

SKRYPT DO LABORATORIUM

Wymiana i Składowanie Danych Multimodalnych

ĆWICZENIE 4: Kompresja dźwięku i obrazu

**autor:
dr inż. Adam Bujnowski**



KAPITAŁ LUDZKI
NARODOWA STRATEGIA SPÓJNOŚCI

UNIA EUROPEJSKA
EUROPEJSKI
FUNDUSZ SPOŁECZNY



Gdańsk, 2010

1. Wymagania wstępne

1.1. Ustawienia

Umiejętność posługiwania się komputerem w zakresie obsługi systemu plików w systemie Windows. Znajomość podstaw programu Matlab. Posługiwanie się programem Audacity i VirtualDub

Cele ćwiczenia:

Wykaz przyrządów, materiałów i aparatury niezbędnej do przeprowadzenia ćwiczenia:

Komputer klasy PC z systemem operacyjnym co najmniej MS Windows XP, z zainstalowanym środowiskiem MatLab, programem Audacity z biblioteką Lame oraz program VirtualDub. Dostęp do internetu. Wskazany zestaw głośników multimedialnych.

Spodziewane efekty kształcenia - umiejętności i kompetencje:

Znajomość podstawowych metod kompresji stratnej i bez stratnej dźwięku i obrazu video. Umiejętność oceny współczynnika kompresji. Umiejętność oszacowania jakości kompresji.

Metody dydaktyczne:

Student dokona kompresji wybranych fragmentów utworów muzycznych i video z użyciem dostępnych bibliotek do kompresji z różnymi ustawieniami ich parametrów. Następnie dokona wyznaczenia współczynnika kompresji oraz wyznaczenia podstawowych parametrów współczynników kompresji.

Zasady oceniania/warunek zaliczenia ćwiczenia

Podstawą zaliczenia ćwiczenia jest zrealizowanie wszystkich obowiązkowych elementów ćwiczenia i zaprezentowanie ich prowadzącemu. Ocenę bardzo dobrą można uzyskać realizując wszystkie zadania dodatkowe zawarte w instrukcji oraz zlecone przez prowadzącego.

Wykaz literatury podstawowej do ćwiczenia:

1	Skrypt do wykładu
2	

2. Przebieg ćwiczenia

L.p	Zadanie
1.	Badanie współczynnika kompresji audio dla formatów mp3 i ogg
2.	Analiza podobieństwa materiałów skompresowanych i po dekompresji dla formatów mp3 i ogg
3.	Badanie maksymalnego możliwego do osiągnięcia współczynnika kompresji dla mp3 i ogg
4.	Inne metody poprawy współczynnika kompresji
5.	Analiza wymagań przepustowości kanału dla transmisji video z różnymi formatami
6.	Szacowanie i badanie współczynnika kompresji dla wybranych fragmentów utworów video
7.	Testy porównawcze różnych dostępnych algorytmów kompresji video
8.	Szacowanie udziału kanału audio w strumieniu video
9.	Zadania dodatkowe

UWAGI!**3. Wprowadzenie do ćwiczenia**

Kompresja jest działaniem, którego celem jest zmniejszenie rozmiarów materiału podlegającego kompresji. Jeśli K jest materiałem po kompresji, O - oryginałem to F w przekształceniu:

$$K = F(O)$$

będzie funkcją realizującą algorytm kompresji. Kompresja ma miejsce wtedy, gdy rozmiar K jest mniejszy od rozmiaru O :

$$\text{size}(K) < \text{size}(O)$$

Wówczas możemy mówić o współczynniku kompresji definiowanym min jako:

$$R = \frac{\text{size}(K)}{\text{size}(O)}$$

często wyrażanym w procentach:

$$R = \frac{\text{size}(K)}{\text{size}(O)} \cdot 100 \%$$

Tak zdefiniowany współczynnik kompresji określa ile procent z oryginału zajmuje jego skompresowany obraz.

Aby można było skorzystać z K musimy taki obiekt danych poddać dekompresji. Dekompresja D jest działaniem odwrotnym do Kompresji F i dostarcza nam O' takiego że:

$$O' = D(K)$$

Jeśli O' jest tożsame z O to mówimy o kompresji bez stratnej. Jeśli O' tylko przypomina O - to mamy do czynienia z kompresją stratną.

Ponieważ w kompresji stratnej O' nie jest tożsame z O to oznacza, że w istocie są to dwa różne obiekty (danych) podobne do siebie. Na ile podobne? Jeśli potraktujemy O i O' jako ciągi bajtów (tak są zapisywane w komputerze) to możliwe będzie wprowadzenie kilku miar podobieństwa.

MSE (Mean Square Error) jest miarą podobieństwa wykorzystywaną powszechnie do szacowania podobieństwa w przetwarzaniu obrazów. Zaadoptujmy to dla szacowania podobieństwa dźwięków:

$$MSE = \frac{1}{N^2} \sum_{n=1}^N \sum_{m=1}^2 [O(m, n) - O'(m, n)]^2$$

gdzie N jest długością wektora reprezentującego dany sygnał dźwiękowy, M=2 dla dwóch kanałów (zakładamy dźwięk stereofoniczny). Im MSE jest mniejsze tym O i O' są bardziej podobne. MSE wyraża się w energii sygnału, aby wyrazić to w wartościach zgodnych z miarą sygnału (dźwięk może być reprezentowany jako chwilowa zmiana napięcia w czasie) używa się RMSE:

$$RMSE = \sqrt{MSE}$$

Ta miara błędu wyrażona jest w wartościach jakimi mierzony jest sygnał. Tutaj podobnie: im mniejsze RMSE tym O i O' jest bardziej podobne.

Inną miarą jest PSNR. Definiuje się ją jako:

$$PSNR = 10 \cdot \log_{10} \left(\frac{MAX_I^2}{MSE} \right) = 20 \cdot \log_{10} \left(\frac{MAX_I}{RMSE} \right)$$

gdzie MAX_I jest największą możliwą wartością sygnału w O. PSNR wyrażany jest w decybelach. Tutaj im większe PSNR tym O i O' są bardziej podobne.

PSNR nie mogą być jednak używane do porównywania algorytmów przetwarzania sygnałów dla różnych wzorców. Konieczne jest badanie algorytmów względem takiego samego, arbitralnie wybranego wzorca.

4. Realizacja ćwiczenia

Przygotowanie do zajęć.

Aby możliwa była realizacja ćwiczeń należy zaopatrzyć się w dane źródłowe oraz sprawdzić stan posiadanego oprogramowania.

UWAGA ! Ze względu na duże rozmiary wszystkich przygotowywanych plików należy wszystkie testy przeprowadzać na dysku D: komputera PC dostępnego w laboratorium. W tym celu zaleca się założenie katalogu D:\WISDM\LAB4.

Do tego katalogu należy skopiować pliki:

MikeNewCar.avi , Track03.wav lub Track06.wav oraz anh_enhanced_avi1.zip

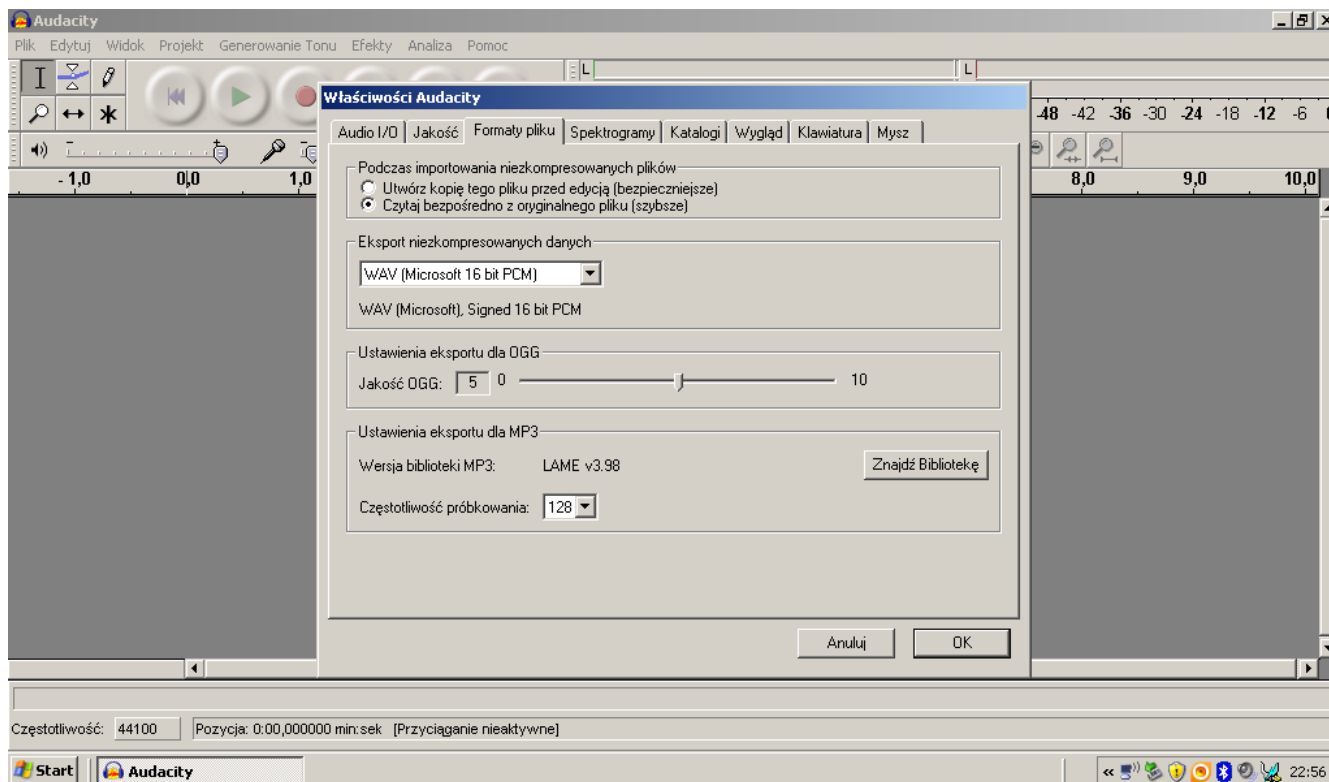
Pliki dostępne są na serwerze, adres wskazany przez prowadzącego.

Należy skontrolować dostępność programów :

Audacity, VirtualDub oraz DivXCreate na lokalnym komputerze. W przypadku braku potrzebnych programów należy je zainstalować, korzystając z uprawnień przydzielonych przez prowadzącego.

Kompresja dźwięku.

Zapoznać się z obsługą programu Audacity.



Zaimportować wybrany utwór muzyczny, dokonać jego eksportu do pliku *.mp3 oraz do pliku *.ogg z domyślnymi ustawieniami eksportu (dla mp3 zastosować bitrate 128kbps, dla ogg współczynnik 4).

Zadanie: Dla tak otrzymanych plików wyznaczyć współczynniki kompresji.

Zadanie dodatkowe: Wyznaczyć zależność współczynnika kompresji dla plików ogg od parametru Q – wykreślić dane na wykresie.

Zadanie dodatkowe: Przygotować podobny wykres dla mp3. Jako oś x użyć parametru bitrate.

Badanie jakości kompresji stratnej:

Z badanego utworu wyciąć fragment ok 15 sekund utworu. Fragment ten zapisać jako „orig.wav”.

Przywrócić oryginalne ustawienia kompresji dla mp3 (128kbps) i ogg (Q=4).

Wczytać do programu Audacity plik „orig.wav”. Dokonać eksportu do formatu mp3 zapisując go jako „fr.mp3” oraz do ogg jako „fr.ogg”.

Teraz należy zamknąć wszystkie otwarte pliki w programie Audacity, lub zamknąć i ponownie uruchomić ten program. Wczytać do niego plik „fr.mp3” i wyeksportować go jako „frmp3.wav” - w formacie wav. Podobnie - wczytać plik „fr.ogg” i wyeksportować go jako wav do pliku „frogg.wav”.

W rzeczywistości odtwarzanie pliku mp3 lub ogg polega na jego dekompresji na postać przypominającą oryginał i wysłanie takiego ciągu na głośnik. W tej części ćwiczenia zapisaliśmy do plików to co usłyszelibyśmy podczas normalnego odtwarzania tak nagranych utworów. Pozwala to na oszacowanie parametrów MSE i PSNR dla tak przygotowanych plików.

Tą część ćwiczenia zrealizujemy w programie Matlab. W tym celu uruchomimy program Matlab i przejdźmy do katalogu, gdzie zapisywane były wszystkie pliki:

```
>> cd d:\wisdm\lab4
>> pwd
>> ans

d:\wisdm\lab4
```

Teraz można wczytać wszystkie potrzebne pliki. Matlab pozwala na wczytywanie plików przy pomocy polecenia wavread. Aby wczytać plik orig.wav należy posłużyć się poleceniem:

```
>> o = wavread('orig.wav');
```

spowoduje to wczytanie zawartości tego pliku do zmiennej o. W przypadku dźwięku stereo o będzie macierzą o rozmiarach Nx2.

Zadanie: Wczytaj wszystkie pliki do programu Matlab. Wyznacz MSE i PSNR pomiędzy „orig.wav” i „frmp3.wav” oraz pomiędzy „orig.wav” i „frogg.wav”.

Uwaga ! Wczytane pliki orig i mp3 różnią się rozmiarami ! Do analizy wybierz tylko fragment pliku mp3 o długości równej długości pliku orig .

Wskazówka: Warto sobie przygotować plik wsadowy:

```
o = wavread('orig.wav');
op = wavread('frmp3.wav'); % tu przerabiamy dla ogg
op = op(size(o));
MSE = 1/(size(o,1)*size(o,2)) * sum(sum( (o-op).^2 ));
```

```
RMSE = sqrt(MSE);  
MX = max(max(o));  
PSNR = 20*log10(MX/RMSE);  
disp(['MSE=' num2str(MSE) 'RMSE=' num2str(RMSE) 'PSNR=' num2str(PSNR)  
]);
```

Mając tak przygotowane pliki możliwe jest odsłuchanie różnic pomiędzy sygnałami:

```
wavplay(o-op,44100);
```

Użycie tej funkcji będzie miarodajne jeśli do zestawu komputerowego podłączy się zestaw głośników. Zaleca się wykonanie tego punktu na jednym, wybranym komputerze w formie demonstracji.

Porównać w ten sposób kompresję ogg i mp3 (odsłuch, miary podobieństwa).

Zadanie dodatkowe: Wykonać podobne porównanie dla innych ustawień kompresji

Zadanie dodatkowe: Przygotować wykres $MSE = f(Q)$ dla ogg (RMSE, PSNR)

Zadanie dodatkowe: Przygotować wykres $MSE = f(\text{bitrate})$ dla MP3

Sprawdzić możliwość zastosowania zmniejszenia częstotliwości próbkowania i dynamiki sygnału, skomentować wpływ tych zmian na ilość zapisanej informacji. Odnieść to do możliwych obszarów zastosowań poszczególnych technik w praktyce.

Parametry dla ogg vorbis

skala Q (ang. Quality)	przepływność nominalna
-2 (tylko aoTuV b2 i nowszy)	32 Kbps -1 45 Kbps
0	64 Kbps
1	80 Kbps
2	96 Kbps
3	112 Kbps
4	128 Kbps
5	160 Kbps
6	192 Kbps
7	224 Kbps
8	256 Kbps
9	320 Kbps
10	499,9 Kbps

Kompresja obrazów video.

Konieczność kompresji obrazów ruchomych.

Zadanie. Wyznaczyć rozmiar 1 minuty obrazu ruchomego zapisanego w postaci nieskompresowanych obrazów o rozdzielczości 640x480 px w 24-bitowej głębi kolorów. Zakładamy 25 klatek na sekundę.

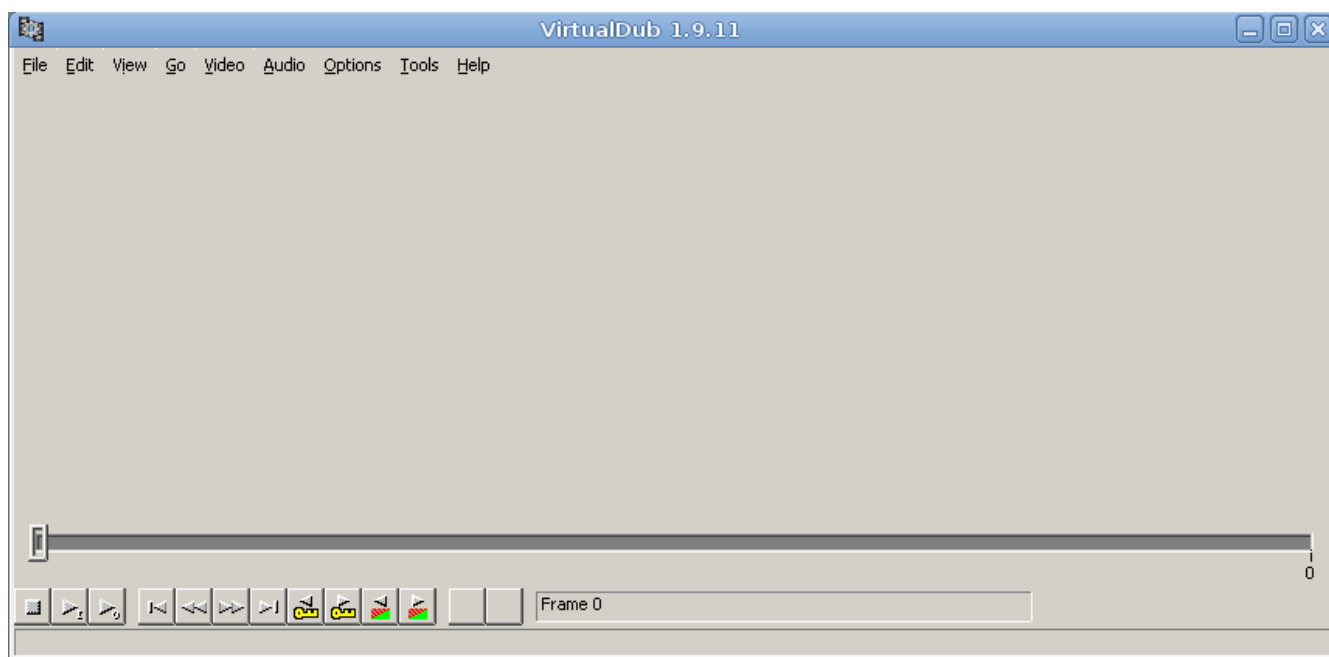
Rozwiązanie tego zadania wymaga umiejętności mnożenia.

Zastanówmy się jaki dysk twardy potrzebny jest na godzinę takiego nagrania ?
Ile miejsca zajmie jedna minuta w HD (1024x768px)

Odpowiedź na te pytania jednoznacznie wskazuje na potrzebę kompresji – zwłaszcza video.

Zapoznać się z programem VirtualDub (<http://www.virtualdub.org/>)

Uruchomić program VirtualDub. Wczytać jeden z pobranych plików (File->Open Video File...). Sprawdzić ile powinno zajmować niekompresowane video w takim pliku. Parametry pliku otrzymuje się w menu File->File Information



Wykorzystując program VirtualDub usunąć z pliku ścieżkę dźwiękową.

W tym celu zaznacza się Audio->No audio oraz Video->DirectStream Copy .. po czym wybiera się opcję File->Save as...

Zapisać nagranie do innego pliku (np. o nazwie orig_nosound.avi).

Znając parametry - rozdzielczość, FPS, zakładając 3 bajty/pixel oszacować rozmiar nieskompresowanego strumienia. Sprawdzić wyjściowy stopień kompresji. Zapisać plik video do postaci nieskompresowanej - porównać rozmiar uzyskanego pliku z danymi z pkt 2. W tym celu w zakładce video zaznacza się Fast recompress oraz ponownie Video->Compression wybiera się Uncompressed RGB.

Stosując różnorodne kodeki i ich ustawienia sprawdzić możliwości zapisu obrazu i zbadać jakość odtwarzania obrazu.

Jak powinien kształtować się stosunek czasu kompresji do dekompresji danych w przypadku konieczności ich składowania ?

Oszacować stosunek informacji audio w sygnale video dla różnych stopni kompresji.

Sprawdzić działanie kodeka DiviX. O ile kodek umożliwia przetestować wykrywanie wektorów ruchu i kompresję progresywną.

Opcjonalnie: Zapoznać się z podobnym do VirtualDub programem Super (zainstalowany w laboratorium) - ocenić możliwości, interfejs użytkownika, działanie tych samych kodeków w różnych narzędziach itp.

W przypadku analizy video nie możliwe będzie oszacowanie miar jakości obrazu - po pierwsze obraz pobrany został w skompresowanej postaci - nie posiadamy zatem oryginału, po drugie - złożoność obliczeniowa takich obliczeń byłaby znaczna.

W sprawozdaniu proszę odpowiedzieć na pytania:

- 1.** W jaki sposób dokonuje się kompresji stratnej sygnału audio ?
- 2.** Jakie zjawisko wykorzystuje się przy kompresji obrazu video ?
- 3.** Jaki jest udział ścieżki dźwiękowej w rozpatrywanych przykładach ?

- 4.** Co to są klatki kluczowe ? Jakie inne klatki zapisuje się w skompresowanym sygnale video ?
- 5.** Dlaczego w kompresji obrazu wykorzystuje się detekcję wektorów ruchu ?

Wyniki obliczeń oraz uzyskanych danych wraz z komentarzem zamieścić w sprawozdaniu, które należy okazać prowadzącemu.