



WYKORZYSTANIE DESKRYPTORÓW MPEG-7 W KODOWANIU PRZEJŚCIA EDYCYJNEGO W KODERZE WIZYJNYM

Streszczenie: Aktualnie poddawane standaryzacji algorytmy kodowania sekwencji wizyjnych są bardzo zaawansowane. Wprowadzenie modyfikacji jest bardzo trudne i charakteryzuje się niewielkim wzrostem efektywności kodowania. Dlatego też z dużym zainteresowaniem należy spojrzeć na rozwiązania wykorzystujące informacje towarzyszące kodowanej sekwencji wizyjnej. Artykuł przedstawia propozycje zwiększenia efektywności kodowania poprzez wykorzystanie metadanych MPEG-7, opisujących kodowaną sekwencję wizyjną, w algorytmie sterowania pracą koderza wizyjnego.

1. WPROWADZENIE

Standardy kodowania sekwencji wizyjnych MPEG-2 i MPEG-4 odniosły duży sukces, będąc szeroko stosowanymi w cyfrowej telewizji, transmisji strumieniowej i w wielu usługach interaktywnych. Aktualnie dokonuje się rewolucja wprowadzona nowym standardem H.264/MPEG-4 AVC/MPEG-4 część 10. Standard H.264/AVC oferuje znaczne poprawienie efektywności kodowania w porównaniu do starszych standardów kodowania takich jak MPEG-2.

Można oczekiwać, że w przyszłości bardzo duża liczba plików audiowizualnych będzie zindeksowana i metadane będzie można łatwo wyznaczyć. W rezultacie materiał audiowizualny będzie dostępny łącznie z metadanymi opisującymi jego zawartość. Standard MPEG-7 [1-3] dostarcza narzędzi służących temu opisiowi. W tym kontekście każdy typ materiału audiowizualnego jest charakteryzowany przez zestaw wyróżniających ją cech. Istnieje duża potrzeba efektywnego opisu treści multimedialnej, ponieważ wykorzystując metadane można wykorzystać je i zoptymalizować strategię kodowania w koderze wizyjnym [10].

Jednym z charakterystycznych elementów treści wizyjnej jest przejście edycyjne. Przejścia są powszechnie używane w celu połączenia dwóch różnych materiałów wizyjnych w tym w szczególności reklam, które cechują się innymi właściwościami niż główna przekazywana w sekwencji wizyjnej treść. Przejścia nie są naturalną częścią sekwencji wizyjnej a są one wprowadzane na etapie montażu. Biorąc pod uwagę te fakty można zrozumieć, że koder nie zawsze może dobrze przewidzieć wszystkie elementy obrazu i w wyniku jakość kodowanej sekwencji wizyjnej jest gorsza.

Pomimo stosowania bardzo zaawansowanych technik kodowania, wydaje się być słuszne wykorzystanie dodatkowych informacji towarzyszących sekwencji wizyjnej np. strumienia deskryptorów standardu MPEG-7

do zwiększenia efektywności kodowania materiału wizyjnego.

Przejście edycyjne typu nachodzącego (ang. wipe) [5-7] jest specjalnym efektem, w którym dwa obrazy z różnych źródeł są wyświetlane w jednym kadrze. Przejście to jest trudne do zakodowania, ponieważ zawartość makrobloków i wektory ruchu w makroblokach zmieniają się a nowa treść obrazu pojawia się nagle na krawędzi przejścia. Koder musi zmieniać wiele parametrów takich jak parametr kwantyzacji, strukturę makrobloków, typ stosowanej predykcji w makroblokach itd.

Opis w postaci metadanych może pomóc w wyborze decyzji w koderze i podnieść jego efektywność. W pracy przedstawiony jest algorytm wykorzystania metadanych do opisu przejścia edycyjnego, które następnie można wykorzystać na etapie kodowania sekwencji.

2. DETEKcja PRZEJŚCIA

Podczas trwania przejścia edycyjnego, każda ramka obrazu zawiera fragment starej sceny i fragment nowej. Dla przejścia zdefiniowanego w kierunku poziomym, przejście cechuje się pionowym paskiem obrazu, który rozszerza się w kierunku poziomym. Dla przejścia zdefiniowanego w kierunku pionowym, pasek poziomy rozszerza się w kierunku pionowym. Podczas występowania przejścia liczba zmian wartości punktów występujących w pasku jest znacząco większa niż w całym pozostałym obszarze. Lokalizacja zmieniających się punktów może być rejestrowana a następnie wykonywana jest analiza rozkładu przestrzennego zmian.

Aktywność ruchowa, która określa stopień aktywności punktów w sekwencji wizyjnej została zdefiniowana jako deskryptor w standardzie MPEG-7. Standard nie określa sposobu wyznaczania deskryptora a jedynie składnię i semantykę strumienia danych MPEG-7.

W proponowanym rozwiązaniu, sposób wyznaczania deskryptora aktywności ruchowej MAD oparty jest na automatycznym generowaniu parametru aktywności ruchowej [8]. Jednakże autor zaadaptował tę metodę do detekcji przejścia edycyjnego, uzyskując dobrą efektywność poprzez rozszerzenie liczby wektorów, które są sumami wartości w kolumnach – SK i sumami wartości w wierszach – SW.

Macierz TAR (Rys. 2) reprezentuje zakumulowane różnice pomiędzy dwoma kolejnymi ramkami obrazu.

Poniższe równania definiują w jaki sposób obliczane są wartości wektorów SK i SW:

- wektor SK1:

$$SK_1 = \left(\sum_{i=0}^{\frac{N-1}{2}} a_{i1}, \sum_{i=0}^{\frac{N-1}{2}} a_{i2}, \dots, \sum_{i=0}^{\frac{N-1}{2}} a_{iM} \right) \quad (1)$$

- wektor SK_2 :

$$SK_2 = \left(\sum_{i=\frac{N}{2}}^N a_{i1}, \sum_{i=\frac{N}{2}}^N a_{i2}, \dots, \sum_{i=\frac{N}{2}}^N a_{iM} \right) \quad (2)$$

- a wektor SW_j :

$$SW_1 = \left(\sum_{j=0}^{\frac{M-1}{2}} a_{1j}, \sum_{j=0}^{\frac{M-1}{2}} a_{2j}, \dots, \sum_{j=0}^{\frac{M-1}{2}} a_{Nj} \right) \quad (3)$$

- wektor SW_2 :

$$SW_2 = \left(\sum_{j=\frac{N}{2}}^M a_{1j}, \sum_{j=\frac{N}{2}}^M a_{2j}, \dots, \sum_{j=\frac{N}{2}}^M a_{Nj} \right) \quad (4)$$

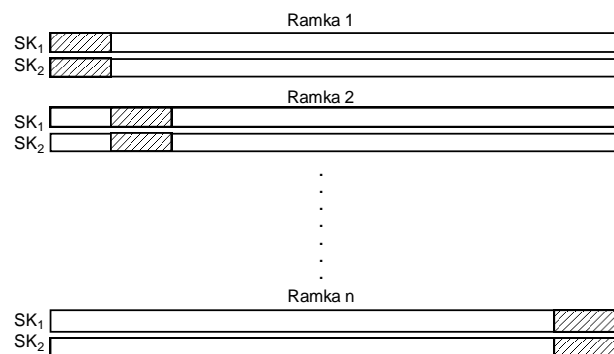
gdzie:

a_{ij} – zakumulowane różnice pomiędzy dwoma kolejnymi obrazami,

M – rozdzielczość pozioma obrazu,

N – rozdzielczość pionowa obrazu.

Analiza dwóch wektorów SK poprzedzona jest analizą całkowitej aktywności ruchu w obrazie. Wyniki eksperymentalne pokazują, że ten parametr mocno zależy od treści sekwencji. Lokalizacja poruszającej się krawędzi przejścia edycyjnego jest oparta o analizę rozkładu przestrzennego wartości w wektorze SK. Wektory SK reprezentują górną i dolną część obrazu.

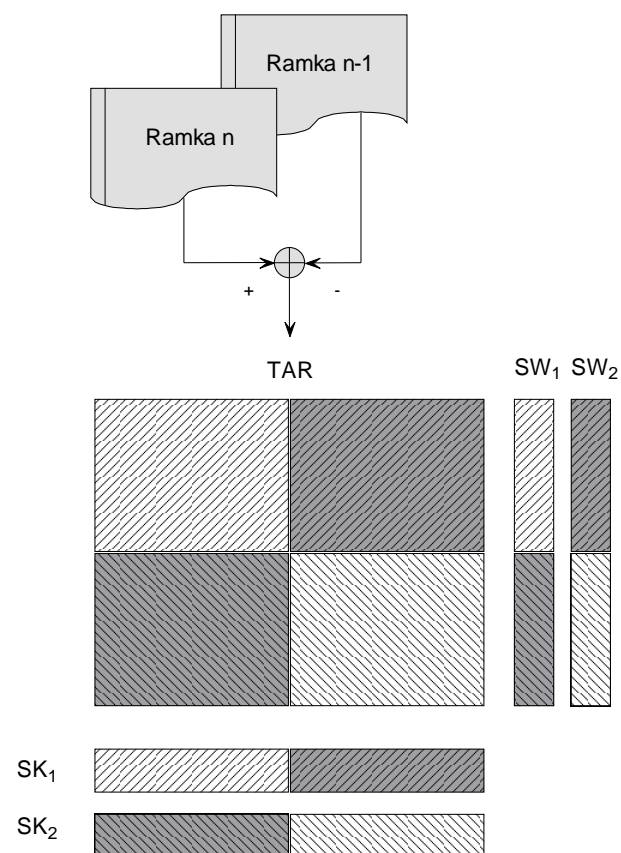


Rys. 1. Wektory SK_1 i SK_2 określające położenie krawędzi przemieszczającego się przejścia.

Ponieważ efektywność detekcji przejścia edycyjnego oparta o deskryptor aktywności ruchowej MAD jest niewystarczająca, dlatego do algorytmu dodany został deskryptor koloru dominującego DCD.

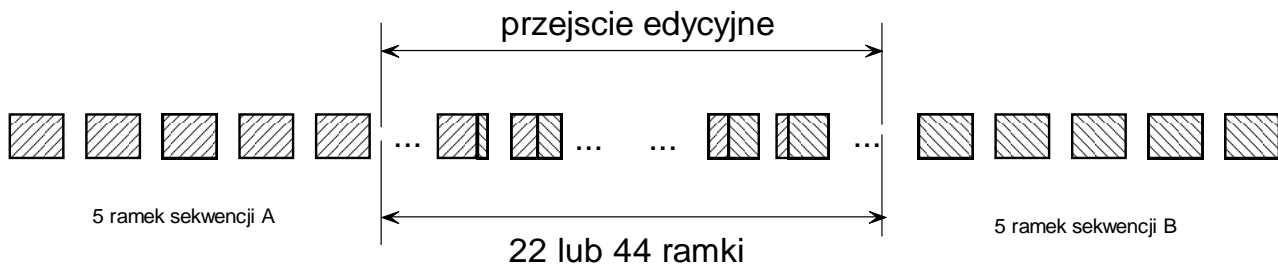
Deskryptor koloru dominującego jest najbardziej odpowiedni dla reprezentacji cech obszaru obrazu, ponieważ mała liczba kolorów jednoznacznie charakteryzuje obszar zainteresowania. Kwantyzacja kolorów jest używana do wydzielenia reprezentantów kolorów w każdym obszarze.

Cały obraz podzielony jest na dwa obszary, lewy i prawy. W proponowanym algorytmie, deskryptor koloru dominującego jest niezależnie obliczany dla każdego regionu. Jeżeli skwantowane wartości kolorów dominujących przekraczają próg, przejście edycyjne jest wykrywane.



Rys. 2. Graficzna ilustracja sposobu definiowania macierzy TAR i wektorów SK_1 , SK_2 , SW_1 i SW_2 .

Bardzo ważnym elementem algorytmu jest określenie progu akceptacji. Autor proponuje ustawić próg akceptacji w oparciu o wartości deskryptorów uzyskane dla pierwszego obrazu. W przypadku deskryptora MAD, najpierw definiowany jest parametr określający ruch całkowity a następnie jako wartość maksymalna ustalana jest wartość o 15% większa od wartości określającej ruch całkowity a wartość minimalna progu o 15% niższa od wartości określającej ruch całkowity.



Rys. 3. Struktura sekwencji wizyjnych użytych w eksperymentach.

3. WYNIKI EKSPERYMENTALNE

Badania zostały przeprowadzone dla sekwencji testowych 352 x 288, 25 Hz, 4:2:0.

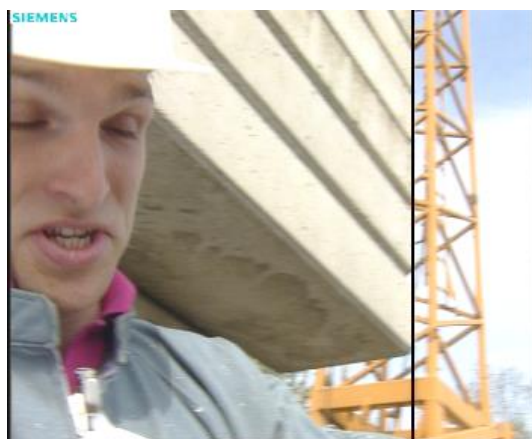
Deskryptor aktywności ruchowej i koloru dominującego zostały zaimplementowane jako oprogramowanie w języku C. Algorytm detekcji przejścia był testowany z sekwencjami testowymi CIF. Sekwencje testowe złożone były z dwóch sekwencji wizyjnych połączonych pomiędzy sobą przejściem edycyjnym (Rys. 3).

Efektywność algorytmu detekcji przejścia opartego wyłącznie na deskrytorze aktywności ruchowej MAD w dużej mierze zależy od treści sekwencji. Błędne decyzje są popełniane w średnio 9% przypadków (Tabela 1). Dlatego algorytm uzupełniono o wykorzystanie drugiego deskryptora.

Tabela 1. Efektywność detekcji przejścia wyznaczanego wyłącznie na podstawie analizy deskryptora aktywności ruchu MAD.

Sekwencja testowa	Liczba ramek, które zostały wykryte jako zawierające przejście edycyjne	Całkowita liczba ramek sekwencji	Numer ramki	Rysunek ilustrujący
<i>Basket</i>	0	496	-	-
<i>Buf</i>	0	296	-	-
<i>Cheerf</i>	0	296	-	-
<i>Football</i>	0	396	-	-
<i>Foreman</i>	4	298	182, 208, 216, 217	4, 5
<i>Fun</i>	413	496	6
<i>Hall_monitor</i>	0	328	-	-
<i>Paris</i>	0	1063	-	-
<i>Stefan</i>	0	596	-	-
<i>Table</i>	1	298	130	-
suma	418	4563		

Efektywność algorytmu detekcji przejścia edycyjnego oparta o analizę deskryptora koloru dominującego DCD jest gorsza niż na podstawie analizy deskryptora MAD. Liczba błędnych decyzji wynosi 28% całkowitej liczby przypadków (tabela 2). W proponowanym rozwiązaniu 100% efektywność detekcji i lokalizacji położenia przejścia edycyjnego w kierunku poziomym osiągnięta jest przez wykorzystanie kombinacji obu deskryptorów.



Rys. 4. Sekwencja *Foreman.cif*, ramka 182 (czarna linia oznacza niepoprawnie wykrytą krawędź przejścia edycyjnego).



Rys. 5. Sekwencja *Foreman.cif*, ramka 216 (czarna linia oznacza niepoprawnie wykrytą krawędź przejścia edycyjnego).

Jak pokazują wyniki badań eksperymentalnych (Tabela 3), para deskryptorów poprawnie określa ramki zawierające przejście. Tylko w przypadku jednej sekwencji *Table2Paris.cif*, w której przejście trwa 22 ramki, algorytm błędnie określił lokalizację położenia przejścia w jednej ramce. W sekwencji *Fun2Paris.cif* algorytm nie określił końca przejścia edycyjnego. Algorytm wskazuje pionowy element karuzeli jako krawędź przejścia łączącego treść dwóch sekwencji (rys. 6). Próg akceptacji definiowany był dla pierwszej ramki sekwencji występującej przed przejściem.

Sekwencja *Fun.cif* charakteryzuje się większą aktywnością ruchową treści wizyjnej i większymi zmianami

w kolorach dominujących niż ma to miejsce w przypadku sekwencji *Paris.cif*. Te dwa fakty wpływają znacząco na podejmowanie błędnych decyzji.



Rys. 6. Sekwencja *Fun.cif* (czarna linia oznacza niepoprawnie wykrytą krawędź przejścia edycyjnego).

Tabela 2. Efektywność detekcji przejścia edycyjnego wyznaczonego wyłącznie na podstawie analizy deskryptora koloru dominującego DCD.

Sekwencja testowa	Liczba ramek, które zostały wykryte jako zawierające przejście edycyjne	Całkowita liczba ramek sekwencji
<i>Basket</i>	111	496
<i>Bus</i>	296	296
<i>Cheer</i>	21	296
<i>Football</i>	21	396
<i>Foreman</i>	32	298
<i>Fun</i>	0	496
<i>Hall_monitor</i>	18	328
<i>Paris</i>	287	1063
<i>Stefan</i>	332	596
<i>Table</i>	171	298
sum	1289	4563

Tabela 3. Efektywność detekcji przejścia edycyjnego na podstawie deskryptorów MAD i DCD.

Sekwencja testowa	Czas przejścia [ramki]	Całkowita liczba ramek w sekwencji	Liczba ramek z poprawnie wykrytym przejściem edycyjnym	Liczba ramek z niepoprawnie wykrytym przejściem edycyjnym
<i>Table2Paris</i>	44	52	44	0
<i>Table2Paris</i>	22	33	21	0
<i>Foreman2-Paris</i>	44	52	44	0
<i>Foreman2-Paris</i>	22	33	22	0
<i>Fun2Paris</i>	44	52	44	3
<i>Fun2Paris</i>	22	33	22	3

Pomimo tego, aktualna pozycja przejścia edycyjnego jest poprawnie określona we wszystkich przypadkach. Należy zauważyć, że pozycja przejścia jest określona w kierunku poziomym z dokładnością punktu. W mechanizmie sterującym koderem, ta informacja może być określana wystarczająco z dokładnością makrobloku.

4. PODSUMOWANIE

Przyszłe kodery obrazu i sekwencji wizyjnych będą zdolne wykorzystywać metadane opisujące obraz/sekwencje wizyjną w celu poprawy efektywności ich kodowania lub w celu optymalizacji strategii kodowania. W artykule przedstawiono algorytm wykrywania przejścia edycyjnego, który może być użyteczny w algorytmach kompresji MPEG-2 [10] lub AVC/H.264, poprawiając efektywność ich kodowania. Zaproponowany algorytm oparty tylko na dwóch deskryptorach wizyjnych cechuje się blisko 100% efektywnością określenia pozycji przejścia edycyjnego w obrazie.

LITERATURA

- [1] Text of ISO/IEC 15938-3/FCD Information technology – Multimedia content description interface – Part 3 Visual, ISO/IEC JTC1/SC29/WG11/N4062, Singapore, March 2001.
- [2] B. S. Manjunath, T. Sikora, "Introduction to MPEG-7 Multimedia Content Description Interface, Chapter 6 – Overview of Visual Descriptors", John Wiley & Sons, Ltd, 2002.
- [3] Wang, Z. Liu and J. Huang, "Multimedia content analysis using audio and visual information," *IEEE Signal Processing Magazine*. vol. 17, no. 6, Invited paper in the special issue on joint audio-visual processing., Nov. 2000, pp. 12-36.
- [4] MPEG-7 Visual part of eXperimentation Model Version 10.0, ISO/IEC JTC1/SC29/WG11/N4063, Singapore, March 2001.
- [5] M. Wu, W. Wolf, B. Liu, "An Algorithm for Wipe Detection", In Proceeding of IEEE Int. Conf. on Image Processing (ICIP'98), Chicago, 1998.
- [6] M. S. Drew, Z.-N. Li, and X. Zhong, "Video dissolve and wipe detection via spatio-temporal images of chromatic histogram differences", In Proceeding of IEEE Int. Conf. on Image Processing (ICIP 2000), vol. 3, pp. 929–932, 2000.
- [7] A. M. Alattar, "Wipe Scene Change Detector for Segmenting Uncompressed Video Sequences", In Proc. IEEE Int. Sym. Circuits and Systems, Vol. 4, pp. 249-252, 1998.
- [8] V. Gavirneni, J.H. Oh; "Automatic Generation of Motion Activity Descriptors in Video Sequences", Department of Computer Science and Engineering, University of Texas at Arlington.
- [9] J. Ruiz and P. Salembier, "Metadata-based coding tools for hybrid video codecs", In 23rd Picture Coding Symposium, PCS'2003, pp. 473-477, Saint-Malo, France, April 2003.
- [10] S. Maćkowiak, M. Relewicz, "Metadata-based Coding Tools for MPEG-2 Video Coder", International Workshop on Systems, Signals and Image Processing, IWSSIP'04, pp. 159-162, Poznań, 2004.