STUDIA INFORMATICA 2000
Volume 21 Number 1 (39)

Agnieszka CHODOREK
Politechnika Świętokrzyska, Katedra Elektroniki i Telekomunikacji

PREDYKTOR SYGNAŁU WIDEO SKOMPRESOWANEGO W STANDARDZIE MPEG ZREALIZOWANY ZA POMOCĄ ROZMYTEGO SYSTEMU WNIOSKUJĄCEGO

Streszczenie. Przebiegi czasowe sygnału wideo kodowanego w standardzie MPEG charakteryzują się znaczną nieregularnością, przy czym trudno jest precyzyjnie określić, jak zmieniać się będą wartości chwilowe parametrów takiego strumienia. W artykule zostanie przedstawiona analiza sygnału wideo MPEG prowadząca do sformułowania opisu w logice rozmytej oraz system wnioskujący typu Takagi-Sugeno realizujący funkcję predyktora sygnału wideo MPEG.

FUZZY PREDICTOR OF THE MPEG-ENCODED VIDEO STREAM

Summary. The MPEG video stream is very bursty, so in many technical problems classical modelling methods fail or are inefficient. In the paper we present an analysis of MPEG video stream and its properties which led to form fuzzy rules. The Takagi-Sugeno fuzzy inference system for prediction of MPEG video traffic is proposed.

1. Wprowadzenie

W ostatnich latach możemy zaobserwować rosnące zainteresowanie sieciowymi aplikacjami multimedialnymi, jak np. *TV Broadcast* czy VoD. Sygnał audio-wideo posiada dużą redundancję, toteż, aby zmniejszyć ilość przesyłanej informacji, jest on poddawany kompresji. Jednym z częściej wykorzystywanych standardów kodowania i kompresji sygnału multimedialnego jest standard MPEG. Sygnał po kompresji charakteryzuje się dużą nieregularnością, co sprawia, że trudno jest go opisać klasycznymi metodami.

W artykule zostanie zaprezentowana analiza sygnału wideo MPEG oraz rozmyty system wnioskujący typu Takagi-Sugeno realizujący predykcję sygnału wideo MPEG.

2. Strumień wideo skompresowany w standardzie MPEG

W chwili obecnej ruch generowany przez aplikacje multimedialne stanowi znaczną część ruchu w sieciach teleinformatycznych [1], a jednym z ważniejszych standardów kodowania i kompresji sygnału multimedialnego jest standard MPEG [9][8][10][22].

2.1. Kompresja sygnału wideo zgodna ze standardem MPEG

Na wejście kodera MPEG jest podawana seria ramek (pojedynczych obrazów nieruchomych), z których każda zawiera dwuwymiarową tablicę pikseli (ang. *picture elements*, *pixels*, *pels*). Podstawę algorytmu kompresji MPEG stanowią dwuwymiarowe (N x N), ortonormalne dyskretne transformaty kosinusowe (ang. *Discrete Cosine Transform*, *DCT*) (1), przy czym wielkość wejściowa transformaty DCT jest kodowana na 9 bitach, natomiast współczynniki transformaty są 12-bitowymi liczbami z zakresu <–2048: +2047>.

$$F(u, v) = \frac{2}{N} C(u)C(v) \sum_{x=0}^{N-1} \sum_{y=0}^{N-1} f(x, y)\cos\frac{(2x+1)u\pi}{2N} \cos\frac{(2y+1)v\pi}{2N} \qquad (1)$$

przy czym: u, v, x, y = 0, 1, 2, ... N – 1

gdzie: x, y - współrzędne w przestrzeni sygnału dyskretnego,

u, v - współrzędne w przestrzeni widma dyskretnego,

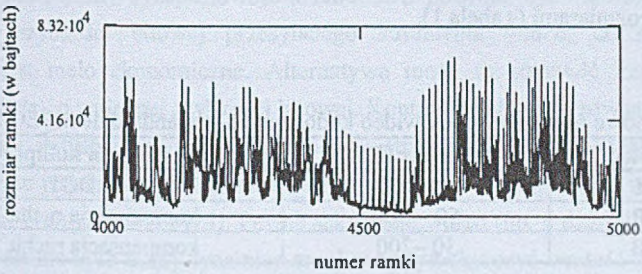
$$C(u), C(v) = \begin{cases} \frac{1}{\sqrt{2}} & \text{dla } u, v = 0 \\ 1 & \text{dla } u, v \neq 0 \end{cases}$$

Oparty na DCT adaptacyjny, hybrydowy algorytm kompresji pozwala usunąć:

- korelację przestrzenną za pomocą dyskretnej transformaty kosinusowej, która operuje na blokach (ang. block) 8 x 8 pikseli (próbek);
- korelację czasową za pomocą kompensacji ruchu operującej na makroblokach (ang. *macroblock*), tj. fragmentach obrazu o rozmiarach 2 x 2 bloki;
- nadmiarowość statystyczną za pomocą kodowania entropii;
- nadmiarowość subiektywną, dzięki kwantyzacji współczynników transformaty.

Każda skompresowana ramka zawiera pojedynczy obraz nieruchomy, który może być reprezentowany przez ramkę należącą do jednego z trzech typów ramek: I, B, P. Ramki intrakodowane – typu I (ang. *Intraframe coding*) – charakteryzują się minimalną kompresją. Nie mają one powiązań z żadną z ramek, co pozwala na prostą realizację swobodnego dostępu do poszczególnych sekwencji wideo. Algorytm kompresji ramek z predykcją ruchu - typu P (ang. *Predictive coding*) – opiera się na DCT i tzw. kompensacji ruchu (wykorzystuje

podobieństwa kolejnych obrazów w sekwencji tworzącej przekaz filmowy), co sprawia, iż ramki typu P są powiązane z wcześniejszymi ramkami w sekwencji wideo. Największym stopniem kompresji charakteryzują się ramki interpolowane - typu B (ang. *Bidirectional coding*) - z kompresją opartą na DCT i dwukierunkowej kompensacji ruchu, wykorzystującą podobieństwo bloków w dwóch obrazach odniesienia: wcześniejszym i późniejszym.



Rys.1. Przykładowe przebiegi czasowe strumienia MPEG
Fig.1. An example of MPEG video traces

Skompresowane ramki są formowane w deterministyczną strukturę, tzw. grupę obrazów (ang. *Group of Pictures* GOP), zdefiniowaną jako uporządkowany ciąg ramek zawarty między dwoma nagłówkami GOP. Standard wymaga, by ciąg ramek występujących po nagłówku GOP rozpoczynał się od ramki typu I. Sama struktura GOP nie została ujęta w standardzie, a w skrajnym przypadku może się zmieniać także podczas transmisji sekwencji obrazów. Przykładowo, GOP może się składać zarówno z samych ramek typu I, jak i z 13 ramek I, P, B. Najczęściej stosowana jest konstrukcja, w której ramki I są przesyłane co 12 (sekwencja: ***I-B-B-P-B-B-P-B-B-P-B-B***) lub 15 ramek (sekwencja: ***I-B-B-P-B-B-P-B-B-P-B-B-P-B-B***). Na rysunku 1 zostały przedstawione przykładowe długości kolejnych ramek filmu *Gwiezdne wojny*[1]. Ramki są zgrupowane w 12-elementową strukturę GOP.

2.2. Własności sygnału wideo MPEG

Przebiegi czasowe sygnału wideo kodowanego w standardzie MPEG są superpozycją trzech różnych przebiegów, odniesionych do trzech typów ramek i powtarzanych okresowo zgodnie z przyjętym wzorcem GOP. Strumień zmultipleksowany, jak również trzy strumienie składowe, charakteryzują się znaczną nieregularnością, zależną w dużej mierze od dynamiki

[1] Materiał filmowy *starwars*. Jest to jedna z sekwencji filmowych zakodowanych w standardzie MPEG-1 przez Olivera Rose. Są one publicznie dostępne na serwerze Uniwersytetu w Wuerzburgu pod adresem: `ftp-info3.informatik.uni-wuerzburg.de/pub`. Każda sekwencja zawiera 40 000 ramek o rozdzielczości obrazu oryginalnego 384x288 pikseli, zgrupowanych w 12-ramkowe struktury GOP.

scen i szczegółowości obrazu (ang. *content*) źródłowego materiału filmowego. Materiał filmowy o znacznej dynamice scen charakteryzuje się na ogół dużymi rozmiarami ramek. Dotyczy to zarówno ramek z zakodowanym całym obrazem (ramki typu I), jak i ramek z kompensacją ruchu (ramki typu P, ramki typu B). Duża szczegółowość obrazu wpływa przede wszystkim na zwiększenie rozmiaru ramek typu I. Generalnie, ramki typu I mają większe rozmiary (w bajtach) niż ramki typu P, natomiast ramki typu B charakteryzują się najmniejszymi rozmiarami (tabela 1).

Tabela 1

Typowe rozmiary ramek wideo kodowanych w standardzie MPEG-2

Typ ramki	Rozmiar (KB)	Metoda kompresji
ramki I	100 – 600	DCT
ramki P	50 – 300	kompensacja ruchu, DCT
ramki B	30 –200	kompensacja ruchu, DCT

Znaczna nieregularność strumienia wideo MPEG w połączeniu z jego silnymi właściwościami pseudookresowymi sprawia, że tego typu strumień trudno jest zamodelować klasycznymi metodami. Charakterystyki statystyczne bazujące na metodach statystyki opisowej (tabela 2) nie umożliwiają dobrego oszacowania wartości chwilowych. Brak jest również wyraźnych związków zarówno między charakterystykami statystycznymi strumienia wynikowego a charakterystykami jego trzech strumieni składowych, jak i między charakterystykami statystycznymi strumieni (wynikowego, bądź któregokolwiek ze strumieni ramek) pochodzących z różnego materiału filmowego (spostrzeżenia te znajdują również potwierdzenie w literaturze - np. [6][13][14][17][25]).

Tabela 2

Charakterystyki statystyki opisowej wyznaczone dla materiału filmowego *starwars*

Charakterystyka statystyczna	GOP	wszystkie ramki	ramka typu I	ramka typu P	ramka typu B
minimum	$1.886*10^4$	272	$1.373*10^4$	344	272
maksimum	$4.495*10^5$	$1.24*10^5$	$1.248*10^5$	$6.764*10^4$	$5.202*10^4$
średnia arytmetyczna	$1.117*10^5$	9313	$4.401*10^4$	$1.013*10^4$	4669
mediana	$9.757*10^4$	4432	$4.342*10^4$	6848	3465
kwartyl dolny	$6.788*10^4$	2392	$3.339*10^4$	3460	2008
kwartyl górny	$1.372*10^5$	9160	$5.29*10^4$	$1.342*10^4$	5840
rozstęp	$4.306*10^5$	$1.24*10^5$	$1.111*10^5$	$6.73*10^4$	$5.174*10^4$
wariancja	$4.156*10^9$	$1.66*10^8$	$1.999*10^8$	$9.11*10^7$	$1.82*10^7$
odchylenie standardowe	$6.447*10^4$	$1.29*10^4$	$1.414*10^4$	9547	4266
odch.przec. od średniej arytmet.	$4.713*10^4$	8431	$1.073*10^4$	7116	2891
odchyl. przeciętne od mediany	$4.557*10^4$	6923	$1.072*10^4$	6575	2702
odchylenie ćwiartkowe	$3.466*10^4$	3384	9756	4982	1916
współczynnik asymetrii	1.688	2.739	1.383	1.745	2.737
współczynnik skupienia	6.961	11.9	8.013	6.172	15.48
współczynnik zmienności	57.69	138.5	32.12	94.25	91.38
współcz. nierównomierności	42.17	90.52	24.37	70.25	61.93

2.3. Transmisja sygnału wideo skompresowanego w standardzie MPEG

Skompresowany sygnał wideo w standardzie MPEG charakteryzuje się m.in. losowo zmienną w czasie szybkością bitową, przy czym trudno jest precyzyjnie określić, jak zmieniać się będą wartości chwilowe parametrów takiego strumienia. Aby przeprowadzić transmisję strumienia wideo MPEG, można zastosować kanał o stałej przepustowości (ang. *Constant Bit Rate, CBR*). Przepustowość kanału CBR powinna być określona na poziomie maksymalnej szybkości bitowej przesyłanego strumienia wideo, co sprawia, że to rozwiązanie jest mało ekonomiczne. Alternatywę może tu stanowić kanał VBR (ang. *Variable Bit Rate*), o zmiennej szybkości bitowej. Kontrakt VBR precyzowany jest w oparciu o parametry statystyczne transmitowanego strumienia. Tego typu rozwiązanie wykorzystywane jest często do transmisji wideo w czasie rzeczywistym w sieciach ATM (*Asynchronous Transfer Mode*) [24]. Pewne sekwencje wideo (np. o dużej ilości szczegółów kodowanego obrazu i/lub dużej dynamice scen) mogą przekroczyć wartość kontraktu. Część komórek ATM zostaje wówczas odrzucona przez komutator, powodując przypadkowe uszkodzenia ramek wideo. W wielu przypadkach ma to znaczący wpływ na pogorszenie jakości transmitowanego obrazu.

Podjęcie decyzji o przyjęciu lub odrzuceniu zgłoszenia następuje na podstawie założonego modelu ruchowego źródła, opartego np. na teorii masowej obsługi, typowych procesach stochastycznych (Poissona, Gaussa i inne) czy też metodach analitycznych. Ze względu na wymaganą dużą szybkość działania model ruchowy źródła powinien charakteryzować się możliwie małą ilością parametrów i szybkim algorytmem przetwarzania. Ważne jest również, by był on łatwo realizowalny w postaci układów VLSI.

3. Rozmyte systemy wnioskujące

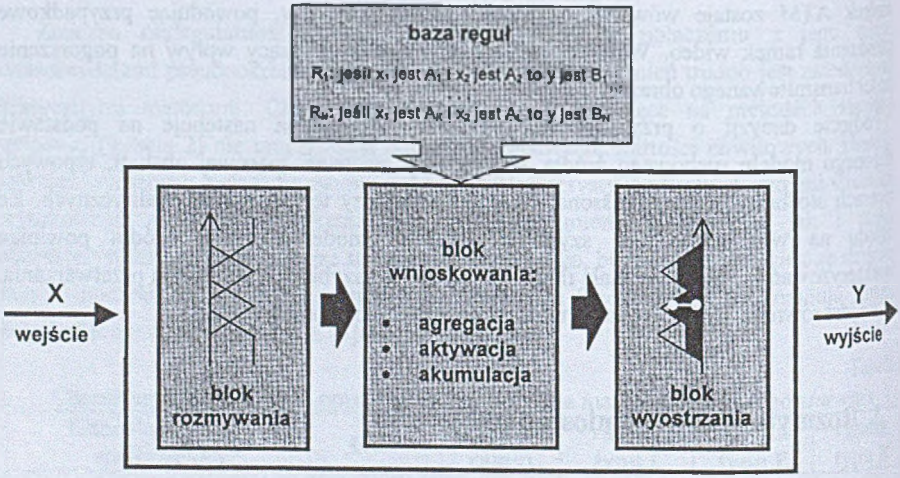
Teoria zbiorów rozmytych, sformułowana w 1965 roku przez L.A. Zadeha, znalazła zastosowanie w wielu dziedzinach nauki i techniki jako narzędzie do modelowania złożonych systemów oraz procesów podejmowania decyzji. Rozmyte systemy wnioskujące (ang. *fuzzy inference system, FIS*) [11][18][19][21] pozwalają na osiągnięcie zadowalających rezultatów w warunkach braku precyzji lub niejednoznaczności danych wejściowych. Często bywają one skuteczne tam, gdzie zawodzą klasyczne metody przetwarzania informacji – np. metody statystyczne. Systemy tego typu znajdują zastosowanie w wielu dziedzinach nauki i techniki, w tym również w sterowaniu i predykcji ruchu sieci teleinformatycznych [2].

3.1. Struktura i elementy funkcjonalne rozmytego systemu wnioskującego

Rozmyty system wnioskujący (rys.2.) składa się z czterech bloków funkcjonalnych [7][12][18][19]: bloku rozmywania bazy reguł, bloku wnioskowania i bloku wyostrzania.

Blok rozmywania realizuje operację fuzyfikacji (ang. *fuzzyfication*), tj. odwzorowania wielkości wejściowej (ostrej) w zbiór rozmyty A', podawany na wejście bloku wnioskowania. Odwzorowanie to polega na oszacowaniu stopnia zgodności zmiennej wejściowej systemu wnioskującego z wartością zmiennej lingwistycznej.

Baza reguł (ang. *rule base*) nazywana jest również modelem lingwistycznym (ang. *linguistic model*) i zawiera empiryczną wiedzę o procesie będącym przedmiotem wnioskowania. Wiedza ta jest reprezentowana poprzez zbiór reguł rozmytych w postaci (2). W literaturze przedmiotu z bazy reguł wyodrębnia się czasami bazę danych (słownik) zawierającą opisy funkcji przynależności zbiorów rozmytych występujących w poprzednikach i następnikach reguł R_i [4].



Rys. 2. Struktura i elementy funkcjonalne rozmytego systemu wnioskującego
Fig. 2. Structure and functional elements of a fuzzy inference system

Blok wnioskowania realizuje mechanizm uogólnionego wnioskowania typu *modus ponens*, dla którego przesłanki stanowią dane wejściowe (zazwyczaj poddane procesowi fuzyfikacji), natomiast implikację – reguły rozmyte zawarte w bazie reguł. Proces wnioskowania przebiega w trzech etapach [7]: agregacji reguł elementarnych w regułę złożoną; aktywacji reguły (skojarzenia zmiennej wyjściowej z odpowiednią regułą rozmytą); akumulacji wyników w jeden zbiór rozmyty.

Na wyjściu bloku wnioskowania otrzymujemy zbiór rozmyty B', jednakże w wielu zastosowaniach wymagane jest, by na wyjściu systemu wnioskującego uzyskać ostrą wielkość wyjściową. Zbiór rozmyty B' należy wówczas poddać operacji wyostrzania

(defuzyfikacji – ang. *defuzzyfication*). Realizowana jest ona w bloku wyostrzania i ma na celu wyznaczenie takiej wielkości wyjściowej (ostrej) systemu wnioskującego, która w jak najlepszy sposób odda charakter wyjściowego zbioru rozmytego.

3.2. Systemy typu Mamdaniego i systemy typu Sugeno

Rozmyte systemy wnioskujące mają dość szczegółowo zdefiniowany mechanizm wnioskowania, nie ma jednakże ścisłych wytycznych co do sposobu implementacji (w szczególności niektóre spośród bloków funkcjonalnych mogą być, w całości lub częściowo, realizowane za pomocą sieci neuronowej) ani co do typu funkcji przynależności zbiorów rozmytych występujących w przesłankach i konkluzjach reguł. Tym niemniej, w literaturze przedmiotu wyróżnia się kilka typów rozmytych systemów wnioskujących, z których najczęściej stosowanymi są: system typu Mamdaniego oraz system typu Sugeno.

Pierwsze zastosowania rozmytych systemów wnioskujących pojawiły się w połowie lat siedemdziesiątych. Były to prace zespołu pod kierownictwem E.H. Mamdaniego. Systemy te, nazwane później systemami Mamdaniego, nie narzucają żadnych ograniczeń na system rozmyty – w tym również ograniczeń na funkcje przynależności zbiorów rozmytych. Wnioskowanie w systemach typu Mamdaniego jest zgodne z (2). W przesłankach i konkluzjach reguł występują zbiory rozmyte. Systemy typu Mamdaniego są konstruowane na podstawie przyczynowo-skutkowej zależności miedzy wejściami a wyjściami systemu. Nie zakłada się znajomości innych,niż zbiór danych typu wejścia/wyjście, opisów systemu.

$$\text{jeśli } (x \text{ jest } A) \text{ to } (y \text{ jest } B) \tag{2}$$

Zastosowanie technik bazujących na logice rozmytej pozwoliło uzyskać wyniki, jakich nie dało się osiągnąć klasycznymi metodami. Dotyczyło to zwłaszcza zagadnień modelowania i sterowania nieliniowego. Jednak rewolucyjna zmiana, jaką niosło ze sobą nowe podejście, wiązała się z brakiem możliwości wykorzystania całej dotychczasowej wiedzy o obiekcie. Tymczasem na przestrzeni lat (a w wielu przypadkach nawet wieków) powstało wiele lokalnie poprawnych modeli liniowych, dających dobre przybliżenie zjawisk nieliniowych. Modele te można wykorzystać stosując rozmyty system wnioskujący typu Sugeno[2] [23], który powstał w dziesięć lat po systemie Mamdaniego. W systemie Sugeno w przesłankach reguł występują zbiory rozmyte, natomiast we wnioskach – zależności funkcyjne (3), dobierane najczęściej na podstawie teoretycznej wiedzy o systemie.

$$\text{jeśli } (x \text{ jest } A) \text{ to } (y=f(x)) \tag{3}$$

Funkcje przynależności zbiorów rozmytych i funkcje występujące we wnioskach mogą być dowolne. We wnioskach stosuje się najczęściej funkcje stałe (i wówczas system typu Sugeno jest równoważny systemowi Mamdaniego) lub funkcje liniowe.

---

[2] stosowane są również nazwy: system Takagi-Sugeno i Takagi-Sugeno-Kanga

4. Rozmyty predyktor sygnału wideo skompresowanego w standardzie MPEG

Jak wspomniano w rozdziale 2.2, pomimo silnych właściwości pseudookresowych wykresy czasowe strumienia wideo MPEG charakteryzują się znaczną nieregularnością. Z tego też względu próby sformułowania opisu - matematycznego bądź lingwistycznego - strumienia MPEG napotykały na duże trudności, wynikające głównie z braku wyraźnych powiązań pomiędzy ramkami nie należącymi do tego samego GOP. Rozwiązaniem tego problemu może być m.in. zastosowanie metod sztucznej inteligencji - w tym metod opartych na logice rozmytej.

4.1. Analiza sygnału wideo MPEG pod kątem możliwości opisu w logice rozmytej

W ogólnym przypadku rozmyty system wnioskujący jest obiektem o n wejściach i m wyjściach, realizującym wnioskowanie rozmyte. Określenie ilości wejść i wyjść ma wpływ zarówno na dokładność systemu wnioskującego (im większa ilość wejść, tym system - zazwyczaj - jest dokładniejszy), jak i na szybkość działania (im mniejsza ilość wejść i wyjść, tym system jest szybszy).

4.1.1. Pozyskiwanie informacji z rozmytego systemu wnioskującego

W przypadku transmisji wideo w sieci ATM ilość komórek potrzebnych do przetransmitowania pojedynczej ramki sygnału wideo skompresowanego w standardzie MPEG waha się od kilkuset do kilkunastu tysięcy (tabela 3). Aplikacja multimedialna czasu rzeczywistego narzuca ścisłe ograniczenia czasowe: w zależności od systemu (PAL, NTSC) ramki są odtwarzane w urządzeniu końcowym z częstotliwością 24-30 ramek na sekundę. W przypadku przesyłania ramek z prędkością 25 ramek/s pojedyncza ramka musi zostać przetransmitowana w czasie nie dłuższym niż 40 milisekund. Stanowi to wystarczający horyzont czasowy dla systemów zarządzających sieciami.

Dla sprawnego zarządzania siecią wystarczy zatem, aby system zarządzający mógł dokonać predykcji rozmiaru następnej transmitowanej ramki wideo.

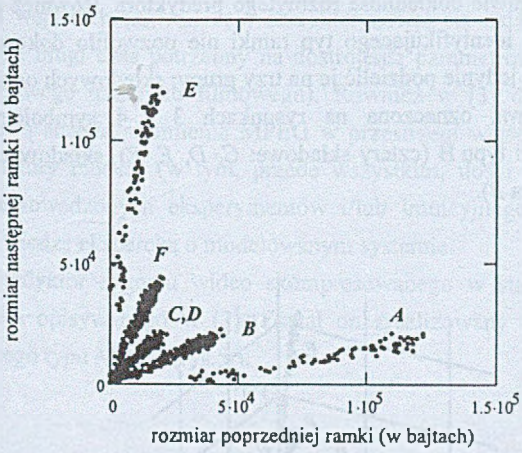
Tabela 3

Ilość komórek ATM potrzebna do przetransmitowania ramki wideo MPEG

Ramka	Typowy rozmiar (KB)	Orientacyjna ilość komórek ATM
ramki I	100 – 600	2 100 - 12 800
ramki P	50 – 300	1 000 - 6 400
ramki B	30 –200	600 - 1 000

4.1.2. Rozmyty wielomodel jako przybliżenie sygnału wideo MPEG

Zgodnie z analizą strumienia MPEG przedstawioną w rozdziale 2.2. długości kolejnych ramek strumienia w znacznej mierze zależą od dynamiki i szczegółowości scen zakodowanej sekwencji wideo. Zazwyczaj zaraz po "dużych" ramkach typu I występują "duże" ramki typu B, a po nich "duże" ramki typu P (itd.). Stąd też, analizując strumień MPEG pod kątem określenia minimalnej ilości wejść niezbędnych do prawidłowego opisu, można przyjąć wstępne założenie, że strumień ten nie posiada własności pamięci, a co za tym idzie - dla danej chwili czasu t przewidywany rozmiar następnej ramki możemy uzależnić tylko od rozmiaru ramki bieżącej. Proponowany rozmyty system wnioskujący będzie zatem obiektem typu SISO.



Rys. 3. Charakterystyka strumienia wideo MPEG w dwuwymiarowej przestrzeni wejść i wyjść (model SISO). Charakterystykę wyznaczono dla pierwszych 70 sekund filmu *Gwiezdne wojny* (*starwars*)

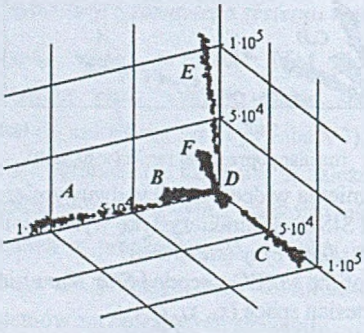
Fig. 3. The first 70 minutes of the MPEG-encoded *Star Wars* video traces into the two-dimensional Cartesian space (x_k, x_{k+1})

Jednakże, jak wynika z rysunku 3, charakterystyka strumienia nie jest jednoznaczna względem wejścia x. Stan wyjścia nie może być określony jednoznacznie - w najgorszym przypadku jednej wartości x mogą odpowiadać nawet 3 wartości wyjścia y. Przedstawiona na rys.3. charakterystyka strumienia MPEG wskazuje na to, iż mamy do czynienia z wielomodelem[3], złożonym z (co najmniej) pięciu modeli rozmytych. Rozmyty system wnioskujący należy zatem uzupełnić o mechanizm przełączania modeli składowych,

[3] Pojęcie wielomodelu rozmytego wprowadził w 1996 roku W. Pedrycz [15]. Wielomodel jest to zbiór modeli rozmytych wraz z mechanizmem przełączania pomiędzy tymi modelami.

działający na podstawie dodatkowych informacji o systemie [15][16]. Informacje te muszą zostać wprowadzone na dodatkowe wejścia systemu wnioskującego.

Nasuwające się rozwiązanie zastosowania systemu dwuwejściowego (x_k, x_{k-1}), w którym na wejście podawane są wartości ramek: bieżącej i poprzedniej, nie rozwiązuje jednak problemu. Charakterystyka strumienia wideo MPEG w trójwymiarowej przestrzeni kartezjańskiej (x_k, x_{k-1}, x_{k+1}) wykazuje bowiem istnienie szóstego modelu składowego (rys.4.), który nie jest widoczny w rzucie na płaszczyznę (x_k, $y=x_{k+1}$). Model ten wprowadza nową niejednoznaczność wyjścia względem pary wejść (x_k, x_{k-1}). Zjawiska tego nie daje się wyeliminować poprzez podawanie na wejście przełączające sygnału (x_{k-2}) - uzyskuje się wówczas tylko zmianę położenia szóstego składnika wielomodelu w przestrzeni wejść i wyjść, chociaż (zgodnie z [20]) przesunięcie w czasie sygnału przełączającego może, w pewnym stopniu, poprawić dokładność rozmytego predyktora. Również podanie na wejście przełączające sygnału identyfikującego typ ramki nie pozwoliło dokonać pełnej separacji składowych modelu, a jedynie podzielić je na trzy grupy: składowych odniesionych do ramki typu I (jedna składowa, oznaczona na rysunkach 3 i 4 symbolem *A*), składowych odniesionych do ramki typu B (cztery składowe: *C, D, E, F*), składowych odniesionych do ramki typu P (składowa *B*).



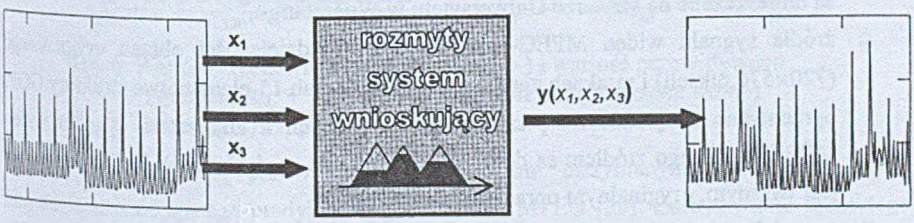
Rys. 4. Charakterystyka strumienia wideo MPEG w trójwymiarowej przestrzeni wejść i wyjść (model MISO). Charakterystykę wyznaczono dla pierwszych 70 sekund filmu *Gwiezdne wojny* (*starwars*)

Fig. 4. The first 70 minutes of the MPEG-encoded *Star Wars* video traces into the two-dimensional Cartesian space (x_k, x_{k-1}, x_{k+1})

4.2. Struktura rozmytego systemu wnioskującego realizującego predykcję sygnału wideo kodowanego w standardzie MPEG

Zagadnienie modelowania - w tym również modelowania rozmytego - strumienia wideo kodowanego w standardzie MPEG było wielokrotnie poruszane w literaturze. Opracowane modele rozmyte są zwykle modelami dwuwejściowymi, chociaż można spotkać również modele 12-wejściowe (najdokładniejsze, lecz równocześnie najmniej podatne na dostrajanie parametrów i, jak zostanie za chwilę wykazane, obarczone bardzo dużą redundancją). Do ciekawszych rozwiązań można zaliczyć [5], gdzie wprowadzono dodatkową, ostrą[4] informację o typie ramki (I, B, P). W [20] zaproponowano z kolei modyfikację modelu dwuwejściowego (x_k, x_{k-1}) uzależniającego wyjście od ramki bieżącej i poprzedniej, w którym na wejście podawano wartości ramek: bieżącej i przedostatniej (x_k, x_{k-2}). Zaproponowany w [3] model o trzech wejściach (x_k, x_{k-1} x_{k-2}) charakteryzował się największą dokładnością, jednak jego wadą był długi czas potrzebny na dostrojenie parametrów modelu do nowych warunków pracy (nowego materiału filmowego). Również w [3] dobór struktury został poprzedzony skrótową analizą strumienia MPEG w przestrzeni wejść i wyjść. Najczęściej bowiem wybór struktury modelu (w tym, przede wszystkim, ilości wejść i wyjść) bywa efektem wielu przeprowadzonych eksperymentów i/lub intuicyjnego podejścia twórców, posiadających dużą wiedzę ekspercką o modelowanym systemie.

Proponowany predyktor sygnału wideo skompresowanego w standardzie MPEG jest rozwinięciem modelu opisywanego w [3]. Został on zrealizowany w postaci rozmytego systemu wnioskującego typu MIMO (rys.5.).



Rys. 5. Rozmyty system wnioskujący realizujący predykcję strumienia wideo MPEG

Fig. 5. The fuzzy inference system for prediction of the MPEG video stream

Wektor wejściowy systemu posiada trzy składowe: x_1 ... x_3, gdzie: x_1 jest zmienną wejściową odniesioną do zmiennej lingwistycznej "bieżąca ramka" i reprezentuje rozmiar

[4] nie rozmytą.

bieżącej ramki, podczas gdy zmienne podawane na wejścia przełączające, tj. x_2 ("poprzednia ramka") i x_3 ("następna ramka") reprezentują typ ramki (I, B, P). Funkcje przynależności zbiorów rozmytych odniesionych do zmiennej x_1 zostały przybliżone funkcją Gaussa. Funkcje przynależności zbiorów rozmytych odniesionych do zmiennych x_2, x_3 zostały przybliżone funkcją trójkątną.

Wielkość wyjściowa y reprezentuje przewidywany rozmiar następnej ramki. Podobnie, jak miało to miejsce w [3], wyjściowe funkcje przynależności są liniowo zależne od wektora wejściowego (rozmyty system wnioskujący typu Takagi-Sugeno). Decyzja o zastosowaniu rozmytego systemu wnioskującego typu Takagi-Sugeno została podjęta ze względu na pseudoliniowy charakter składowych modeli rozmytych (por. rys. 3 i 4). Z kolei ze względu na możliwość łatwej dekompozycji wielomodelu na modele składowe (obserwowany wielomodel jest efektem grupowania różnych typów ramek w struktury GOP) wielkość wyjściowa y została uzależniona tylko od zmiennej wejściowej x_1. Zmienne wejściowe x_2, x_3 dostarczają informacji o bieżącym stanie systemu. Informacje te są wykorzystywane przez mechanizm przełączania modeli składowych wielomodelu rozmytego. Parametry funkcji przynależności wejściowych zbiorów rozmytych zostały oszacowane metodą *fuzzy c-means*.

Rozmyty predyktor został zrealizowany w postaci rozmytej sieci neuronowej (ang. *neuro-fuzzy network*). System wnioskujący został następnie dostrojony zgodnie z algorytmem *ANFIS*.

W ramach prowadzonych prac sieć rozmyta była dostrajana przy użyciu ruchu generowanego przez źródła sygnału wideo zgodnego ze standardem MPEG (Tabela 4):

- źródła sygnału wideo MPEG-1, o niskiej rozdzielczości obrazu oryginalnego (384x288 pikseli) i ramkach zgrupowanych w 12-elementowe struktury GOP; dane te są umieszczone na serwerze Uniwersytetu w Wuerzburgu,
- źródła sygnału wideo MPEG-2, o wysokiej rozdzielczości obrazu oryginalnego (720x576 pikseli) i ramkach zgrupowanych w 12- lub 15-elementowe struktury GOP; opracowano na podstawie przebiegów otrzymanych z analizatora sygnału wideo MPEG-2, którego źródłem są dane przechowywane na dyskach DVD (analizator ten jest własnym, oryginalnym opracowaniem autorki).

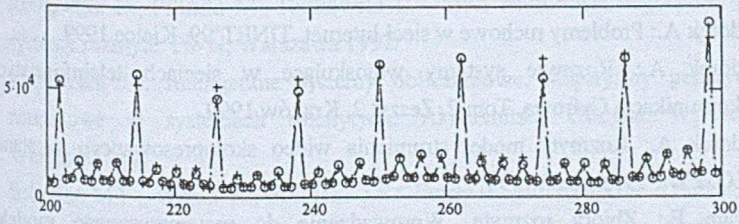
W przypadku zaistnienia sytuacji, która występowała w zbiorze uczącym, różnica pomiędzy wartością przewidywaną przez system a wartością rzeczywistą była przeważnie bardzo mała (najlepszy przypadek: 0,03%, najgorszy przypadek: 3,85%, średnio 1,31%). W przypadku sekwencji wideo, która nie brała udziału w dostrajaniu sieci, błędy nie przekraczały 10%. Największe błędy bezwzględne zaobserwowano podczas predykcji ramek typu I.

Tabela 4

Źródła sygnału wideo zgodnego ze standardem MPEG

nazwa materiału filmowego	tytuł oryginalny	standard kodowania	rozdzielczość	rozmiar GOP
armageddon	Armageddon	MPEG-2	720x576	15
avengers	The Avengers	MPEG-2	720x576	15
bond	Goldfinger	MPEG-1	384x288	12
dino	Jurassic Park	MPEG-1	384x288	12
lambs	The Silence of the Lambs	MPEG-1	384x288	12
microcosmos	Microcosmos: Le peuple de l'herbe	MPEG-2	720x576	12
postman	The Postman	MPEG-2	720x576	15
rejs	Rejs	MPEG-2	720x576	12
starwars	Star Wars	MPEG-1	384x288	12
terminator	The Terminator	MPEG-1	384x288	12

Przykładowa odpowiedź rozmytego predyktora sygnału wideo MPEG została zamieszczona na rys. 6. Jak widać na rysunku, system jest w miarę dokładny, chociaż przejawia tendencję do uśredniania rozmiarów ramek (szczególnie jest to widoczne w przypadku ramek typu I).



Rys. 6. Wartość rzeczywista (symbol *okrąg:* o) i wartość przewidywana (symbol *plus:* +) strumienia wideo MPEG (*Gwiezdne wojny*)

Fig. 6. The real (o's) vs. predicted (+'s) value of the frame size (*Star Wars*)

Uzyskane wyniki są porównywalne z wynikami otrzymanymi dla systemu typu Mamdaniego realizującego predykcję sygnału wideo MPEG [20]. Zastosowanie rozmytego systemu wnioskującego typu Sugeno pozwoliło wykorzystać wiedzę o modelowanym obiekcie (strumieniu wideo MPEG - por r. 4.1.), co przyspieszyło znacznie proces uczenia rozmytej sieci neuronowej, a co za tym idzie - skróciło czas dostrajania predyktora.

5. Zakończenie

Efektywne zarządzanie sieciami teleinformatycznymi wymaga znajomości parametrów ruchu oferowanego. Współczesne aplikacje multimedialne generują strumień danych o złożonej charakterystyce, stąd też opis takiego strumienia trudno jest zrealizować metodami klasycznymi.

W artykule przedstawiono predyktor sygnału wideo MPEG zrealizowany za pomocą logiki rozmytej. Dokonano analizy strumienia wideo kodowanego w standardzie MPEG w przestrzeni wejść i wyjść modelu rozmytego, po czym wprowadzono pojęcie rozmytego wielomodelu strumienia wideo MPEG. Zaproponowana struktura predyktora wideo MPEG została zrealizowana w oparciu o rozmyty system wnioskujący typu Takagi-Sugeno. Zamieszczone wyniki świadczą o tym, że opisywany predyktor potrafi stosunkowo dobrze wyznaczać wielkości kolejnych transmitowanych ramek.

Wkomponowanie stworzonego predyktora w systemy zarządzania sieciami pozwoli na efektywne zarządzanie transmisją multimedialną w sieciach teleinformatycznych.

LITERATURA

1. Chodorek A.: Problemy ruchowe w sieci Internet. TINET'99, Kielce 1999.
2. Chodorek A.: Rozmyte systemy wnioskujące w sieciach teleinformatycznych. Telekomunikacja Cyfrowa, Tom 2, Zeszyt 2, Kraków 1999.
3. Chodorek A.: Rozmyty model strumienia wideo skompresowanego w standardzie MPEG. PWT'99, Poznań 1999.
4. Czogała E.: Zbiory rozmyte. Wprowadzenie do matematycznego modelowania niejednoznaczności. Strona WWW: http://alfa.iele.polsl.gliwice.pl/kalejdoskop/1997/czogala/ wyklad1.html.
5. Genco A., Lombardo G.: MPEG Video Frame Discard by Fuzzy Control. Global Telecommunications Conference GLOBECOM 1998. vol. 2, pp. 690 -695.
6. Heyman D. P., Tabatabai A., Lakshman T. V.:Statistical Analysis of MPEG-2 Coded VBR VideoTraffic. 6th Int. Workshop on Packet Video, September 1994.
7. IEC 1131: Programmable Controllers. Part 7 – Fuzzy Control Programming. Committee Draft CD 1.0 (Rel. 19 Jan. 1997).
8. ISO/IEC 11172-2:1993 Information technology - Coding of moving pictures and associated audio for digital storage media at up to about 1,5 Mbit/s - Part 2: Video.
9. ISO/IEC 13818/H.222: Information Technology - Generic Coding of Moving Pictures and Associated Audio Information - Part 2: Video.

10. ITU-T H.262: Information technology – generic coding of moving pictures and associated audio information: video. 1995.
11. Kacprzyk J.: Zbiory rozmyte w analizie systemowej. PWN, Warszawa 1986.
12. Kosko B.: Neural networks and fuzzy systems. A dynamical systems approach to machine inteligence. Englewood Cliffs: Prentice-Hall International, Inc., 1992.
13. Mellaney E., Orozco-Barbosa L., Gagnon G.: Study of MPEG-2 video traffic in a multimedia LAN/ATM internetwork system, IEEE Trans. Circuit Syst. Video Technol., vol. CSVT-7, pp. 663 - 674, August 1997.
14. Natkaniec M., Wajda K.: Statistical properties of MPEG streams and issues for transmission of video information in high speed networks. ATM Technology Users Symposium, February 1999, pp. 145-150.
15. Pedrycz W.: Fuzzy multimodels. IEEE Transactions on Fuzzy Systems. vol.4., No.2, 1996, pp. 139-148.
16. Piegat A.: Modelowanie i sterowanie rozmyte. Akademicka Oficyna Wydawnicza EXIT, Warszawa 1999.
17. Rose O.: Statistical properties of MPEG video traffic and their impact on traffic modeling in ATM systems. University of Wuerzburg Institute of Computer Science Research Report Series, Report No. 101, Feb. 1995, pp. 1-25.
18. Rutkowska D., Piliński M., Rutkowski L.: Sieci neuronowe, algorytmy genetyczne i systemy rozmyte. PWN, Warszawa 1997.
19. Rutkowska D.: Inteligentne systemy obliczeniowe. Algorytmy genetyczne i sieci neuronowe w systemach rozmytych. Akademicka Oficyna Wydawnicza PLJ, Warszawa 1997.
20. Scheffer M.F., Wajda K., Kunicki J.S.: Fuzzy Logic Adaptive Traffic Enforcement Mechanisms for ATM Networks. Proc. of International Teletraffic Conference, Pretoria, South Africa 1995.
21. Stacewicz P.: Zastosowania sieci neuronowych do wnioskowań rozmytych. Prace IPI PAN nr 782, Wyd. Instytutu Podstaw Informatyki Polskiej Akademii Nauk, Warszawa 1995.
22. Steinbrink Bernd, Multimedia u progu XXI wieku. Wyd. Robomatics, Wrocław 1993.
23. Takagi T., Sugeno M.: Fuzzy Identyfication of System and Its Applications to Modeling and Control. IEEE Transactions on Systems, Man and Cybernetics, vol. 15, January/February 1985.
24. Wajda K.: Sieci szerokopasmowe. Wydawnictwo Fundacji Postępu Telekomunikacji, Kraków 1994.
25. Wajda K., Pacyna P.: Statystyczne właściwości strumienia MPEG i ich wpływ na realizację usług multimedialnych. KST'97, Bydgoszcz 1997.

Recenzent: Dr inż. Ryszard Maceluch

Wpłynęło do Redakcji 7 kwietnia 2000 r.

Abstract

There is growing interest in telecommunication applications using video information, as for example DVB or VoD systems. Most of them use MPEG standard for compression of video signal. The MPEG video traffic stream is very burst, so in many technical problems conventional prediction approaches fail or are inefficient.

MPEG video traces are superposition of three various traces, corresponding with three types of frames and repeated by the GOP pattern. The multiplexed stream, and three component streams, presents a strong pseudo-periodic property. As reported in the literature (e.g. [14][17]), it is difficult to consider any correlation aspects of MPEG streams, when methods of analysis based on descriptive statistics are used.

However, if we insert MPEG traces into the three-dimensional space (Fig. 4) we observe the pseudo-linear character of the graph. Therefore the size of the next frame can be evaluated according to one of "lines" depicted on Fig. 1.

Described phenomena led to define a structure of fuzzy model of the MPEG encoded video stream (Fig. 5). Due to pseudo-linear dependence of the output quantity on inputs, we decide to apply a Takagi-Sugeno model of fuzzy inference system.

The input vector consists of three components x_1 ... x_3, where: x_1 is corresponding to linguistic variable "current frame" and represents the size of the current frame, whereas x_2 ("previous frame") and x_3 ("next frame") represent the frame type (I, B, P). Fuzzy sets corresponding to x_1 are defined by Gaussian membership functions and corresponding to x_2, x_3 are defined by triangular membership functions. The output quantity y represents the predicted size of the next frame. Output membership functions are the linear dependence of the input variable x_1. Fuzzy rules are defined according to properties of MPEG video stream.