Metadata-based Coding Tools for MPEG-2 Video Coder

S. Maćkowiak, M. Relewicz

Poznań University of Technology, Institute of Electronics and Telecommunications,
Piotrowo 3A, 60-965 Poznań, POLAND
e-mail: 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. The paper reports the results of work which concerns the improving of MPEG-2 coder efficiency. The encoder will be able to access to the Motion Activity and Dominant Colour Descriptors information and use it in order to optimize their encoding strategy.

Keywords: MPEG-2, video coding, MPEG-7, metadata

I. INTRODUCTION

It can be expected that in the future, a very large amount of audio-visual documents 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 improve coding efficiency of video coder [9] or introduce new technique of error masking in video decoder.

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

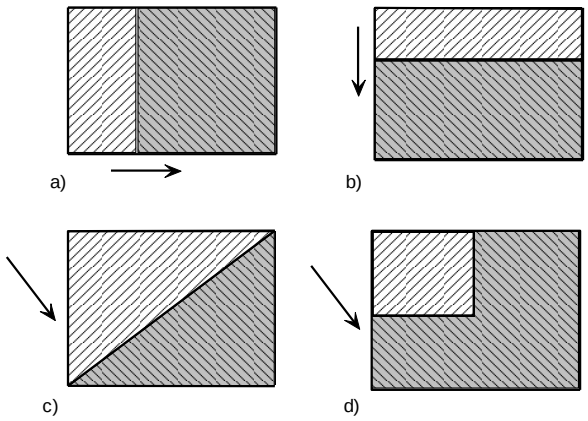

Fig.1. In wipe transitions, the edge may be different shapes, such as a straight line, a wedge, or a zigzag: a) a bar moves from left to right, b) a bar moves from top to bottom, C) a diagonal line moves from the upper-left corner to the lower-right corner, D) a box expands from the upper-left corner to the lower-right corner.

Wipe transition [5-7] is a special effect in which two pictures from different video sources are displayed on one screen (Fig. 1). Special effects generators provide numerous wipe patterns varying from simple horizontal and vertical wipes to multi-shaped, multi-colored arrangements. Wipe is very difficult to encode by the video coder. 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 and 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, that model does not generalize well across the broad variety of transitions currently in use today. Therefore in order to maintain simplicity of this model the author think about left to right wipe transition.

II. 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 that method to the wipe detection and improve 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. 3) represents the accumulated differences between two consecutive frames.

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

- a vector SK_1:

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

- 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:

a_{ij} – 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 shows, that this parameter depends on video sequences. Therefore the algorithm defines the up and down thresholds for a total motion activity. The thresholds are setting based on first encoding frames. However this parameter gives only information about the motions occurs. The localization of moving edge of wipe transition in frames is based on spatial distribution analyzed of values in vectors *SK*. Vectors *SK* represent a top and bottom part of frame. The local motion activity only in bottom or a top part of frame does not cause a false decision in wipe detection.

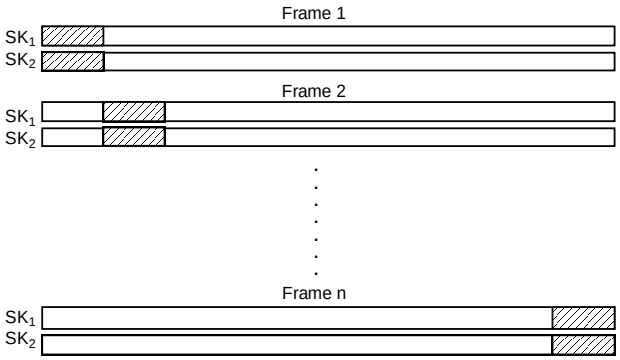

Fig. 2.The vectors SK_1 and SK_2 define a position of moving edge of wipe.

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

The Dominant Colour Descriptor is best suitable for representing image region features when a small number of colours are 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 independent calculated in each region. If quantized values of colours cross a threshold, the wipe is detected. The efficiency of wipe detection based on DCDs analysis is worse than MADs analysis. The false decision efficiency is about 28 % of total frames. In proposed solution the 100% efficiency of wipe detection and wipe localisation in horizontal dimension is achieving by exploiting combination of both descriptors.

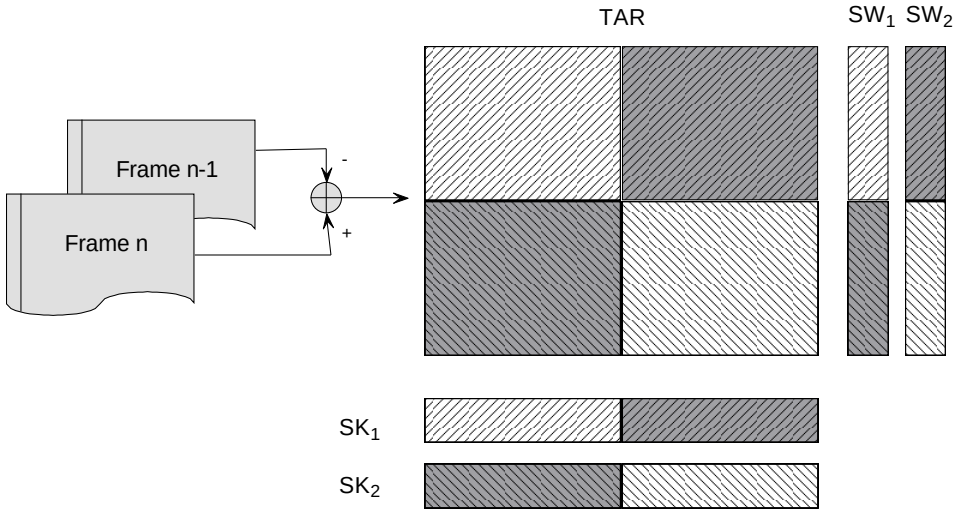

Fig. 3. A graphic illustration, how the TAR matrix and SK_1, SK_2, SW_1 and SW_2 vectors are defining.

III. SCENARIOS OF MPEG-2 VIDEO CODING

General structure of the MPEG-2 coder is not modified. The authors proposed only a several different scenarios of macroblock encoding in P-frames. The modification of encoding strategy is applied only to GOP inclusive a wipe transition.

The macroblocks of the frame with wipe transition are more difficult to encode because prediction is not good. Therefore authors propose encoding of macroblocks in P-frame as an I-type macroblocks. The application of this scenarios lead to higher SNR, compared to standard strategy. The higher SNR of a P-frame denote that the reference frames for the B-frames are also better and the B-frames have lower bitrates and higher SNR.

The authors proposed the following scenarios of macroblock encoding:

a) All macroblocks on the left side from the moving wipe edge are intraframe encoding (Fig. 4.a).

b) Every second row of the macroblocks on the left side from the moving wipe edge are intraframe encoding (Fig. 4.b). This scenario is a compromise between the number of intraframe coding macroblocks and a quality.

c) All macroblocks in the first column of the frame are intraframe encoding (Fig. 4.c). Since wipe is from the left side to the right side, all new contents ingoing into the frame is difficult to encoding.

d) Every second macroblock in the first column of the frame are interframe encoding (Fig. 4.d).

e) The left part of the frame is intraframe encoding (Fig. 4.e).

f) All macroblocks in P-frame are encoding as a I-type macroblocks (Fig.4.f).

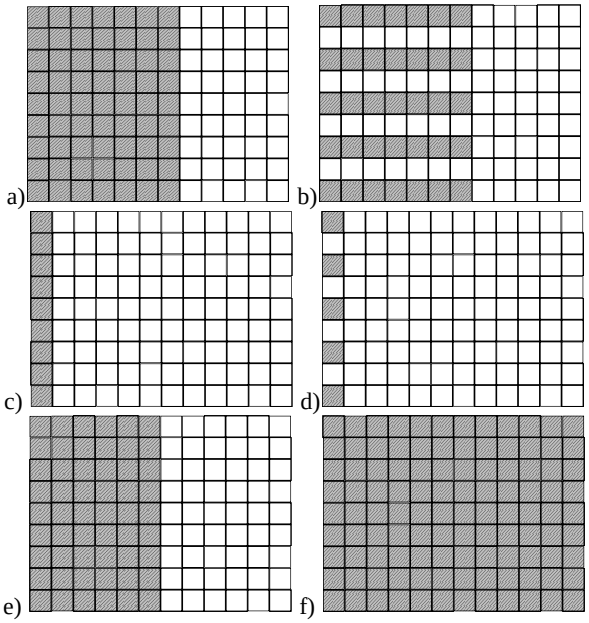

Fig. 4. The proposed scenarios of macroblock coding for P-frames. The grey fields represent macroblocks, which are intraframe encoding.

Another modification of encoding scenario consists in change all I-frames into P-frames in a GOP which contain wipe transistion. This reallocation of bits budget provide more bits for P-frames and an averaged SNR of frames is higher. In all P-frames, locally the encoder also can choose the above scenarios for macroblocks encoding. Therefore this strategy is more efficient for a wipe transition encoding.

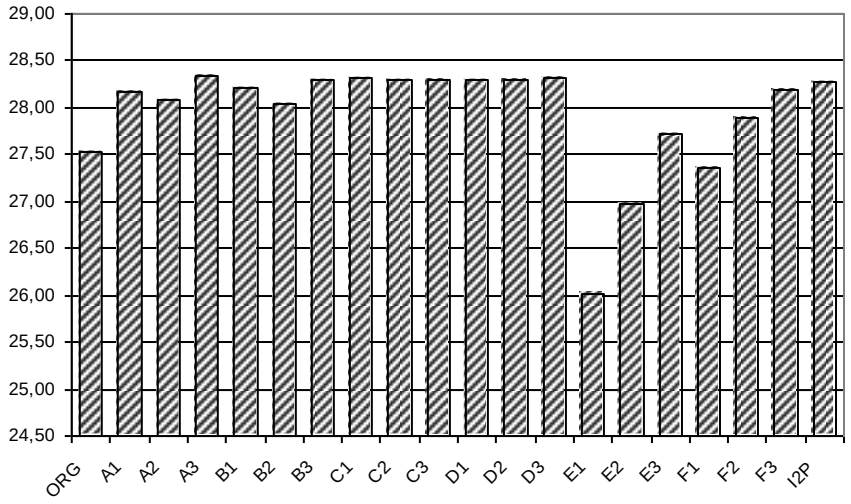

Fig. 5. The SNR results for experiment with transition between *Paris* and *Foreman* sequences, the time transition is 44 frames (a letter denotes the type of considered scenarios of macroblock encoding, a digit denotes every, every second or every third P-frames with scenario applied, I2P – I-frame to P-frame modification in a GOP contain a wipe transition).

IV. EXPERIMENTAL RESULTS

The Motion Activity and Dominant Colour Descriptors have been implemented as software written in C language. This software was included to the MPEG-2 verification model [10].

The coder is aimed at processing of progressive 352 x 288, 25 Hz test sequences. The experiments have been performed with 4:2:0 sequences as the verification model implements the coder that is an augmented version of Main Profile @ Main Level MPEG-2 coder.

Several experiments performed by the authors prove that proposed scenarios result in more precise prediction. This improvement in quality is exploited in the predicition of the next pictures. Figure 5 contains experimental results for presented earlier scenarios.

In the experiments the authors achieved a SNR gain up to 0,8dB contrary to encoding without proposed decision strategy.

V. 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 strategy. The paper reports the first stage of work which concerns the improving of MPEG-2 coder efficiency. A set of metadata based coding tools have been presented.

The results reported here show promising gains when enabling metadata based coding into current standards video codecs such as MPEG-2.

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, pp. 12-36, Nov. 2000. Invited paper in the special issue on joint audio-visual processing.

[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] Vasavi Gavirneni, JungHwan 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] html://www.mpeg.org/MSSG/tm5