Wipe Transition Detection based on Motion Activity and Dominant Colours Descriptors

Sławomir Maćkowiak, Maciej Relewicz
Institute of Electronics and Telecommunications, Poznań University of Technology, Poznań,
smack@et.put.poznan.pl, relu@simplusnet.pl

Abstract

Future image and video sequence encoders will be able to use the metadata information in order to improve their efficiency or introduce a new technique of error masking in a video decoder. The paper presents a wipe transition detection algorithm based on two video descriptors: Motion Activity and Dominant Colour. The information about transition could be useful in encoder in order to optimize their encoding strategy.

1. Introduction

MPEG-2 and MPEG-4 have been successful standards which have given a rise to widely adopted commercial products such as CD-interactive, digital audio broadcasting, digital television and many video-on-demand trials. The state of the art in video compression has just experienced a revolution with the new standard H.264/MPEG-4 AVC/MPEG-4 Part 10. Version 1 of the new video coding standard AVC/H.264 has been already developed. H.264/AVC offers a significant improvement of coding efficiency as compared to older compression standards such as MPEG-2.

It can be expected that in the future a very large amount of audio-visual files will be indexed and that metadata information will be rather easy to create. As a result, in many circumstances, audio-visual material will be available together with the metadata describing its content. The goal of MPEG-7 [1-3] is to provide tools for the description of multimedia content. In this context, each type of multimedia material is characterized by a set of distinctive features.

There exists a great demand for efficient description of multimedia content because the exploiting metadata could optimize encoding strategy of audio-video encoder [10].

Edition passages are used mostly in order to connect two different visual materials: scenes and commercials. They are not natural parts of visual sequences; they are introduced during the installation process. Owing to this fact, an encoder in the process of prediction cannot predict them. This causes a decrease in the quality of coded visual sequence. Due to the employment of the advanced video coding techniques it became necessary to use additional information accompanying the visual sequence (e.g MPEG-7 stream) to improve the efficiency of coding.

Wipe transition [5-7] is a special effect in which two pictures from different video sources are displayed on one screen. Wipe is a transition from one image to another, either by having the second image slide over and cover the first image, or by having it push the first image out of the display area. Wipe is very difficult to encode by the video encoder. In wipe we have a motion edge which moves from one side of the picture to another. This causes that contents of macroblocks and motion vectors in macroblocks are changing, The encoder must change many parameters of encoding like a quantization parameter, a structure of mackroblocks, the type of predictions in macroblocks, etc. The metadata description could help in a choice of decision in encoder and improve coding efficiency.

In some cases, it would be very convenient to think of modeling a particular property unique to each type of wipe transition. However, this model does not generalize well across the broad variety of transitions that are currently in use nowadays. Therefore, in order to maintain simplicity of this model the authors think about left to right wipe transition.

2. Wipe detection

During a wipe, each frame will have a portion of the old scene and the new scene. A single strip of the image changes between adjacent frames. For a horizontal wipe there is a vertical strip, and for a vertical wipe there is a horizontal strip. Since the scene transition occurs in the strip, the number of changing pixels within the strip should be higher than those in the rest of the image. The location of the changing pixels can be recorded and their spatial distribution analyzed.

Motion activity which is defined as a degree of activity, or amount of motion, in video sequence, has been included as a descriptor in MPEG-7 standard. Standard does not define of descriptor calculating method, only a syntax and semantics of MPEG-7 bitstream is defined.

In this paper, MADs calculating is based on automatic generation of motion activity descriptors [8]. However, the authors adapted this method to the wipe detection and improved their efficiency by extending the number of vectors, which are the summation of column values – SK and summation of row values - SW.

The matrix TAR (Fig. 2) represents the accumulated differences between two consecutive frames.

The following equations show how to compute values of vectors SK and SW:

- a vector SK1:

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

- a vector SK_2:

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

- a vector SW_1:

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

- a vector SW_2:

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

where:

aij – an accumulated differences between two consecutive frames,

M - a horizontal resolution of the frame,

N - a vertical resolution of the frame.

The analysis of two vectors SK precedes an analysis of total motion activity. The experimental results show, that this parameter depends on video sequences. Therefore, the algorithm defines the up and down thresholds for a total motion activity. The thresholds are set on the basis of first encoding frames. However, this parameter gives information only about the motions occur. The localization of moving edge of wipe transition in frames is based on spatial distribution analyzed of values in vectors SK. Vectors SK represent the top and the bottom part of the frame. The local motion activity only in the bottom or the top part of the frame does not cause a false decision in wipe detection.

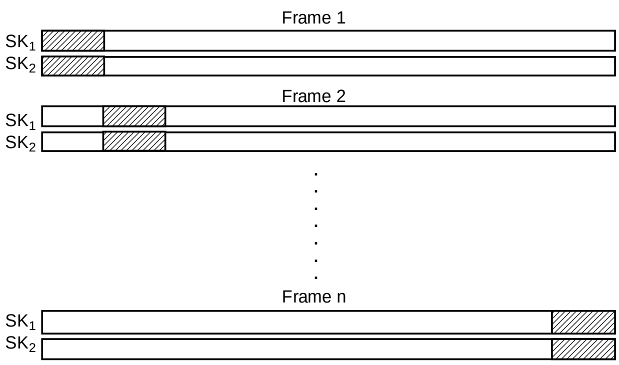

Fig. 1.The vectors SK1 and SK2 define a position of moving edge of wipe.

Since the efficiency of wipe detection based only on MADs analysis is not ideal, the Dominant Colour Descriptor is used.

The Dominant Colour Descriptor is most suitable for representing image region features when a small number of colours is enough to characterize the colour information in the region of interest. Colour quantization is used to extract a small number of representing colours in each region.

The whole image is divided into two image regions, left and right. In proposed algorithm, the Dominant Colour Descriptor DCDs is independently calculated in each region. If quantized values of colours cross a threshold, the wipe is detected.

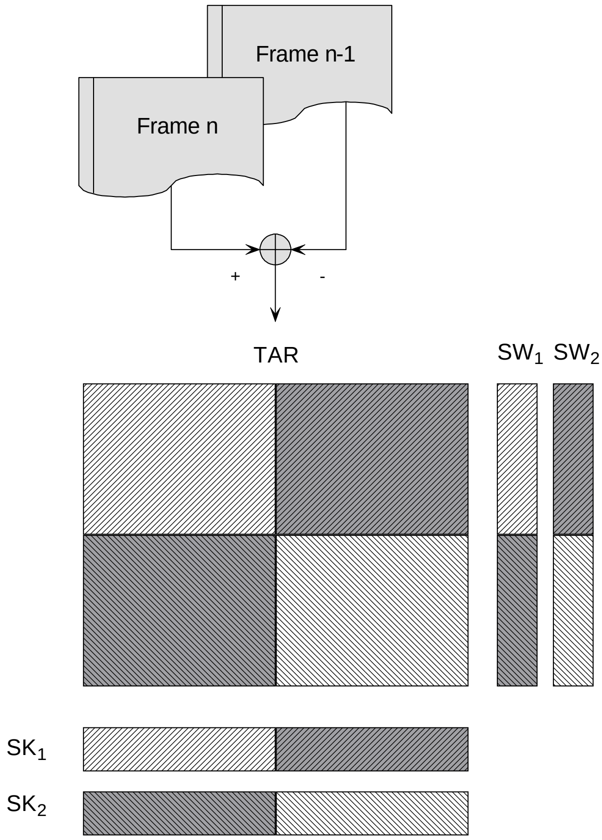

Fig. 2. A graphic illustration, how the TAR matrix and SK1, SK2, SW1 and SW2 vectors are defined.

Very important information in the detection algorithm is a level of the acceptance threshold. The authors propose to set the level of acceptance on the basis of the first frame in the sequence. In the case of the MAD, the first, the total motion is calculated and, next, we set the maximum value of the level of the acceptance threshold at the 15% higher than the total motion value and minimum value of the level of the acceptance threshold at the 15% lower than the total motion value.

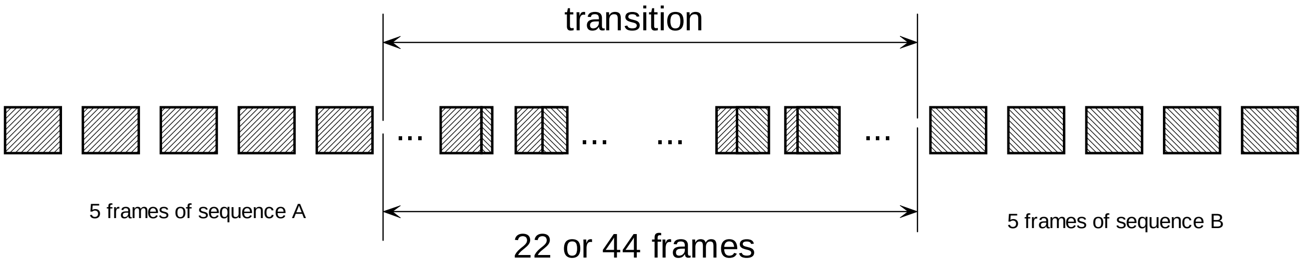

Fig. 3. The structure of sequence composition in experiments.

3. Experimental results

The experiments have been performed with 352 x 288, 25 Hz, 4:2:0 test sequences.

The Motion Activity and Dominant Colour Descriptors have been implemented as software written in C language. The detection algorithm was tested with CIF video sequences. The experimental test sequences consist of two sequences and wipe transition between them (Fig.3).

The efficiency of wipe detection based only on MADs analysis depends mostly on video sequence contents. The false decision efficiency is about 9 % of total frames (Table 1). Therefore, second descriptor has been added in algorithm.

Table 1. The efficiency of wipe detection based only on MAD analysis.

Sequence	The number of frames which was detected as a wipe	Total number of frames	Number frame	Figure
Basket	0	496	-	-
Buf	0	296	-	-
Cheerf	0	296	-	-
Football	0	396	-	-
Foreman	4	298	182, 208,216, 217	*4,5*
Fun	413	496	.....	*6*
Hall_monitor	0	328	-	-
Paris	0	1063	-	-
Stefan	0	596	-	-
Table	1	298	130	-
sum	418	4563		

The efficiency of wipe detection based on DCDs analysis is worse than MADs analysis. The false decision efficiency is about 28 % of total frames (Table 2). In proposed solution the 100% efficiency of wipe detection and wipe localization in horizontal dimension is achieved by exploiting combination of both descriptors (Table 3).

Fig. 4. The *Foreman.cif* sequence, the frame number 182 (the black vertical line denotes the incorrectly detected edge of the wipe).

Fig. 5. The *Foreman.cif* sequence, the frame number 216 (the black vertical line denotes the incorrectly detected edge of the wipe).

As the experimental results show (Table 3), the pair of descriptors correctly detects the frames including wipe. Only in one test sequence Table2Paris.cif, in which wipe go on across 22 frames, the algorithm cannot detect 1 frame. In the Fun2Paris.cif sequence the algorithm cannot indicate the end point of the transition. The algorithm indicates moving edges of the carrousel as a wipe (Fig.6). The acceptation threshold was defined for the first frame of the Paris.cif sequence. The Fun.cif sequence is characterized by more motion activity and more changes of the dominant colors than the Paris.cif sequence. These two facts impose that the detection decision is failed.

Fig. 6. The *Fun.cif* sequence (the black vertical line denotes the incorrectly detected edge of the wipe).

Table 2. The efficiency of wipe detection based only on DCD analysis.

Sequence	The number of frames which was detected as a wipe	Total number of frames
Basket	111	496
Bus	296	296
Cheer	21	296
Football	21	396
Foreman	32	298
Fun	0	496
Hall_monitor	18	328
Paris	287	1063
Stefan	332	596
Table	171	298
sum	1289	4563

Table 3. The efficiency of wipe detection based on both MAD and DCD analysis.

Sequence	Transition time [frames]	Total number of frames	The number of frames with correctly detected wipes	The number of frames with incorrectly detected wipes
Table2Paris	44	52	44	0
Table2Paris	22	33	21	0
Foreman2-Paris	44	52	44	0
Foreman2-Paris	22	33	22	0
Fun2Paris	44	52	44	3
Fun2Paris	22	33	22	3

Nevertheless, the current position of the wipe is correctly detected in all test cases. It is important to note, that the position of the wipe is defined with accuracy to the horizontal position of the pixel. In steering mechanism in the encoder this information can be defined with macroblock position accuracy.

4. Conclusions

Future image and video sequence encoders will be able to use the metadata information in order to improve their efficiency or to optimize their encoding strategy. The paper presents a wipe transition detection algorithm which could be used in MPEG-2 [10] or AVC/H.264 encoder in order to improve their efficiency. This algorithm based only on two descriptors gives us a 100 % precise position of the right-to-left wipe transition.

5. References

[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.