

Wydział Elektryczny

Katedra Telekomunikacji i Aparatury Elektronicznej

Instrukcja do pracowni specjalistycznej:
Techniki Multimedialne

Temat ćwiczenia: Metody kompresji sygnału wideo

Numer ćwiczenia: 5 – 6

Opracowanie:
dr hab. inż. Ewa Świercz
mgr inż. Grzegorz Kraszewski

Białystok 2014

1. Metody kompresji sekwencji wideo

Przy transmisji sygnałów telewizyjnych DVB-T (naziemna emisja sygnałów telewizyjnych) stosowane są głównie dwa standardy kodowania sygnału wizyjnego: MPEG-2 i MPEG-4 (AVC/H.264). Sygnał telewizyjny jest najbardziej reprezentatywnym przykładem sygnału wideo, dlatego też wymienione algorytmy kompresji wykorzystywane praktycznie do przetwarzania sygnału telewizyjnego będą przedmiotem rozważań na pracowni specjalistycznej. Inne zastosowania sygnału video wraz ze standardami kompresji są podane w tabeli nr 1.

Tabela 1. Przykłady zastosowań sygnału wideo.

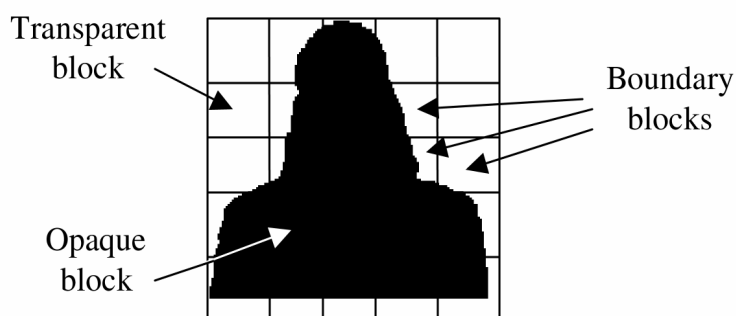
Cyfrowa transmisja telewizyjna	2 ... 5 Mbps (10 ... 20 Mbps dla HD)	MPEG-2 H.264/AVC
Płyty wideo DVD, HD-DVD, Blu-ray	4 ... 8 Mbps (10 ... 20 Mbps dla HD)	MPEG-2 H.264/AVC
Transmisja strumieniowa w Internecie	20 ... 600 kbps	Kodeki własnościowe podobne do H.263, MPEG-4, H.264/AVC, VC-1
Wideokonferencje, wideotelefony	20 ... 320 kbps	H.261, H.263, H.264/AVC
Transmisja przez łącza telefonii komórkowej 3G	20 ... 200 kbps	H.263, MPEG-4, H.264/AVC, VC-1

1.1 Cechy standardu MPEG-4

Podstawowym założeniem standardu MPEG-4 jest uniwersalność odnosząca się do:

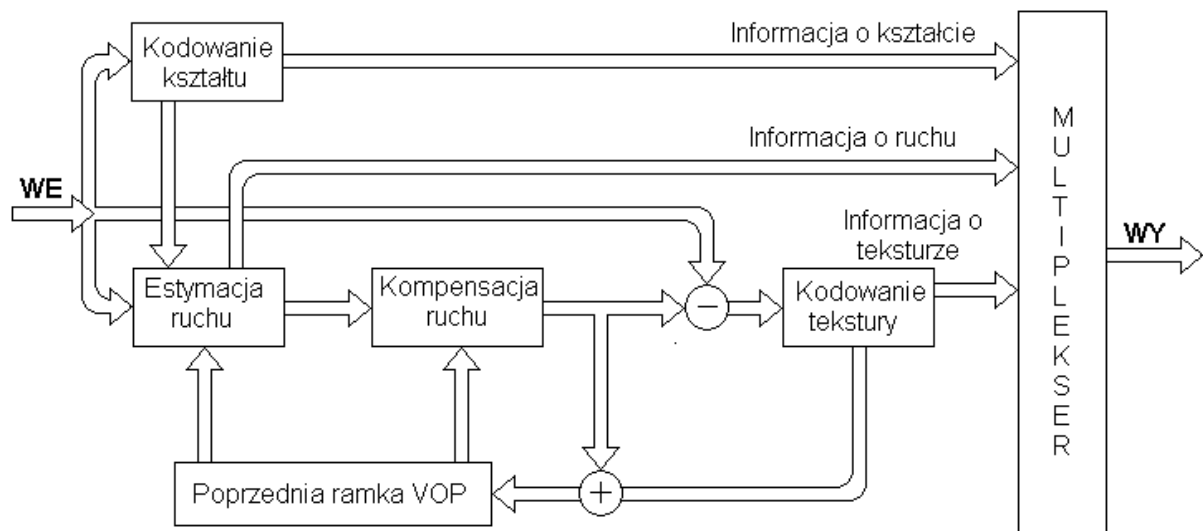
- rozdzielczości – od 176×144 (QCIF, quarter common intermediate format) do rozdzielczości przewyższającej HDTV (1280×720),
- przepływności binarnej – od 5 kB/s do 10 MB/s,
- stosowania lub nie przepłotu.

MPEG-4 jest oparty na koncepcji obiektowej, ze sceną, obiektami audio-wideo, możliwością ingerencji użytkownika. Na pracowni będą analizowane tylko sygnały wideo bez skojarzonego sygnału audio, więc same obiekty wideo będą oznaczane skrótem VO. Wyodrębnione obiekty o dowolnych kształtach, złożone z wielu makrobloków, mogą być kodowane niezależnie od nieruchomego tła poprzez kodowanie kształtu i tekstury, co pozwala na zwiększenie stopnia upakowania danych poprzez opis przesuwania się obiektów względem tła w kolejnych chwilach czasu.



Rys. 1. Kształt obiektu VO na tle otaczających makrobloków.

Przez teksturę należy rozumieć obraz po transformacji DCT (dyskretna transformata kosinusowa) lub transformacji falkowej. Standard MPEG-4 wprowadza elementy VOP (*Video Object Planes*) typu I, P i B, analogicznie do modelu GOP (*Group of Pictures*) znane z MPEG-2. VOP odpowiada zdjęciom obiektu VO w poszczególnych chwilach czasowych. Dzięki zastosowaniu VOP możliwe jest zapisywanie zmian położenia, kształtu i tekstury elementów obrazu w oszczędny sposób, jak w przypadku GOP.

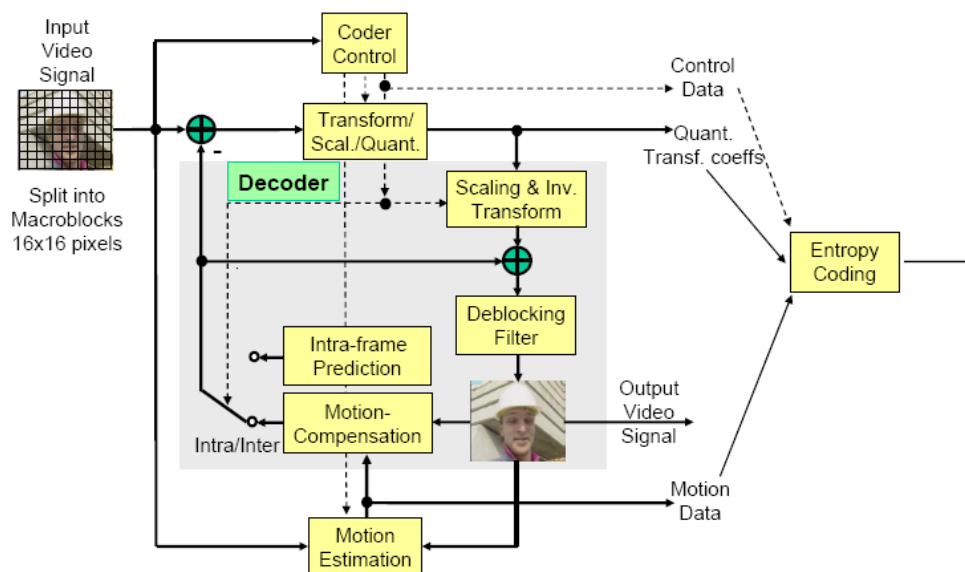


Rys. 2. Uproszczony schemat standardu MPEG-4.

Sekwencja VOP należąca do tego samego fizycznego obiektu sceny tworzy obiekt wideo VO. Zdefiniowano trzy typy VOP:

- I-VOP (ang. *Intra VOP*),
- P-VOP (ang. *Predicted VOP*),
- B-VOP (ang. *Bidirectionally VOP*).

W strumieniu danych zawarte są informacje potrzebne do syntezy i prezentacji kompletnej sceny wideo.



Rys. 3. Uproszczony schemat kodeka H.264/AVC, jednego z wielu kodeków MPEG-4.

1.2 Metody oceny jakości obrazu cyfrowego

Ocena jakości obrazów rekonstruowanych na podstawie reprezentacji utworzonej w procesie stratnej kompresji jest zagadnieniem wieloaspektowym, silnie zależnym od konkretnej aplikacji (klasy obrazów i sposobów ich wykorzystania). Oceny można dokonać poprzez subiektywne odczucia odbiorcy lub wykorzystać miary obiektywne, wynikające z modelu matematycznego obrazu. Obiektywna obliczeniowo miara jakości winna mieć następujące podstawowe cechy:

- łatwość obliczeniową,
- dużą korelację z wartością perceptualną (psychowizualna ocena jakości przez obserwatorów).

Tabela 2. Najczęściej stosowane obiektywne miary jakości obrazów.

1	średnia różnica	$AD = \frac{\sum_{x=1}^M \sum_{y=1}^N [f(x, y) - \hat{f}(x, y)]}{MN}$
2	zawartość strukturalna (<i>structural content</i>)	$AC = \frac{\sum_{x=1}^M \sum_{y=1}^N [f(x, y)]^2}{\sum_{x=1}^M \sum_{y=1}^N [\hat{f}(x, y)]^2}$
3	znormalizowana korelacja wzajemna (<i>normalized crosscorrelation</i>)	$NK = \frac{\sum_{x=1}^M \sum_{y=1}^N [f(x, y) \hat{f}(x, y)]}{\sum_{x=1}^M \sum_{y=1}^N [f(x, y)]^2}$
4	jakość korelacji (<i>correlation quality</i>)	$CQ = \frac{\sum_{x=1}^M \sum_{y=1}^N [f(x, y) \hat{f}(x, y)]}{\sum_{x=1}^M \sum_{y=1}^N [f(x, y)]}$
5	maksymalna różnica (<i>maximum difference</i>) lub szczytowy błąd bezwzględny (<i>peak absolute error</i>)	$MD = \text{Max} \{ f(x, y) - \hat{f}(x, y) \}$
6	wierność obrazu (<i>image fidelity</i>)	$IF = 1 - \frac{\sum_{x=1}^M \sum_{y=1}^N [f(x, y) - \hat{f}(x, y)]^2}{\sum_{x=1}^M \sum_{y=1}^N [f(x, y)]^2}$
7	laplasjanowy błąd średniokwadratowy (<i>laplacian mean square error</i>)	$LMSE = \frac{\sum_{x=1}^M \sum_{y=1}^N [O\{f(x, y)\} - O\{\hat{f}(x, y)\}]^2}{\sum_{x=1}^M \sum_{y=1}^N [O\{f(x, y)\}]^2}$
8	błąd średniokwadratowy (<i>mean square error</i>)	$MSE = \frac{1}{MN} \sum_{x=1}^M \sum_{y=1}^N [f(x, y) - \hat{f}(x, y)]^2$
9	szczytowy błąd średniokwadratowy (<i>peak mean square error</i>)	$PMSE = \frac{1}{MN} \frac{\sum_{x=1}^M \sum_{y=1}^N [f(x, y) - \hat{f}(x, y)]^2}{\sum_{x=1}^M \sum_{y=1}^N [\max\{f(x, y)\}]^2}$

10	znormalizowany błąd bezwzględny (<i>normalized absolute error</i>)	$NAE = \frac{\sum_{x=1}^M \sum_{y=1}^N f(x, y) - \hat{f}(x, y) }{\sum_{x=1}^M \sum_{y=1}^N f(x, y) }$
11	znormalizowany błąd średniokwadratowy (<i>normalized mean square error</i>)	$NMSE = \frac{\sum_{x=1}^M \sum_{y=1}^N [f(x, y) - \hat{f}(x, y)]^2}{\sum_{x=1}^M \sum_{y=1}^N [f(x, y)]^2}$
12	norma L_p (Minkowskiego)	$L_p = \left\{ \sum_{x=1}^M \sum_{y=1}^N f(x, y) - \hat{f}(x, y) ^p \right\}^{\frac{1}{p}}$
13	stosunek sygnału do szumu (<i>signal to noise ratio</i>)	$SNR = 10 \log_{10} \left(\frac{\sum_{x=1}^M \sum_{y=1}^N [f(x, y)]^2}{\sum_{x=1}^M \sum_{y=1}^N [f(x, y) - \hat{f}(x, y)]^2} \right)$
14	szczytowy stosunek sygnału do szumu (<i>peak signal to noise ratio</i>)	$PSNR = 10 \log_{10} \left(\frac{[\max\{f(x, y)\}]^2}{\sum_{x=1}^M \sum_{y=1}^N [f(x, y) - \hat{f}(x, y)]^2} \right)$

2. Badania wybranych algorytmów kompresji sekwencji wideo

2.1. Zasady korzystania z komputerów w pracowni

- W czasie ćwiczeń korzystamy z konta *Student*, logując się hasłem *multimedia*.
- Dyskiem roboczym do pracy z materiałami multimedialnymi jest dysk *D*:
- Pliki umieszczane w innych miejscach (dysk *C*:, w tym pulpit Windows, karta SD aparatu) mogą być bez ostrzeżenia skasowane między kolejnymi zajęciami.
- Dysk *D*: nie jest czyszczony po każdych zajęciach, ale nie zabezpiecza to przed skasowaniem plików np. przez „życzliwych” kolegów z innych grup. Chcąc zatem zachować pliki dla potomności, trzeba je zapisać na własny nośnik USB.

2.2. Przygotowanie sekwencji wideo

Do badań potrzebne będą dwie sekwencje obrazów, każda o czasie trwania około 30 sekund, pozyskane aparatem cyfrowym. Pierwsza sekwencja powinna być statyczna, czyli zachować należy stałe tło, z niewielkim ruchem aktora. Druga sekwencja, dynamiczna, powinna mieć zmieniające się w czasie tło, co można łatwo uzyskać np. powoli chodząc z aparatem po sali w czasie filmowania i wykonując łagodnie zmieniające się, panoramiczne ujęcie.

Przed filmowaniem zdejmujemy osłonę obiektywu i włączamy aparat przyciskiem *POWER* na górze. Pokrętkiem ustawiamy tryb pracy kamery. Sprawdzamy czy ustawiona jest rozdzielczość filmu 720p (1280 × 720 pikseli), jest to sygnalizowane napisem „HD” w prawym górnym rogu wyświetlacza. Filmowanie rozpoczynamy i zatrzymujemy przez wciśnięcie do końca przycisku fotografowania, albo przyciskiem nagrywania (z czerwoną kropką) położonym obok.

W czasie filmowania ujęcia należy unikać gwałtownych ruchów aparatem. Po sfilmowaniu mniej więcej trzydziestosekundowego fragmentu zatrzymujemy zapis, aparat zapisuje sekwencję na karcie pamięci w formacie MPEG-4.



Rys. 4. Aparat fotograficzny w laboratorium, elementy sterujące.

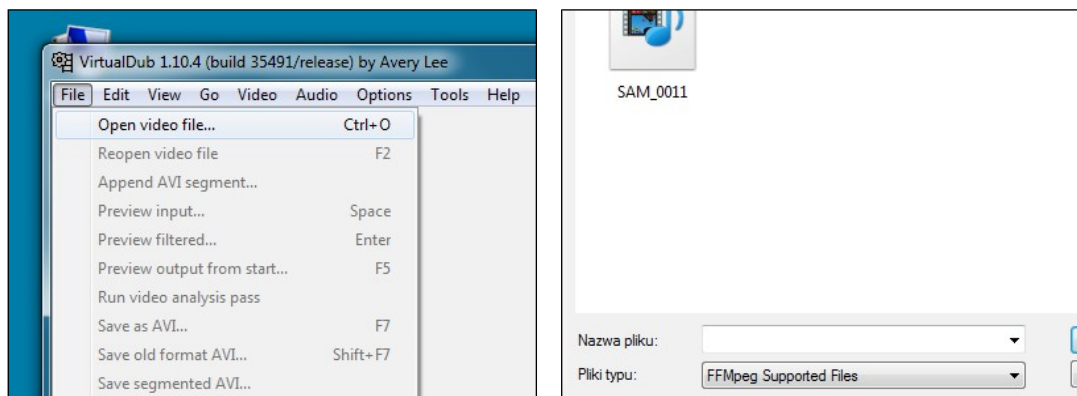
Kolejnym krokiem jest skopiowanie filmu na komputer. Otwieramy klapkę z lewej strony aparatu i podłączamy kabel do gniazda micro USB. Drugi koniec oczywiście do komputera, do jednego z gniazd USB na przednim panelu. Następnie wciskamy przycisk *POWER* w aparacie. Komputer rozpoznaje aparat jako dwa urządzenia: napęd CD (zazwyczaj *F:*) zawierający oprogramowanie, oraz dysk wymienny (zazwyczaj *G:*) zawierający zdjęcia i filmy. Film, nad którym będziemy pracować, kopiujemy na dysk *D:* komputera, po skopiowaniu możemy odłączyć aparat korzystając z normalnej procedury bezpiecznego odłączenia w systemie Windows.



Rys. 5. Aparat fotograficzny w laboratorium, gniazda.

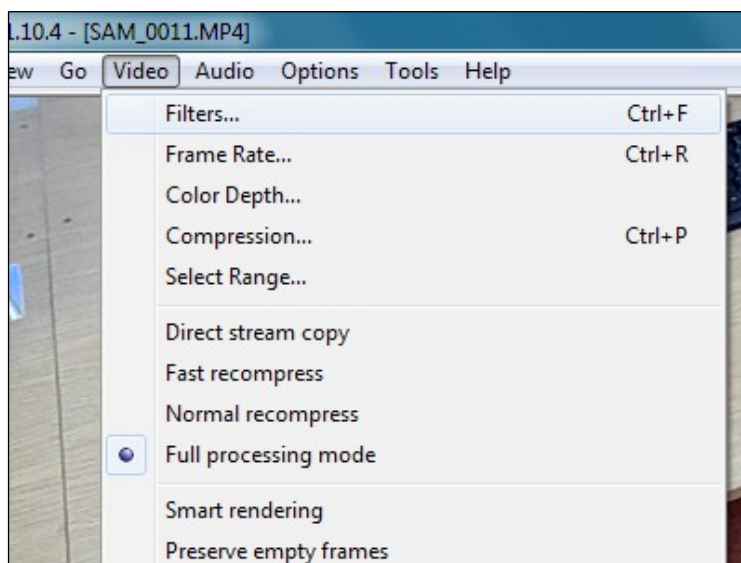
2.3. Operacje w programie *VirtualDub*

Filmy w formacie MPEG-4 przekształcamy do sekwencji obrazków PNG, która będzie materiałem źródłowym do wykonania obliczeniowej części ćwiczenia. Posługujemy się w tym celu programem *VirtualDub*. Oprócz konwersji do sekwencji klatek, skalujemy materiał do rozdzielczości 640×360 pikseli. To działanie ma dwa cele. Po pierwsze przetwarzanie dużych klatek w *Matlabie* jest znacznie wolniejsze, po drugie zaś usuwamy z materiału artefakty kompresji MPEG-4 wprowadzone przez aparat. Te ślady kompresji mogłyby nakładać się na zniekształcenia wprowadzane przez badane w ćwiczeniu kodeki.



Rys. 6. Ładowanie filmu do programu *VirtualDub*.

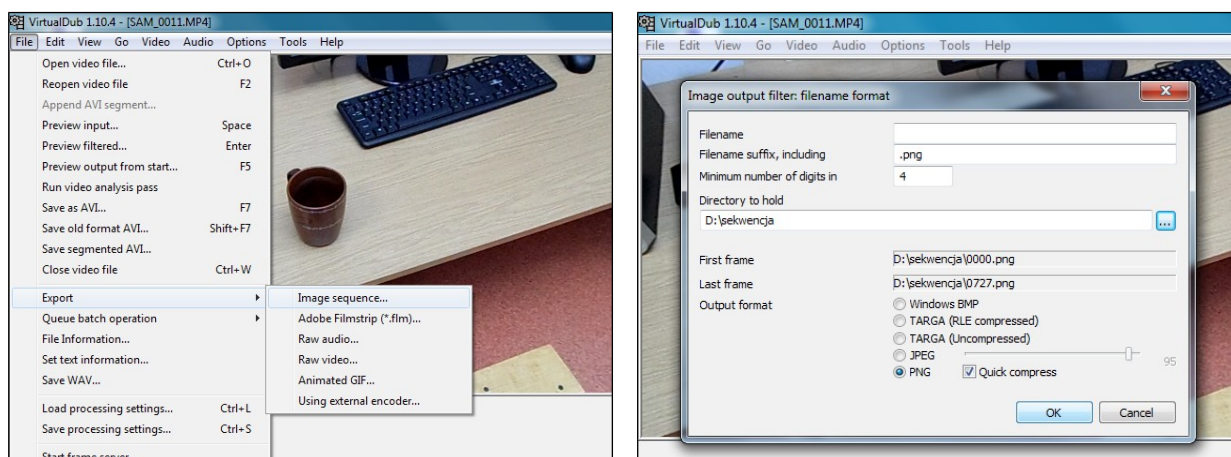
Materiał video wczytujemy do *VirtualDuba* opcją *Open video file*. W oknie wyboru pliku trzeba wybrać typ „FFMpeg supported files”. W przeciwnym wypadku otrzymamy komunikat o braku kodeka. Aparat zapisuje filmy używając kodeka *Advanced Video Coding (AVC1)*, format ten jest rozpoznawany przez wtyczkę do *VirtualDuba* opartą na kodekach z programu *FFMpeg*. Po załadowaniu filmu w oknie programu pojawiają się dwa podglądy: po lewej materiał źródłowy, po prawej wyjściowy. Na początku oba podglądy są identyczne. Ponieważ dwa podglądy nie mieszczą się w całości na ekranie, można oba zmniejszyć do 50%, korzystając z menu kontekstowego pod prawym przyciskiem myszy na każdym z podglądów.



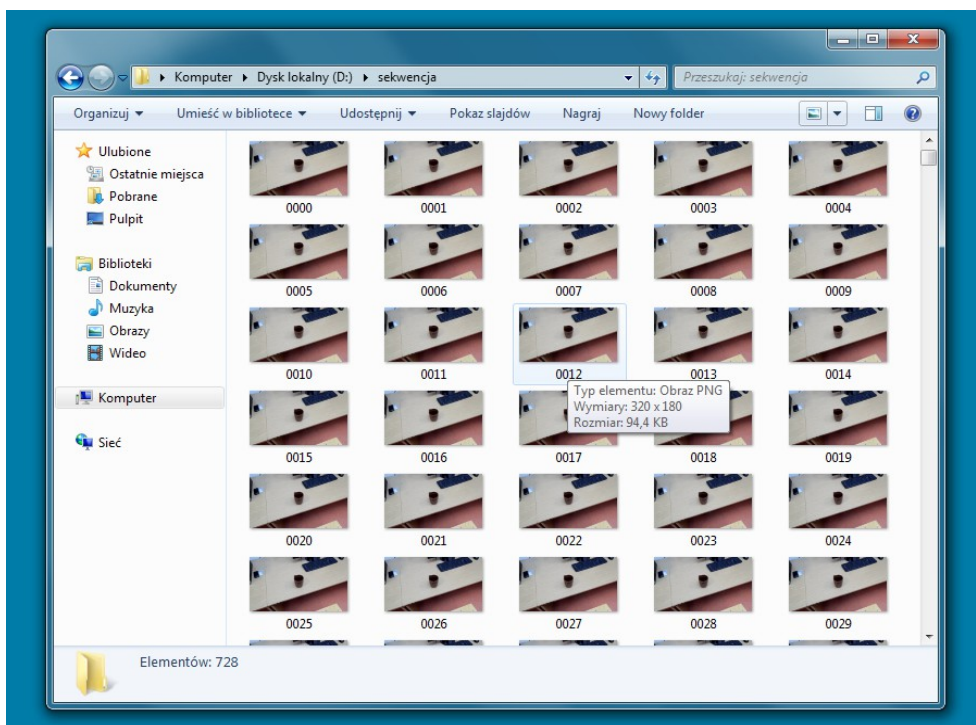
Rys. 7. Dodawanie filtru w *VirtualDubie*.

Kolejnym krokiem jest dodanie filtru skalującego obraz. Z menu *Video* → *Filters* otwieramy okienko filtrów i dodajemy filtr *2:1 reduction (high quality)*. Po dodaniu filtru rozmiary podglądu wyjściowego zmniejszą się dwukrotnie.

Teraz można już wyeksportować materiał video jako sekwencję obrazków w formacie PNG. W oknie konfiguracyjnym eksportu należy przede wszystkim zmienić format na PNG (rysunek po prawej). Obrazki zapisujemy w podkatalogu na dysku *D:*, podkatalog ten musi być stworzony wcześniej.



Rys. 8. Eksport materiału video do serii obrazków PNG.



Rys. 9. Wyeksportowana seria obrazków w folderze Windows.

2.4. Enkodowanie materiału wideo

Zapisanie filmu skompresowanego wybranym kodekiem następuje poprzez zewnętrzny enkoder *MEncoder*. Korzystamy z przygotowanych profili, poprzez menu *File* → *Export* → *Using external encoder*. Po wyborze profilu kodeka program otworzy okno wyboru pliku docelowego. Po wskazaniu tego pliku rozpocznie się kodowanie. Na tym etapie należy pamiętać o następujących detalach:

1. W nazwie pliku wynikowego nie może być spacji.
2. Ponieważ korzystamy z zewnętrznego kodeka, *VirtualDub* nie nadaje automatycznie rozszerzeń plikom. Należy zrobić to ręcznie, używając rozszerzenia **.mpg* dla kodeka MPEG-1 i MPEG-2, oraz **.mp4* dla kodeka MPEG-4/H.264.
3. Filmy w kodowaniu MPEG-1 nie są odtwarzane przez program *Windows Media Player*, uruchamiany po dwukliku na film. Trzeba je otwierać w programie *SMPlayer*, np. przez menu kontekstowe pod prawym przyciskiem myszy.

Łącznie zdefiniowanych jest 8 profili kodeków (5 profili MPEG-4/H.264 z różnymi przepływnościami bitowymi, dwa profile MPEG-2 i jeden profil MPEG-1). Oba nakręcone filmy kompresujemy wszystkimi profilami, po czym wypełniamy następującą tabelę (dla każdej z dwóch sekwencji):

Tabela 3. Pomiar efektywności kompresji.

Nr	kodek i przepływność [kbps]	wielkość pliku przy kompresji bezstratnej (obrazki PNG)	wielkość pliku po kompresji	efektywność kompresji
1.				
2.				
3.				

2.5. Obiektywna analiza ilościowa

Ocena obiektywna polega na ilościowym porównaniu wybranej klatki z wzorcowej serii obrazków PNG z tą samą klatką wyciętą z filmu skompresowanego jednym z badanych kodeków. Klatki do badań wycinamy z filmów ponownie posługując się programem *VirtualDub*. Po załadowaniu pliku wideo dokonuje się wyboru miejsca z którego nastąpi wycięcie klatki. za pomocą suwaka. Po wyborze klatki, kopiujemy ją do systemowego schowka (menu *Video* → *Copy output frame to clipboard*). Klatkę ze schowka wklejamy do programu *Paint* i zapisujemy w formacie PNG. Uwagi:

1. Przy wycinaniu klatek z filmów należy zwrócić uwagę, czy z okna filtrów programu został usunięty wcześniej używany filtr *2:1 reduction*, jeżeli nie, trzeba go usunąć.
2. Kodek MPEG-1 gubi się w numeracji klatek, jeżeli film przewijany jest do tyłu. Dlatego chcąc wyciąć określoną klatkę, należy przesuwać się suwakiem od początku filmu, zatrzymując się przed żądaną klatką i precyzyjnie dotrzeć do niej klawiszami kursora.

Porównując wybraną klatkę materiału wzorcowego z serii obrazków PNG, z klatką poddaną kompresji, wczytujemy obie do *Matlaba* poleceniem *imread*. Następnie obie klatki konwertujemy do obrazu intensywności, funkcją *rgb2gray*. Tak przygotowane obrazy intensywnościowe porównujemy ze sobą wykorzystując podane miary jakości:

$$\begin{aligned}
 \checkmark \quad AD &= \frac{\sum_{x=1}^M \sum_{y=1}^N [f(x,y) - \hat{f}(x,y)]}{MN} && \text{– średnia różnica} \\
 \checkmark \quad MD &= \text{Max} \{|f(x,y) - \hat{f}(x,y)|\} && \text{– maksymalna różnica} \\
 \checkmark \quad MSE &= \frac{1}{MN} \sum_{x=1}^M \sum_{y=1}^N [f(x,y) - \hat{f}(x,y)]^2 && \text{– błąd średniokwadratowy} \\
 \checkmark \quad NAE &= \frac{\sum_{x=1}^M \sum_{y=1}^N |f(x,y) - \hat{f}(x,y)|}{\sum_{x=1}^M \sum_{y=1}^N |f(x,y)|} && \text{– znormalizowany błąd bezwzględny} \\
 \checkmark \quad SNR &= 10 \log_{10} \left(\frac{\sum_{x=1}^M \sum_{y=1}^N [f(x,y)]^2}{\sum_{x=1}^M \sum_{y=1}^N [f(x,y) - \hat{f}(x,y)]^2} \right) && \text{– stosunek sygnału do szumu}
 \end{aligned}$$

Miara AD odpowiada średniej wartości różnicy intensywności pomiędzy pikselami dwóch rozpatrywanych obrazów. Obraz kompresowany jest lepszej jakości gdy miara ta dąży do zera.

Miara MD odpowiada największej różnicy intensywności pomiędzy pikselami badanych obrazów i w tym przypadku podobnie jak poprzednio, im wynik jest mniejszy tym jakość obrazu skompresowanego jest lepsza.

Miara AC określa stopień podobieństwa rozpatrywanych obrazów. Wartość tej miary powinna oscylować w granicach jedności aby uznać że obrazy są podobne.

Miara MSE reprezentuje kwadrat różnicy między obrazami objętymi badaniami, i tak jak poprzednio im jest ona mniejsza tym jakość obrazu kompresowanego jest lepsza.

Miara NAE różni się tym od MSE, że zamiast różnicę obrazów podnosić do kwadratu, wyciągamy z niej moduł. Wartość tej miary powinna być jak najmniejsza aby kompresowany obraz był jak najlepszej jakości.

Miarę SNR można traktować jako stosunek sygnału do szumu, określa on w jakim stopniu sygnał rozpatrywany, w naszym przypadku obraz skompresowany, jest zaszumiony pasożytniczymi sygnałami. W przypadku tej miary wskazaniem jest aby przybierała ona jak największe wartości.

3. Bibliografia

1. „*Kompresja danych – wprowadzenie*”, Khalid Sayood, wyd. ReadMe, Warszawa 2002
2. „*Multimedia. Algorytmy i standardy kompresji*” /pod red. W. Skarbka/ Wawa, Ak. Oficyna Wydawn. PLJ, 1998.
3. „*Metody kompresji obrazu dla potrzeb przetwarzania w czasie rzeczywistym*”, M. Antońkiewicz, Praca dyplomowa magisterska Politechnika Białostocka, Wydział Elektryczny, Białystok 2003 r.
4. „*Sprzętowe architektury kompresji obrazu wizyjnego dla potrzeb przetwarzania w czasie rzeczywistym*”, K. Wiatr, P. Russek – Kwartalnik Elektroniki i Telekomunikacji, PAN, T.44, z.2, 1998.
5. *Image Processing Toolbox – for Use with Matlab 6.0.*
6. „*Symulator kodera obrazów z wykorzystaniem elementów standardu MPEG-2*”, M. Jaworska, Praca dyplomowa magisterska, WE, 2007 r.

7. „*Obraz cyfrowy*”, M. Domański, WKŁ, 2010.
8. „*Modele i narzędzia oceny jakości obrazów cyfrowych*”, R. Kujawski, praca dyplomowa – 2001 r.
9. „*Matlab 6 – poradnik użytkownika*”, B. Mrozek, Z. Mrozek, Wyd. Go, Kraków 1994.
10. Internet.