FAST EVENT DETECTION IN MPEG VIDEO

Rui Marcelino, Vitor Silva and Sérgio Faria
School of Technology, University of Algarve, Faro, Portugal

Institute of Telecommunications, DEEC, University of Coimbra, Coimbra, Portugal

School of Technology and Management, Polytechnic Institute of Leiria, Leiria, Portugal

Keywords: MPEG video, video compression, video surveillance, motion detection

Abstract: Many video applications, such as surveillance systems are continuously increasing and the amount of processed and stored data has risen exponentially. In order to manage efficiently this video information, motion detection is necessary. This feature is required to analyze, organize and store compressed video. In this paper, we present an effective video event detection method, which uses information embedded in the MPEG-4 bit stream to detect true motion in the scenario, avoiding other features like scene cuts and camera translation, zooming, pan, tilt and oscillations These events can be detected very fast and with low computational complexity, as only few parameters of the compressed data are processed. This algorithm mainly relies on the amount of signal variation of AC coefficients, between co-localized intra coded blocks, and the amount of motion compensated coded macroblocks within inter coded frames. Our results have shown that this algorithm can perform not only accurate motion detection, but also identifying false motion due to camera movements.

1 INTRODUCTION

Video motion detectors for digital compressed domain interfaces are a key tool in modern surveillance system architectures (C. S. Regazzoni et al, 2001). They allow remote sensing over data networks and the processed data can be easily recorded and integrated in computer applications, for helping human operators in supervision tasks.

A typical surveillance scenario may include a video camera, or a set of cameras, pointing to a quite zone, for long periods of time. Thus, most of the time, the received visual information is useless. However, when an important image change occurs, like human intrusion, the operator should be automatically warned and the recording system must be triggered. When these cameras are installed in remote places, video compression is required, prior to the transmission. At the receiver, the compressed signal (bit stream) can be either decompressed and visualized by an operator or stored. Usually it has to be decompressed down to the pixel domain, either for human visualization or to be processed for feature extraction. However, most of the feature extraction can be performed using parameters included in the compressed data, without completely decompressing the video bit stream. This task can be performed by an efficient and intelligent method using a motion detector.

Several methods and approaches for motion detection have been proposed in the literature (I. Koprinska et al, 2001). Motion detection can be performed in pixel domain, or uncompressed domain, using features extracted from the digital spatio-temporal video representation, with very computational demanding techniques (I. Koprinska et al, 2001), (A. Albiol et al, 2003).

Since we assume the digital video signal is compressed using MPEG-*x* standards (F. Pereira and T. Ebrahimi, 2002), the motion detection can be done using the information embedded in the bit stream (I. Koprinska et al, 2001), (J. Pons et al, 2002), (B. Yeo and B. Liu, 1995), (S. Pei and Y. Chou, 1999), (S. Lee et al, 2000). By partially decoding (parsing) the coded bit stream, one can find quickly and easily the useful information to build a motion detector (J. Pons et al, 2002). Such techniques are designated by compressed domain solutions.

Marcelino R., Silva V. and Faria S. (2004).
FAST EVENT DETECTION IN MPEG VIDEO.
In *Proceedings of the First International Conference on E-Business and Telecommunication Networks*, pages 264-268
DOI: 10.5220/0001386002640268

The main purpose of this work is to use, as much as possible, the embedded information, taking advantage of the huge amount of analysis work performed by the MPEG video encoder. Furthermore, only few parameters have to be adjusted in the detector, regarding to the class of the moving object in scene. This class is related to the dimension of the moving object in scene and to the distance between the object and the camera. This image classification will increase the accuracy of the motion detector.

In the following sections we will describe an efficient and low complexity scene change detector algorithm, which is able to detect significant visual events from a partially decoded MPEG bit stream. In section 2 we introduce MPEG standard and in section 3 the proposed algorithm is described. Some results are shown in section 4 and conclusions are presented in section 5.

2 MPEG BIT STREAM INFORMATION

MPEG encoders use a hybrid algorithm to compress video, by classifying and processing each frame as intra coded (I frame) or motion compensated inter coded (P and B) (F. Pereira and T. Ebrahimi, 2002). Intra frame pictures are encoded only using pixels within a frame, exploring the spatial redundancy with 8×8 DCT (Discrete Cosine Transform) blocks are transformed and DC and AC coefficients are entropy coded. P frames are encoded using motion compensated prediction from a past I/P frame, in order to remove the temporal redundancy. B frames are encoded using motion compensation prediction from both past and/or future encoded I/P frames. Video frames are organized in regular structures called group of pictures (GOP). Each frame (VOP) is divided into blocks of 16×16 pixels, called macroblocks (MB). Furthermore, each macroblock is divided into six 8×8 pixel blocks. After motion compensation, the residual image may also be divided into 8×8 pixel blocks, which are intra coded. Thus, a macroblock contains information about the type of temporal prediction used (or not) for motion compensation, which can be classified as intra coded, forward referenced, backward referenced, interpolated or direct. While MBs inside an I frame are intra coded, each MB in a P frame is either forward predicted, intra coded or skipped. Similarly, each MB in a B frame is either forward predicted, backward predicted, bidirectionally predicted, intra coded or skipped.

3 COMPRESSED DOMAIN MOTION DETECTION

In this section, we explain how motion detection is performed without fully decoding the bit stream. The proposed method mainly relies on the analysis of AC coefficient's signal of I frames (section 3.1) and on the motion vector information of P and B coded frames (section 3.2). The main objective is to detect only motion related to the moving objects in the scene, eliminating camera switching (scene cuts) and some typical camera movements, which occurs in video surveillance scenes.
13

3.1 Motion detection

In most surveillance applications, systems acquire and store images continuously, then a huge amount of information is required to be stored. In this case a high compression ratio is desirable. It is also common, that for long periods of time there are no motions in the scene. Thus, VOPs of type I can be sparser, which increases significantly the compression ratio. In this sense, we propose a hierarchical algorithm that processes the compressed video information in two stages.

At the first stage, only I VOPs are analyzed, in order to check the signal variations between AC coefficients of two co-localized blocks in consecutive I VOPs. In order to speed up the process only a small set of significant coefficients are checked, and blocks with a number of coefficients with signal variation larger than 5 is used. When a number of blocks in this condition exceed a certain threshold, the image is regarded as containing a moving object. This threshold is obtained regarding the average and the variance of the number of blocks containing more than 5 signal variations. We also have to deal with homogeneous surfaces and illumination changes, which tend to be detected as motion. When a VOP of type I is detected with moving objects, the algorithm moves to the second stage for a motion detection refinement.

At the second stage, motion vectors of P and B VOPs are analyzed, in order to check the amount of motion vectors (MV) used to encode each inter frame. If the number of non-zero MVs exceeds a threshold given for that class (section 3.3) of surveillance scene, then the VOP is regarded to contain a moving object.

After this step, it may happen that some motion detections are false, due to camera switching (scene cuts) or camera motions. These false motion detection events have to be eliminated, in order to increase the algorithm efficiency. Thus, a scene cut detection method is used (A. Hanjalicl, 2002), (J. Calic and E. Izquierdo, 2002), (Y. Haoran et al, 2003) and the camera motion is detected (R. Wang and T. Huang, 1999), as explained in section 3.2.

Finally, a report is generated and objects in a moving scene are decoded for visualization.

3.2 Camera motion

Camera movements like: pan, tilt, zoom, shaking and vibration, caused by wind in outdoor environments, are sources of false positives motion detections. In order to remove such false positives, we have incorporated in our method a camera motion estimation module.

The camera motion, mentioned above, is well characterized in a frame analysis, by a large number of forward predicted MB and near homogeneous vector fields.

Figure 1: Motion vectors field for a zoom-in and zoom-out movement (static background MBs are dark colored)

In Figure 1, the CIF "Telex" sequence represents a zoom-in (left image) and zoom-out (right image) camera motion focusing a Telex equipment. This type of motion generates a large amount of motion vectors, which can be divided in various sub-sets of vectors with radial direction. When the image texture is not homogeneous, a large number of MBs are encoded with motion vectors, whose intensity depends on the camera motion. Otherwise, when there are homogeneous texture the number of motion compensated (MC) coded MB is reduced, as can be seen in figure 2, where static background MBs are dark colored.

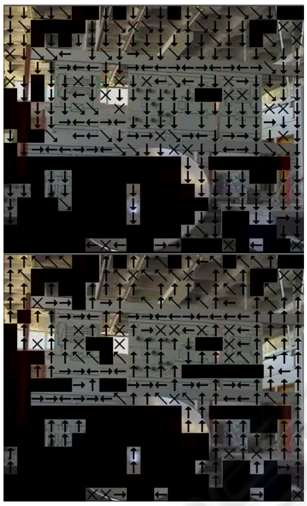

Figure 2: Motion vector's field for a camera oscillation

Using VOPs 216 and 219 of the CIF video sequence called "Room121", we have tested the algorithm for detection of false motion when camera oscillation occurs. In the left image of Figure 2, we have a zoom-in and in the right image a zoom-out. From the analysis of the motion vector's direction, the MVs inversion in the direction can be detected between both images.

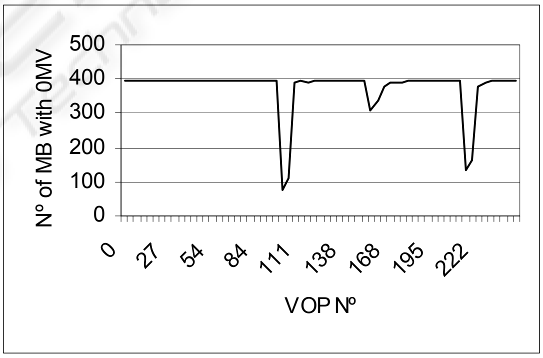

Figure 3: Oscillation detection

Additionally in Figure 3, we show the huge reduction in the amount of MBs with null MVs, when a camera motion occurs. For example the oscillation in VOP 219 can be clearly detected.

3.3 Scene classification

The scene classification is an important issue, as it is directly related to the motion detection. This task must be performed by an operator, regarding the surveillance camera system, namely, distance from the scene, zoom lens and target object's size. Thus, we have divided the surveillance scenes in three classes: A, B and C. These classes have a direct correspondence to the number of expected blocks or macroblocks with non-zero motion vectors, within a VOP. The chosen threshold directly determines the detection performance. Due to such direct mutual dependence, the detection performance is highly sensitive to specified parameter values. Beside the threshold sensitivity, the problem of specifying such a precise value remains and, consequently, the scope of the validity of such an accurate threshold is highly questionable. Clearly, manual threshold specification cannot be avoided in practical implementations. Thus, there must have an installation and set-up phase where the sensitivity of the motion sensor must be adjusted. However, the influence of these parameters on the detection performance can be diminished and the detection can be made more robust if we use lower threshold levels. In fact, it is preferable a false alarm rather than a missed alarm.

4 EXPERIMENTAL RESULTS

In this section, we evaluate our motion detector for video surveillance systems. We have performed a set of experiments using videos obtained from surveillance systems installed in the campus, which have been encoded in MPEG-4 format with CIF spatial resolution.

Figure 4: Test sequences: Pupils, Hall, Door125 and Park, in this order

The length of these videos is between 241 (Pupils) and 846 (Park) frames long. The original sequences are illustrated in figure 4 as Pupils, Hall, Door125 and Park, were carefully obtained in order to include many effects, covering the largest number of different situations. The experimental results demonstrate the efficiency of the proposed motion detection algorithm.

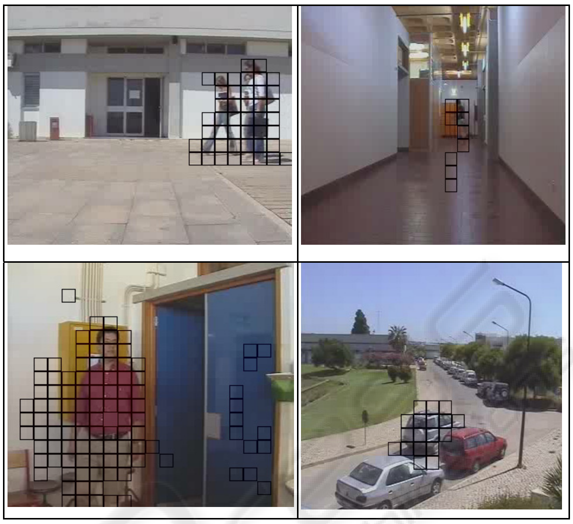

Figure 5: Encoded MBs referenced as black squares

Figure 5 gives an example of motion detection results in various sequences scenes. No filter is used and some MBs have been coded due to noise and object's shadow.

Table 1: Precision and recall results

Sequence	Length	Precision (%)	Recall (%)
Pupils	241	100	90
Hall	308	100	87
Door125	392	97	92
Park	846	79	65

The performance is given in terms of *precision* and *recall* parameters (U. Gargi et al, 2000),

$$precision = \frac{N_C}{N_C + N_E}, \quad recall = \frac{N_C}{N_C + N_M}, \qquad (1)$$

where N_C is the number of correct motion detections, N_E is the number of incorrect motion detections and N_M is the number of missed motion detections.

These results from our experiments presented in table 1, illustrate *precision* and *recall* values very close to 100% and 90%, respectively, for most sequences. Although rates of true and missed detections are not precisely the same for all sequences, there are no outliers in the performance. We can say that the performance of this detector remains relatively consistent over all sequences. Those values of Park sequence are related to the appearance in scene of moving objects of distinct classes. These are cars in the natural plane of the scene and other cars parking at a long distance from the camera, almost indistinguishable points in the scene.

5 CONCLUSIONS

In this paper we propose an efficient and low complexity unsupervised hierarchical motion detection algorithm for surveillance systems, and showed its performance using MPEG-4 video compression data. The key idea of this motion detector is to analyze the motion vector information, embedded in the compressed data and decide if they represent object's motion in the scene.

In case it does not represent true motion in the scene, motion vectors data is analyzed to determine the meaning of the false motion detection. Various techniques have been implemented to detect: scene cuts, zoom, camera translation and camera oscillation. These methods strongly reduce the incorrect motion detection rate.

REFERENCES

C. S. Regazzoni, V. Ramesh and G. L. Foresti (Eds.), "Special Issue on Video Communications, Processing and Understanding for Third Generation Surveillance Systems", *Proceedings of IEEE*, vol. 89, nº. 10, October 2001.

I. Koprinska and S. Carrato, "Temporal Video Segmentation: A Survey", *Signal Processing: Image Communication*, vol. 16, pp. 477-500, 2001.

A. Albiol, C. Sandoval, A. Albiol, V. Naranjo and J. M. Mossi, "Robust Motion Detector for Video Surveillance Applications", in *Proc. IEEE ICIP 2003*, September 2003.

F. Pereira and T. Ebrahimi (Eds.), *The MPEG-4 Book*, Prentice Hall PTR, 2002.

J. Pons, J. P. Nebot, A. Albiol and J. Molina, "Fast Motion Detection in Compressed Domain for Video Surveillance", *Electronics Letters*, vol. 38, nº. 9, pp. 409-411, April 2002.

B. Yeo and B. Liu, "Rapid Scene Analysis on Compressed Video", *IEEE Trans. on Circuits and Systems for Video Technology*, vol. 5, nº. 6, pp. 533-544, December 1995.

S. Pei and Y. Chou, "Efficient MPEG Compressed Video Analysis Using Macroblock Type Information", *IEEE Trans. on Multimedia*, vol. 1, nº. 4, pp. 321-333, Dec. 1999.

S. Lee,Y. Kim and S.Choi,"Fast Scene Detection using Direct Feature Extraction from MPEG Compressed Videos", *IEEE Trans. on Multimedia*, vol. 2, nº. 4, pp. 240-254, Dec. 2000.

A. Hanjalic, "Shot Boundary Detection: Unraveled and Resolved?", *IEEE Trans. on Circuits and Systems for Video Technology*, vol. 12, nº. 2, pp. 90-105, February 2002.

J. Calic and E. Izquierdo, "Temporal Segmentation of MPEG Video Streams", *EURASIP Journal on Applied Signal Processing*, nº. 6, pp. 561-565, 2002.

Y. Haoran, D. Rajan and C. L. Tien, "A Unified Approach to Detection of Shot Boundaries and Subshots in Compressed Video", in *Proc. IEEE ICIP 2003*, Sep. 2003.

R. Wang and T. Huang, "Fast Camera Motion Analysis in MPEG Domain", in *Proc. IEEE ICIP 1999*, vol. 3, pp. 691-694, October 1999.

U. Gargi, R. Kasturi and S. Strayer, "Performance Characterization of Video Shot Change Detection Methods", IEEE Trans. on Circuits and Systems for Video Technology, vol. 10, nº. 1, pp. 1-13, February 2000.