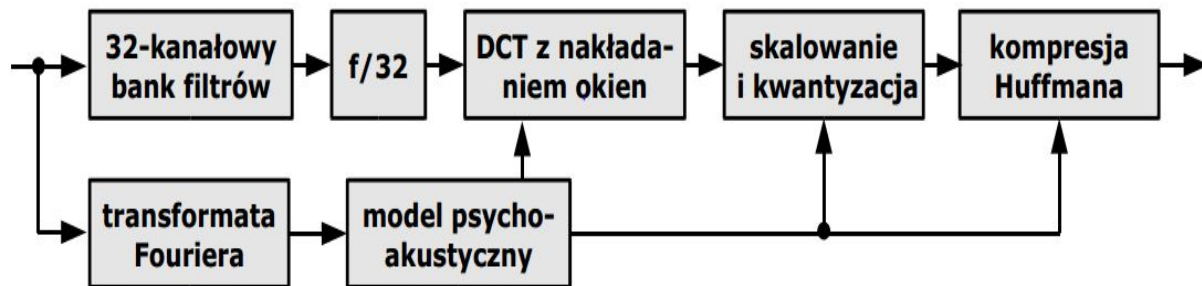
- Używa trzech ramek na filtr,
- korzysta częściowo z maskowania tymczasowego,

Layer3

- Lepszy filtr krytycznych pasm (nierówne częstotliwości)
- model psychoakustyczny używa maskowania tymczasowego,
- używa kodowania huffmana,

KODER MP3





Wykład 5 - formaty audio

Różnice części 3-ciej MPEG-2 w stosunku do MPEG-1

- Zwiększenie liczby kanałów do 5.1,
- Dodatkowe częstotliwości próbkowania (16, 22.05, 24kHz),
- Rozszerzenie rozdzielczości próbki do 16-24 bitów
- Obniżenie przepływności strumienia danych do 8 Kbps
- Dodatkowe kanały komentatorskie (do 8 kanałów) tzn. lektor
- MPEG-2 mogą dekodować MPEG-1
- MPEG-1 mogą dekodować 2 główne kanały MPEG-2

MPEG-2 AAC

- do 48 kanałów
- do 16 kanałów LFE (low frequency enhancement)
- 16 kanałów wielojęzycznych
- 16 strumieni danych
- Możliwe użycie 7.1
- Częstotliwość próbkowania do 96 kHz
- Lepsza jakość niż dla
 - MPEG1 warstwy 3 128Kbps
 - MPEG1 warstwy 2 192Kbps
- Ulepszenia w stosunku do MP3:
 - Lepsze filtry MDCT zamiast hybrydowych
 - Większe rozmiary bloków długich mniejsze krótkich
 - Lepsze tryby łączące kanały (M/S), może być kodowane w podpasmach, a nie jako główny kanał,
 - Wprowadzono predykcję LTP

Ogg Vorbis

- bezstratna kompresja konkurująca z mp3,
- Ogg jest ogólnie metodą przesyłu multimedialnych

- Vorbis jest formatem audio dla Ogg
- Tworzony przez Xiph.org Foundation
- Open source
- Przedział jakości od -1 do 10, domyślnie 3. Określana jako lepsza od 128 MP3 a zajmująca jedynie 10% jej rozmiaru,
-

WMA

- używany jest model psychoakustyczny, format stratny, drugi najpopularniejszy po mp3 format dźwięku
- Korzysta ze zmodyfikowanej dyskretnej transformacji cosinusowej, (MDCT)
- Każda ramka zawiera 128, 256, 512, 1024, 2048 próbek po przetransformowaniu przez MDCT (punkt wyżej)
- kodowanie Huffmanem
- Istnieje wersja bezstratna WMA Lossless
- Ścieżki stereo są łączone za pomocą M/S (midside) - jedna ścieżka zawiera sumę sygnałów audio, a druga różnicę - można kompresować je niezależnie (podobno lepiej). Później można wyłuskać z tego stereo

Mousepack

- M/S encoding
- Ulepszone kodowanie huffmana
- Usuwanie szumów jak w MPEG4 AAC V2
- w pełni modyfikowalny bitrate pomiędzy 0 a 1300 kbit/s
- Całkowicie bezstratne,

Wavpack

- opis jak mousepack

FLAC

- Darmowy bezstratny,
- zmniejszenie transferu bez zmniejszenia integralności,
- free/open source
- silny
- PCM, do 8 kanałów,
- CRC, run length coding,

Monkey's Audio

- .ape, .cue, .apl
- lepsza kompresja,
- Nie otwarty, ani nie darmowy,

AC-3 (Audio Code number 3) (kinowy)

- Inaczej Dolby Digital,
- wierność percepcyjna,
- standard dźwięku HDTV w USA,
- próbkowanie 32, 44.1, 48kHz
- Zakres przepływności od 32 do 640 Kbps,
- Sześć kanałów niezależnych,
- AC3
 - kompresja od 1:3 do 1:13
 - Korzysta z transformaty TDAC (Time DOrain Aliasing Cancelation)
 - zmienna rozdzielczość czasowo-częstotliwościowa,
 - Kontrola CRC
 - Dodatkowe dane określające typ materiału,

Dolby Digital - Surround EX(kinowy)

- Rozwinięcie DD
- dodatkowy kanał surround po bokach,
- polepszenie strefy najlepszego odsłuchu,
- rozwiązania problemów z lokalizacją dźwięków,

DTS - Digital Theater System(kinowy)

- system kinowy i domowy
- do ośmiu kanałów,
- przetwarzanie sygnału algorytmem Coherent Acoustic Coding,
- DTSExtendedSurround
 - Rozszerzony DTS + dodatkowe centralny tylny
 - DTS-ES Tylny niezależny

Dolby Digital Plus

- Możliwość miksowania dodatkowych ścieżek dźwiękowych,
- Wysoka przepustowość dla Blue-Ray
 - 640kbps dla core audio,
 - 1024 kbs dla extension packet

Dolby TrueHD

- Rozwinięcie bezstratnej kompresji dźwięku MLP Lossless na DVD-Audio
- techniki
 - wykorzystanie podobieństw w kanałach,
 - predykcja,
 - kodowanie Huffmana,

- o buforowanie danych
- Do 14 kanałów,
- Przepływność 18 Mbit/s
- Kompresja: 2:1 - 4:1
- Wykorzystanie metadanych, normalizacji dialogów, kompresji dynamiki,

DTS-HD Audio

- Podział na core i extension,
- dodatkowe kanały,
- Większa przepływność
- Dodatkowe ścieżki,

Przegląd formatów

Wykład 6 - Standard JPEG

Opis ogólny

- JPEG - Joint Photographics Expert Group - Fotograficzna grupa eksperckich dżojntów
- Definiuje koder, dekoder i sposoby przejścia,
- Definiuje stratne i bezstratne procesy kompresji,
- PNG i tak wymiata.

JPEG – kompresja obrazów kolorowych

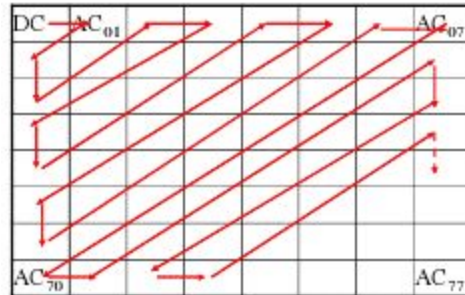
- Proces kompresji obrazu o wielu składowych koloru realizuje się niezależnie dla każdej ze składowych,
- Dlatego kolorowe obrazki trzeba rozdzielić na składowe YCbCr (Y - luminacji, CbCr - chrominacji)
- Zmniejsza to zależność składowych obrazu.
- Jako, że zmiana chrominacji jest mniej widoczna dla ludzi, stosuje się downsampling - tzn, skaczemy sobie co dwie składowe chrominacji w pionie i poziomie, bo skok co jeden byłby zbyt mały żeby go dostrzec... a po co, jak można oszczędzać.

Koder JPEG

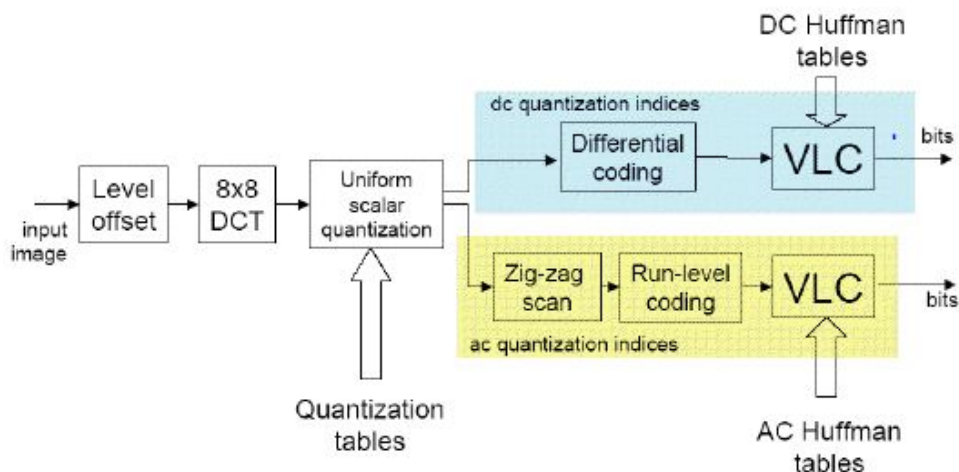
- Obraz dzielimy na bloki 8x8
 - o Przesuwamy ich zera do połowy zakresu, czyli jak mamy piksel 8-bitowy (256) to odejmujemy sobie 128. Czyli z zakresu od 0 do 256, dostajemy zakres od -128

do 128.

- Każdy blok przekształcamy do bloku $S = DsD^T$, gdzie D jest macierza współczynników DCT (rozmiar NxN, N=8).
 - Elementy S_{00} nazywamy współczynnikiem DC (direct current), pozostałe 63 elementy w S noszą nazwę współczynników AC (alternative current)
- AC są później układane w kolejności zig-zag



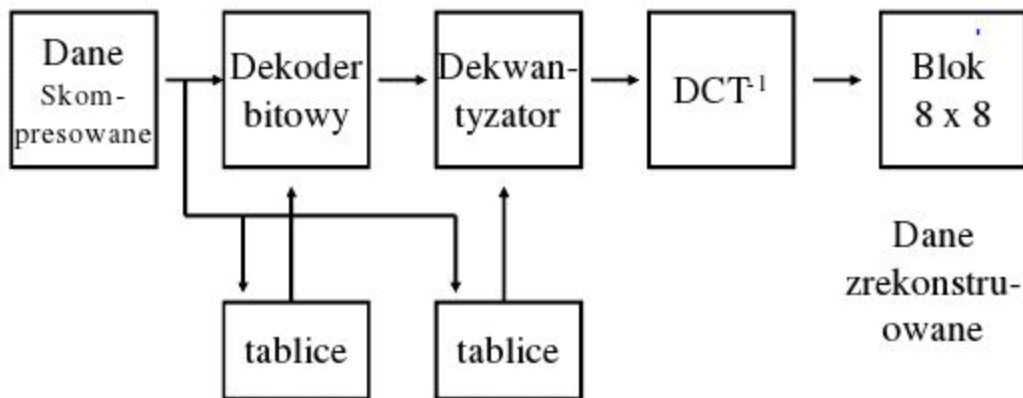
- AC są kodowane za pomocą RLE (Run Length Code) i Huffmana
- W fazie kwantyzacji, każdy współczynnik bloku S dzielimy przez odpowiadający mu element z tablicy kwantyzacji Q, zaokrąglamy i podstawiamy do bloku S_q
 - JPEG określa tablicę kwantyzacji Q, ale nie narzuca jej stosowania, można zrobić własną.
- W i-tym bloku wartość współczynnika DC^i zastępujemy różnicą $DC^i - DC^{i-1}$ i tą różnicę się koduje,
 - Współczynniki AC w każdym bloku ustawiamy wg porządku zyg-zag, a następnie kodujemy mieszanką technik RLE i Huffmana, lub arytmetycznym. Proces jest wspierany tablicami kodowymi.



Dekodowanie JPEG

1. Dostaje blok S_q
2. W fazie dekwantyzacji rekonstruuje blok R mnożąc elementy S_q przez odpowiednie elementy z Q
3. do R stosowana jest odwrotna transformata kosinusowa DCT tworząc blok zrekonstruowany $r = D^T R D$
4. Blok r różni się od s , ponieważ etap kwantyzacji wprowadza stratę informacji (zastępuje przedział liczb pojedynczym reprezentantem tzn. zmienia funkcję ciągłą na zbiór liczb (chyba))

Uproszczony schemat dekodera JPEG



Źródełko o DCT

- Właściwości
 - Utrzymanie poziomu błędów kwantyzacji,
 - Skupienie energii sygnału - wiele współczynników jest małych i łatwo je się kompresuje, ponieważ wartości są skupione,
 - Dekorelacja współczynników traktowanych jako zmienne losowe
 - Istnieją szybkie algorytmy obliczania
- Normalne mnożenie macierzy w DCT wymaga od nas $2N^3$ natomiast dodawań $2N^2(N-1)$
 - Dla $N = 8$ odpowiednio to 1024 mnożenia i 896 dodawań (dużo)
- Dzięki szybkim algorytmom możemy zmniejszyć tę ilość
 - grupowanie działań trygonometrycznych (353* i 448+)
 - Arai (206* i 464+)
 - Dwuwymiarowy algorytm Feiga (54* i 464+ i 6 arytmetycznych przesunięć)

Tryby pracy stratnego kodera JPEG

- SEKWENCYJNY - jeden po drugim, wierszami, po zakodowaniu idą na wyjście, a na wejście wchodzi kolejne,
- PROGRESYWNY - Blok Z uzyskany po kwantyzacji jest kodowany w kilku przeglądkach obrazu.
 - Podział na zakresy spektralne - Z każdym obiegiem dorzucane są informacje o szczegółach bloku, biorąc dodatkowe współczynniki z kolejnych zakresów spektralnych,
 - Sukcesywna aproksymacja wartości współczynników - kolejny przebieg orzuca mniejsze naczające bity współczynników w bloku
- Hierarchiczny - obraz jest reprezentowany przez sekwencję ramek
 - Ramkę wyżej uzyskujemy przez downsampling (jak wcześniej w dźwięku, przewidywanie na podstawie poprzedniej)
 - Proces odwrotny (upsampling) daje rekonstrukcję tej ramki w psotaci obrazu referencyjnego,
 - Za wyjątkiem ramki najwyższej w hierarchii, różnice są kodowane DCT

Segmenty danych

- Dzielimy na:
 - Segmenty parametrów,
 - Segmenty danych kodowych,
- Segment parametrów rozpoczyna się od znacznika (markera), który mówi nam co po nim następuje, Np. rozpoczęcie wartości elementów tablicy kwantyzacyjnej,
- Dane mogą być dzielone markerem restartu. Pozwala to dekodować dalej, gdy wystąpił błąd.
- Poza danymi pikselowymi mamy jeszcze markery dla wszystkich tablic kodowych i kwantyzacji
- Skrócony format pomija niektóre z markerów,

Wykład 7 - Kodowanie obrazów cyfrowych (formaty)

Grafika rastrowa vs. wektorowa

Grafika rastrowa (bitmapowa)-obraz składa się z siatki punktów o określonym kolorze

Grafika wektorowa-obraz składa się z obiektów opisanych równaniami matematycznymi

Typy obrazów

BMP(BitMap Picture)

- Bitmapa-obraz w grafice rastrowej, w którym zakodowana jest wartość(kolor) każdego piksela
- format opracowany pierwotnie dla systemu OS/2 (IBM), obecnie podstawowy format systemów Windows
- rozszerzenie *.bmp
- kompresja
 - z reguły nie stosuje się
 - Obrazy 4- i 8-bitowe można kompresować metodą RLE (Run Length Encoding)
- nagłówek(54B), opcjonalna paleta barw 768B, 3B na piksel, linie zapisywane od dołu do góry(jest to wada)
- zalety: duża paleta barw
- wady:duży rozmiar pliku, długi czas ładowania, brak przezroczystości, brak animacji

GIF(Graphics Interchange Format)

- pozwala na kompresję obrazu(LZW -matematycznie bezstratna). *.gif
- maksymalnie 256 kolorów(8b)
- Palety barw:
 - paleta dokładna- skonwertowana w oparciu o barwy identyczne z występującymi w obrazie i należącymi do RGB, dla ilustracji <=256 kolorów
 - paleta bezpieczna www- opiera się na 216 barwach indeksowanej palety kolorów, zapewnia jednakowe wyświetlanie w dowolnym systemie i przeglądarce
 - paleta systemowa- 8b paleta kolorów opartą na ujednoliconych próbkach kolorów RGB wykorzystywaną przez każdy z systemów operacyjnych
 - paleta adaptacyjna- tworzy paletę barw na podstawie próbkowania kolorów najczęściej występujących w obrazie, redukuje barwy wraz z zachowaniem spektrum kolorystycznego ilustracji
- Roztrzaskanie- uzyskiwanie barw występujących w obrazie ale niedostępnych dla środowiska operacyjnego, oparta na złudzeniu optycznym
- Przezroczystość- ustawienie jednego/kilku kolorów jako niewidocznego tła ,brak regulacji stopnia przezroczystości
- Przeplot-wyswietlanie co n-tej linii podczas ładowania obrazu, mózg uzupełnia braki
- Animacja- wyswietlanie kolejno wielu obrazów, regulacji podlega czas regulacji
- wady- tylko 256v kolorów, licencja na format, ryzyko niewłaściwego wyświetlania
- zalety-mała objętość(dzięki kompresji), przezroczystość, animacja, przeplot, optymalizacja palety kolorów

PNG (Portable Network Graphics)

- odpowiedź na GIF (ma być bezpłatny i pozbawiony ograniczeń GIF-a), *.png
- 48b kolorów, dla szarości 16b

- obsługa korekcji gamma- wspiera prezentowanie jasności i kontrastu bez zmian w różnych systemach prezentacji obrazu
- przezroczystość- możliwość określania stopnia przezroczystości(do 256 stopni)
- Przeplot-pierwszy obraz pojawia się 8x szybciej niż przy GIFie(koszt jakości), tekst 2x szybciej czytelny niż w GIFie
- wyświetlanie sekwencji(progresywne)- najpierw wyświetlane są kontury, potem szczegóły, obraz zapisany progresywnie jest nieco większy
- animacja-png nie obsługuje
- bezstratna 24b kompresja, kombinacja LZ77 i kodowania Huffmana stąd wyniki kompresji są lepsze od 5 do 25% niż w GIFie
- zalety-ilość kolorów, bezstratna kompresja,przezroczystość, przeplot, brak płatnych licencji
- wady-brak animacji, nie obsługiwany przez starsze przeglądarki

DjVu

- konkurencja dla GIF i JPEG, kompresja nierzadko przewyższa wspomniane formaty a przy tym umożliwia zapis bez utraty jakości
- format przewidziany do skanowania książek, konkurencja dla PDF-a

TIFF(Tagged Image File Format)

- stosowany w poligrafii, wiele odmian pod względem palet kolorów(1,2,4,8,24b/px) i kompresji (Huffman, LZW, Fax, Group 3, Fax Group 4)

SVG

- brak kompresji, format wektorowy, oparty na XML, konkurencja GIFa, *.svg
- pliki można tworzyć za pomocą zwykłego edytora tekstu

FLASH

- pliki wektorowe, *.swf, wysoka jakość obrazów, możliwa interaktywność
- filmy *.swf(kilka minut, z dźwiękiem) zajmuje zaledwie kilka kilobajtów

Wykład 8 - Kompresja bezstratna

Ogólnie

- Zapakowane tak informacje można odtworzyć do postaci identycznej jak przed rozpakowaniem,
- Pakują dobrze, jeżeli mamy nadmiarowość informacji

Kodowanie entropowe

Kodowanie binarne

- Strumień kodowany zmieniamy na sekwencję bitów,
- Najłatwiej użyć słownika,
- Jeżeli nie ma w słowniku, to używamy drzew binarnych
- Dekodując kolejne symbole schodzimy w dół 0 do lewego, 1 do prawego,
- W liściu znajduje się nasz symbol
- Sposoby:
 - Słownikowe na stałe w koderze i dekodерze
 - Słowa kodowe wyznaczane na podstawie statystyki występowania symboli
- szukamy drzewa jak najbardziej płyckiego - na to rozwiązanie znalazł Huffman, stąd nazwa.

kodowania Huffmana

Wejście:

- n symboli
- W_i waga i -tego symbolu

Wyjście:

- C_i słowo kodowe i -tego symbolu, takie że poprzednie słowa mają minimalną średnią długość kodu.

Algorytm:

1. Tworzymy las z wag
2. Dla $m = n + 1, 2n - 1$ wybierz z lasu L dwa drzewa o najmniejszych wagach i zastępujemy je drzewem, którego waga jest sumą wag wybranych drzew,
3. Dla $i = 1, n$ tworzymy słowo kodowe idąc od korzenia do i -tego węzła wg zasady na lewo 0, na prawo 1

Huffmana z ograniczeniem na długość słowa kodowego

- Gdy wagi symboli się bardzo różnią słowniki mogą być spore,
- dlatego używamy Huffmana z ograniczoną długością słowa, (w PNG długość kodu = 15)

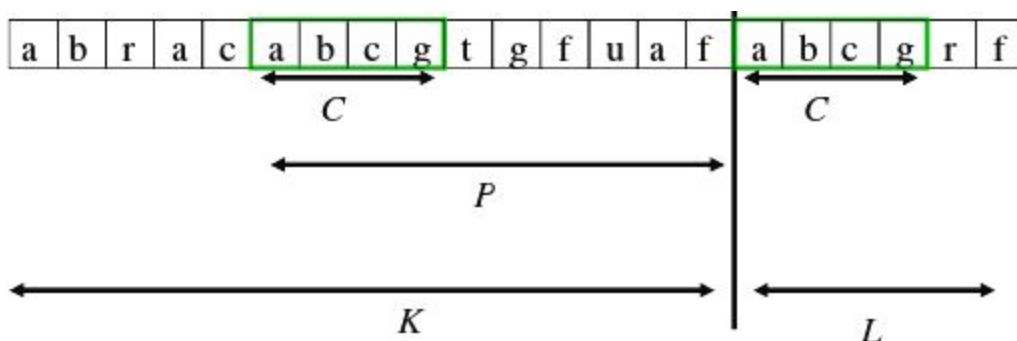
Dynamiczne kodowanie Huffmana

- Budujemy statystykę podczas napływania danych i co znak poprawiamy drzewo,
- Takie samo drzewo buduje się u odbiorcy,
- Mamy dwa algorytmy poprawiające drzewo: Gallera-Gallera-Knuta i algorytm Vittera

Kodowanie słownikowe

- Mamy sekwencję symboli
- Adres sekwencji w dynamicznym słowniku
- Zmieniamy sekwencje strumieni wejściowych na adresy tych sekwencji
- Kod adresu musi być niewielki w porównaniu do sekwencji.
- Koder i dekodery budują taki sam słownik
- Nie potrzebujemy analizy statystycznej,
- Algorytmy głównie różnią się strukturą słownika, jego tworzeniem, aktualizacją i adresowaniem,
- LZ w nazwach pochodzi od twórców,

LZ 77



1. Bufor słownikowy wypełniony pierwszym symbolem wejściowym i przetrzucany na wyjście,
2. Do bufora wejściowego wrzucane jest pierwsze n symboli,
3. Dopóki mamy jakieś dane w buforze wejściowym:
 - a. W buforze słownika szukamy najdłuższy podciąg taki jak początek bufora wejścia. Znajdujemy indeks początku tego podciągu i jego długość. Jeżeli nie znajdziemy podciągu, to ustawiamy długość na 0 :<
 - b. Na wyjście wyrzucamy początek podciągu, długość i symbol następujący po dopasowanym ciągu,
 - c. Przesuwamy bufory o tyle symboli ile zakodowaliśmy.
4. Aby zdekodować bufor słownikowy wypełniamy pierwszym symbolem, doklejamy do niego odpowiednie literki (o zapamiętanej pozycji i długości)

LZ78

- Słownik jest rozszerzany -> korzystamy z przetworzonych wcześniej danych
- Skomplikowany słownik, ale kompresja jest lepsza.

LZW

- Wprowadzony słownik ciągów zakodowanych,

- Elementy słownika uprzednio zakodowanego są ponumerowane,
- Początkowy słownik składa się ze wszystkich możliwych pojedynczych znaków
- W trakcie kodowania do słownika dochodzą ciągi,
- Do koderów są wysyłane tylko numery ciągów ze słownika,
- Koder i dekodek budują ten sam słownik.

Wykład 9 - Standardy H.261 i H.263

Przestrzenie kolorów

- Używa przestrzeni RGB
 - gorsze przedstawienie, gdyż czułość percepcji jest mniejsza niż przypadku luminancji
- YCC, YIQ, YUV
 -

Tryb intra

Tryb inter

Formaty koderów

Przykład analizy wymaganego stopnia kompresji

Koder H.261

Wykład 10 - MPEG-1

Ogólny opis

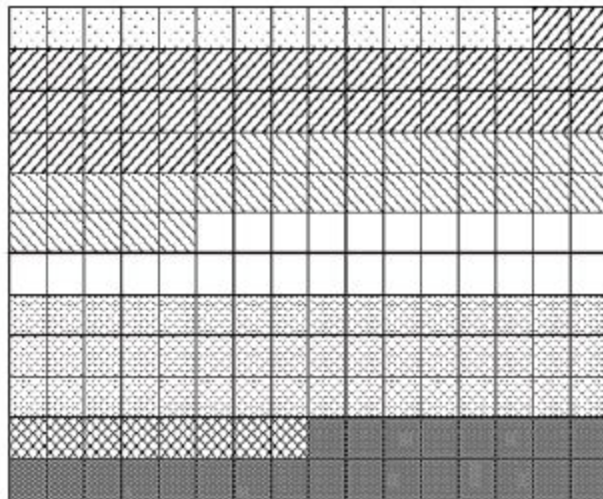
- Szybkie przewijanie w obie strony + swobodny dostęp - dzięki GOP (Group Of Pictures) - każdy blok musi zawierać ramkę typu I
- jakość VHS przy bitrate 1.5Mbps
- względnie szybkie (de)kodowanie
- brak przepływu
- próbkowanie chrominancji w formacie 4:2:0
- Dla tanich dekodek powstał CPB (Constrained Parameters Bitstream), obkrojone

parametry, musi być akceptowane przez każdy dekodery

- Max res: 720x576
- Max FPS: 30
- Max bitrate: 1,86Mbps

Porównanie z H.261

- H.261 służy tylko do kodowania video, a MPEG1 do video i Audio
- H.261 przewiduje tylko predykcje w przód. MPEG-1 ma predykcje w przód i w tył (B-pictures).
- ramki typu D też są tylko w MPEG-1 (miniaturki dla szybkiego podglądu). (tego nie było w wykładach)
- MPEG-1 posiada unikatowy typ klatki, która nie występuje w późniejszych standardach wideo. Nazywane również obrazkami DC (**DC-pictures**) są niezależne od zdjęcia (intra-frames), które zostały zakodowane tylko dla DC. Klatki typu D są bardzo niskiej jakości. Klatki typu D nigdy nie są zamienne z klatkami typu I, P, B. Ten typ klatek jest wykorzystywany jedynie do szybkiego podglądu wideo.
- H.261 obsługuje jedynie rozdzielczości CIF(352x288), QCIF (176x144)
- MPEG-1 obsługuje SIF (352x244 (dla NTSC) lub 352x288 dla PAL)
- Zamiast GOB (Group Of Blocks) w H.261, MPEG-1 może być podzielony na jeden lub więcej pasków (slices) (Slice - inaczej: GOB - Group Of Macroblocks)
 - każdy pasek jest kodowany niezależnie
 - mogą zawierać różną liczbę makrobloków na klatkę
 - mogą zaczynać i kończyć się w dowolnym miejscu
 - umożliwia wielowątkowe przetwarzanie
 - poprawiają odporność na błędy
- MPEG-1 posiada różne tablice kwantyzacji dla kodowania Inter i Intra
- MPEG-1 posiada wektory ruchu (MVs - Motion Vectors) z dokładnością do $\frac{1}{2}$ px lub 1 px
- H.261 posiada max zasięg MV ± 15 px, MPEG-1 [-512, 511.5] dla $\frac{1}{2}$ px dokładności oraz [-1024, 1023] dla 1 px dokładności
- MPEG-1 posiada swobodny dostęp do pliku dzięki GOP (Group Of Pictures), ponieważ każdy GOP posiada znacznik czasu
- H.261 posiada jedynie ramki typu I oraz P, natomiast MPEG-2 ma dodatkowo ramki B



Slices in an MPEG-1 picture

Struktura danych

- Wyróżnia stróktóre pośrednią - slice - złożona z wielu kolejnych w porządku rastrowym makrobloków, może zaczynać i kończyć się w dowolnym miejscu
- Blok 8x8
- Makroblok 4bloki luminacji i 2 chrominacji (4:2:0)
- Obraz
- Grupa obrazów GOP
- Sekwencja

Warstwa zawiera informacje resynchronizacji dla korekcji błędów transmisji

Tryb intra (I)

- standard określa stosowanie tablic kwantyzacji dla współczynników transformaty kosinusowej
- tablica kwantyzacji może być przesłana w nagłówku sekwencji lub może być użyta standardowa
- zalecane jest podanie współczynnika przemnarzającego tablice dla obrazu/warstwy/makrobloku

Tryb inter

- kompensuje ruch na podstawie obrazu referencyjnego
- transformata DCT między aktualnym blokiem 8x8, a najlepiej pasującym referencyjnym
- oprócz delty zapamiętuje wektor przesunięcia z dokładnością do $\frac{1}{2}$ px

Typy obrazów/klatek

- I - obraz intra
- P - obraz z 1 poprzedzającym obrazem referencyjnym typu I lub P
- B - 2 obrazy referencyjne 1 poprzedzający P lub I, kolejny wyprzedza w czasie

Wykład 11 - MPEG-2

Formaty obrazu, poziomy pracy

- Dopuszcza przeplot
- Poziomy pracy:

| Poziom | Max rozdzielczość | FPS |
|-------------------------|----------------------------------------|-----|
| Niski (low) | 352x288 | 30 |
| Główny (main) | 720x576 // 576=2·288, 720=2·360 | 30 |
| Wysoki 1440 (high 1440) | 1440x1152 // 1152=2·576, 1440=2·720 | 60 |
| Wysoki (high) | 1920x1152 | 60 |

Przeplot

Przeplot jaki jest, każdy wie...

- Film 1080i - z przeplotem (interlaced), 1080p - bez przeplotu (progressive)
- przeplot zwiększa płynność przy dużej dynamice (wft!?)
- nakładane są 2 klatki o połowicznej rozdzielczości są przeplatane:



Skalowalność

- Podział na warstwy, warstwa podstawowa jest obowiązkowa i zachowuje zgodność z przestarzałymi odtwarzaczami
- Warstwy rozszerzeń rozbudowują warstwę podstawową dodatkowymi informacjami polepszającymi jakość nagrania (na różne sposoby) w odtwarzaczach obsługujących dane rozszerzenie
- Lepsza odporność na błędy (efekt uboczny, korzystny)
- Rodzaje skalowalności (nazwa skalowalności - co robi rozszerzenie)
 - Przestrzenna - zwiększa rozdzielczość
 - SNR - zwiększa dokładność, lepsze odwzorowanie rzeczywistości
 - Czasowa - dodatkowe ramki typu B, zwiększa FPS
 - Podział danych - więcej współczynników transformaty

Profile

- Proste urządzenia nie muszą obsługiwać pełnej specyfikacji
- Dekoder obsługujący pewien format musi obsługiwać wszystkie niższego poziomu (model hierarchiczny) oraz musi dekodować MPEG-1 w specyfikacji CPB.
- Nie wszystkie kombinacje profili i poziomów są dozwolone

Profile:

1. **Prosty** - nieskalowalny, bez ramek B, format 4:2:0, szybkie (de)kodowanie
2. **Główny** - nieskalowalny, ramki B, format 4:2:0, najczęściej używany
3. **SNR** - skalowalność SNR, ramki B, format 4:2:0
4. **Przestrzenny** - skalowalności: SNR oraz przestrzenna, ramki B, format 4:2:0
5. **Wysoki** - skalowalny, dwie warstwy rozszerzeń (patrz wyżej↑), ramki B, format 4:2:0 lub 4:2:2
6. **4:2:2** - nieskalowalny, ramki B, format 4:2:0 lub 4:2:2, do zastosowań produkcji filmów i programów
7. **Wielowidokowy** - skalowalny, dozwolone 1 rozróżnienie, do kodowania sekwencji stereowizyjnych

| | Prosty | Główny | Snr-skal. | Przest-skal. | Wysoki |
|-------------|--------|--------|-----------|--------------|--------|
| Wysoki | | X | | | X |
| Wysoki 1440 | | X | | X | X |
| Główny | X | X | X | | X |
| Niski | | X | X | | |

Kodowanie

- Współczynniki DC są kodowane predykcyjnie. Kwantowane równomiernie, ze stałym krokiem.
- Współczynniki AC są kodowane krokiem zmiennym:
 - krok określa tablica kwantyzacji przemnożona przez współczynnik skalujący warstwy lub bloku
 - podobnie dla współczynników transformaty
- MPEG-2 definiuje kilka tablic kodów
- Kodek może pracować w 2 trybach:
 - CBR (const bitrate) - koder dopasowuje jakość poszczególnych bloków aby uzyskać określony bitrate
 - VRB (variable bitrate)

Wykład 12 - MPEG-4

Czemu nie MPEG-3?

- MPEG3 był przewidziany do HDTV, ale w TV przyjęto się MPEG-2, więc jego rozwój został porzucony.

Ogólny opis

H.264

- ~50% mniejszy bitrate od MPEG2 (przy tej samej jakości)
- wysoka jakość dla niskich bitrate
- lepsza korekcja błędów
- bitrate 64kbps - 240Mbps
- przystosowany do broadcastu (MPEG2 nie)

Two-Layer Structure

- Video Coding Layer (VCL)
 - efektywnie reprezentuje materiał wideo
 - dzieli wideo na paski o różnych wymiarach, każdy pasek posiada wszystkie informacje do jego prawidłowego zdekodowania, co umożliwia wielowątkowe dekodowanie, oraz minimalizuje błędy
 - paski mogą być wysyłane w dowolnej kolejności, bufor w odbiorniku je układa (podobnie jak przy jitterze)
 - Deblocking Filter: kompresja ruchu, używając transformacji 4x4 oraz block based
- Network Adaptation Layer (NAL) (pojębane to)
 - Umożliwia przesyłanie przez różne rodzaje sieci
 - Głównym założeniem, to w opakowaniu formatów usunąć duplikaty nagłówków

H.264 vs. MPEG-2

| | MPEG-2 | H.264 |
|-------------------------|---------------------------------------------------|------------------------------------------------------------|
| Rozmiar makrobloku | 16x16 (tryb ramki) 16x8 (tryb pola) | 16x16 |
| Transformacja | 8x8 | (16 8 4)x(16 8 4) |
| Kwantyzacja | 8x8 DCT | 4x4, 8x8 int DCT 4x4, 2x2 Hadamard |
| Kodowanie entropii | Skalarna z rozmiarem kroku o stałym przyroście | skalarna z rozmiarem kroku o wzroście na poziomie 12.5% |
| Estymacja i kompensacja | VLC | VLC, CAVLC, CABAC |
| Swobodny dostęp | TAK | TAK |
| Dokładność wektorów | ½ px | ¼ px |

| | | |
|----------------------------|---|----|
| ruchu | | |
| Ramki do predykcji ramek B | 2 | 32 |

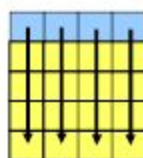
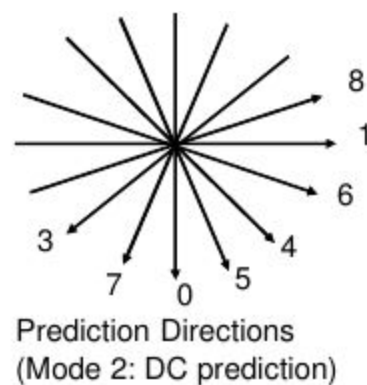
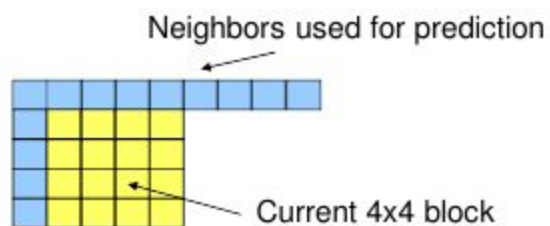
H.264 ma ~50% lepszy stosunek jakość/bitrate

Intra-Picture Prediction

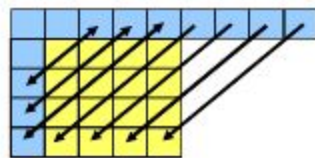
- Generowane przestrzennie
- Tryby:
 - intra 4x4 dla detali
 - intra 16x16 dla bardziej jednolitych
 - I_PCM bez predykcji, dane surowe, stosuje się aby załimitować ilość danych na blok

- **Intra_4x4 Prediction (9 modes)**

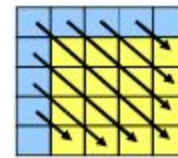
- Predict each 4 x 4 block
- Suitable for details



Mode 0



Mode 3

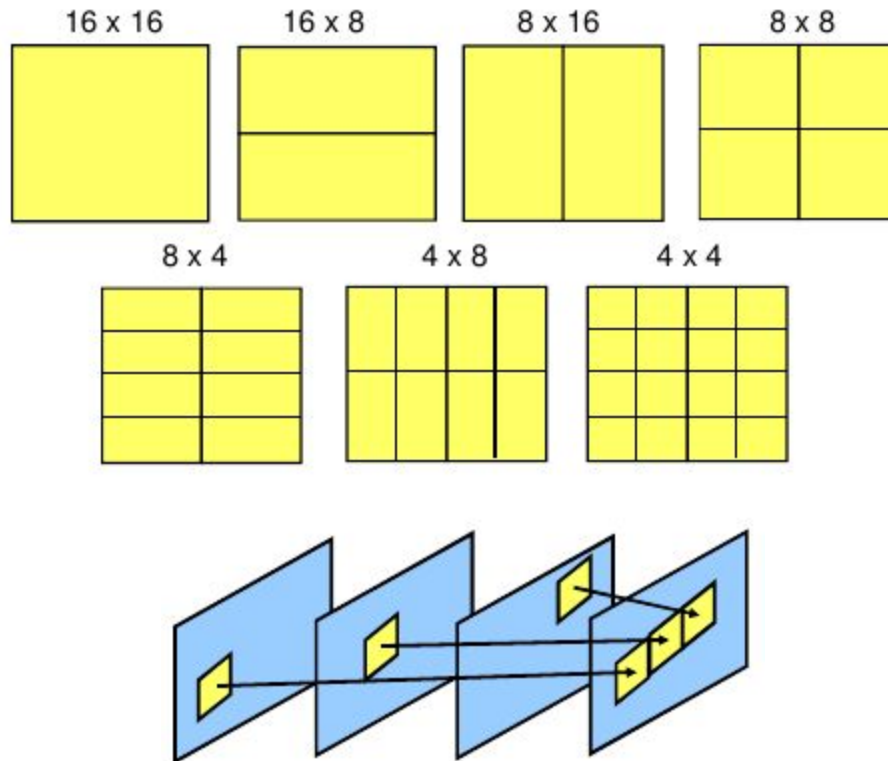


Mode 4

Inter-Picture Prediction

MV - Motion Vector

- do 16 MVs
- MVs są różnie kodowane
- Wymaga dużo optymalizacji aby znaleźć najlepszą konfigurację (ustawienie)



Entropy Coding

- CAVLC (Context Adaptive Variable Length Coding)
 - Na podstawie zakodowanych informacji w sąsiednich blokach
 - Zoptymalizowane tabele VLC są dostarczane dla każdego regionu aby zakodować współczynniki na podstawie analizy statystycznej (?)
- CABAC (Context Adaptive Binary Arithmetic Codes)
 - arytmetyka binarna
 - lepsza kompresja kontekstowa
 - oszczędność bitrate względem CAVLC 5~15%