

EVENT RECOGNITION IN SPORT PROGRAMS USING LOW-LEVEL MOTION INDICES

R. Leonardi, P. Migliorati, A. Bonzanini

DEA - University of Brescia , Via Branze, 38 - 25123, Brescia, Italy

Tel: +39 030 3715434; Fax: +39 030 380014

e-mail: leon@ing.unibs.it

ABSTRACT

In this paper we present a semantic video indexing algorithm based on finite-state machines and low-level motion indices extracted from the MPEG compressed bit-stream. The problem of semantic video indexing is actually of great interest due to the wide diffusion of large video databases. In literature we can find many video indexing algorithms, based on various types of low-level features, but the problem of semantic indexing is less studied and surely it is a great challenging one. The proposed algorithm is an example of solution to the problem of finding a semantic relevant event (e.g., scoring of a goal in a soccer game) in case of specific categories of audiovisual programmes. The simulation results show that the proposed algorithm can effectively detect the presence of goals and other relevant events in sport programs.

1. INTRODUCTION

Nowadays we are able to store and transmit large quantities of audiovisual documents, thanks to the development of devices with large storage and transmission capabilities and to the use of efficient data compression techniques.

The need of an effective management of audiovisual material has brought to the development of semantic indexing techniques [1] [2] useful for various applications, such as, for example, electronic program guides.

To face the semantic indexing problem, a man uses its cognitive skills, while an automatic system can face it in two steps: in the first step some low-level indices are extracted in order to represent low level information in a compact way; in the second step we need a decision-making algorithm to extract a semantic index from the low-level indices.

While the problem of low-level indices extraction is widely discussed in the literature, the problem of finding decision-making algorithms is less addressed [3]. Furthermore it is quite clear that this problem can be partially solved only for particular types of sequences, i.e., with specific program categories.

In this work we have considered the problem of semantic video indexing using only the associated motion information. We have addressed this problem trying to identify a correlation between semantic features and low-level indices associated to a video sequence. In particular we have analyzed soccer video sequences; the standard format chosen for their representation is MPEG-2, which is widely used for storage and transmission of TV signals.

For soccer video sequences the semantic content can be related to the presence of interesting events such as, for example, goals, shots to goal, and so on. These events can be found at the beginning or at the end of the game actions. A good semantic index of a soccer video sequence could be therefore a summary made up of a list of all game actions, each characterized by its beginning and ending event. Such a summary could be very useful to satisfy various types of semantic queries.

In particular we have chosen three low-level indices which represent the following characteristics: (i) lack of motion, (ii) camera operations (represented by pan and zoom parameters) and (iii) the presence of shot-cuts. We have then studied the correlation between these indices and the semantic events defined above, and we have found that they are really meaningful [4].

In this work we propose an algorithm based on finite-state machines that exploits this correlation in order to detect the presence of goals and other relevant events. The simulation results show the good performance of the proposed algorithm.

The paper is organized as follow. In Section 2 we discuss the low-level motion indices, whereas in Section 3 the algorithm used for shot-cut detection is described. In Section 4 the proposed algorithm is presented, while in Section 5 we report some experimental results. Final conclusions are drawn in Section 6.

2. LOW-LEVEL MOTION INDICES

Motion data related to a video sequence are typically the motion vectors associated at each image. In our work the motion information are directly extracted from the com-

pressed MPEG2 domain, where each frame is divided into MacroBlocks and for each MacroBlock is given: the MacroBlock type; a motion vector obtained with a block-matching algorithm with respect to a reference frame.

The MacroBlock type is strictly related to motion vectors as follows: if a MacroBlock is Intra-Coded, no motion vector is given; if a MacroBlock is No-Motion-Coded, the motion vector is null; otherwise a not-null motion vector is given.

The motion field can be compactly represented with various type of descriptors, e.g.: temporal and spatial mean and standard deviation of phase and magnitude of vector field; phase or magnitude histograms; camera motion parameters [5].

These compact representations should anyway be combined with other indices suitable to state its reliability or to add other useful information [6].

As previously mentioned, in our work we have considered three low-level indices which represent the following characteristics: (i) lack of motion, (ii) camera operations (represented by pan and zoom parameters) and (iii) the presence of shot-cuts.

In this section we consider the low-level motion indices, namely: "lack of motion" and "camera motion parameters", pan and zoom.

Lack of motion has been evaluated by thresholding the mean value of motion vector module μ , given for each P-frame by

$$\mu = \frac{1}{MN - I} \sum_{i=0}^{M-1} \sum_{j=0}^{N-1} \sqrt{v_x^2(i, j) + v_y^2(i, j)} \quad (1)$$

$$\mu < S_{\text{no-motion}} \quad (2)$$

where M and N are the frame dimensions (in MacroBlocks), I is the number of Intra-Coded MacroBlocks, v_x and v_y are the horizontal and vertical components of motion vectors and $S_{\text{no-motion}}$ is the threshold value.

We have then evaluated the correlation between this parameter and semantic events for various threshold values, and we have found that with the value of $S_{\text{no-motion}} = 4$ we are able to detect 65 over 92 semantic events with 60 false detections (requiring the presence of lack of motion for at least 3 P-frames) (see Fig. 1). This arises from the fact that qualitatively we have found lack of motion before events at the beginning of game actions or after events at the end of game actions.

Camera motion parameters, represented by horizontal "pan" and "zoom" factors, have been evaluated using a least-mean square method applied to P-frame motion fields [7]. We have detected fast horizontal pan (or fast zoom) by thresholding the pan value (or the zoom factor), using the threshold value S_{pan} (or S_{zoom}).

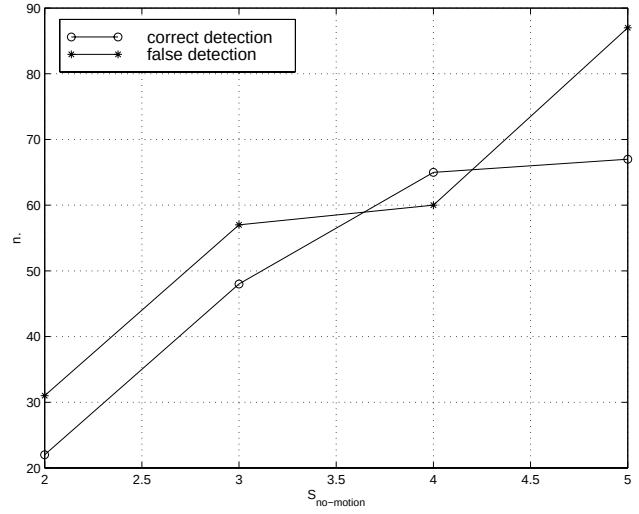


Figure 1: Number of significant events detected (over a total of 92 significant events) and false detections, by setting various values to $S_{\text{no-motion}}$.

Exploiting the correlation between pan and zoom factors and semantic events, we have noticed that we can find fast pan in correspondence with shots toward the goalkeeper or fast ball exchanges, and fast zoom in correspondence with interesting situations according to the perception of camera operator [4]. By requiring the presence of fast pan OR fast zoom for at least 3 P-frames, we are able to detect 53 over 103 semantic events with 67 false detections (see Figures 2, 3).

3. SHOT-CUT DETECTION

In our implementation, shot-cuts have been detected using only motion information too. In particular, we have used the sharp variation of the above mentioned motion indices and of the number of Intra-Coded Macroblocks of P-frames [8].

To evaluate the sharp variation of the motion field we have used the difference between the average value of the motion vectors modules between two adjacent P-frames. This measure is given by

$$\Delta\mu(k) = \mu(k) - \mu(k-1) \quad (3)$$

where $\mu(k)$ is the average value of the motion vectors modules of the P-frame k , given by Eq. 1.

This parameter will assume significantly high values in presence of a shot-cut characterized by an abrupt change in the motion field between the two considered shots.

This information regarding the sharp change in the motion field have been suitably combined with the number of Intra-Coded MacroBlocks of the current P-frames, as follow

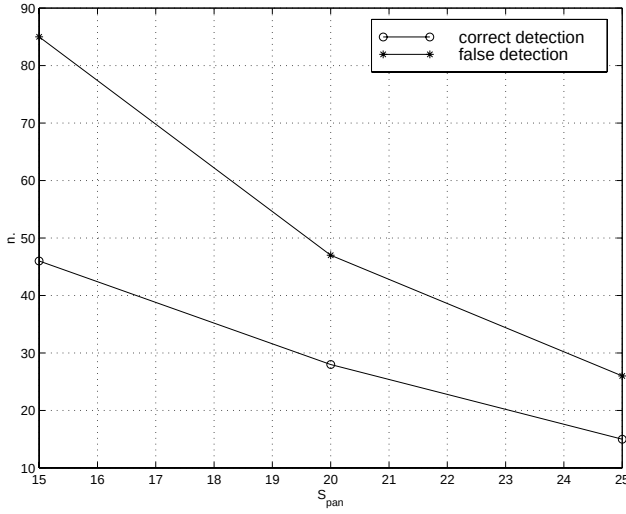


Figure 2: Number of significant events detected (over a total of 103 significant events) and false detections, by setting various values to S_{pan} .

$$\text{Cut}(k) = \text{Intra}(k) + \beta \Delta \mu(k), \quad (4)$$

where $\text{Intra}(k)$ shows the number of the Intra-Coded MacroBlocks of the current P-frame, and β is a proper weighting factor.

When this parameter is greater than a prefixed threshold value S_{cut} , we assume there is a shot-cut. The performance of this methods with various values of β and S_{cut} are presented in Figures 4, 5.

4. THE PROPOSED GOAL-FINDING ALGORITHM

As it can be seen, the above mentioned low-level indices are not sufficient, individually, to reach satisfying results (all the semantic events detected with few false detection). To find particular events, such as, for example, goals or shot toward goal, we have tried to exploit the temporal evolution of motion indices in correspondence with such events. We have noticed that in correspondence with goals we can find fast pan or zoom followed by lack of motion, followed by a shot cut. The concatenation of these low-level events have therefore been detected and exploited using the finite-state machine shown in Fig. 6.

From the initial state SI, the machine goes into state S1 if fast pan or fast zoom is detected for at least 3 consecutive P-frames. Then, from state S1 the machine goes into final state SF, where a goal is detected, if a shot-cut is present; from state S1 it goes into state S2 if lack of motion is detected for at least 3 consecutive P-frames. From state S2 the machine goes into final state SF if a shot-cut is detected, while it returns into state S1 if fast pan or zoom is detected

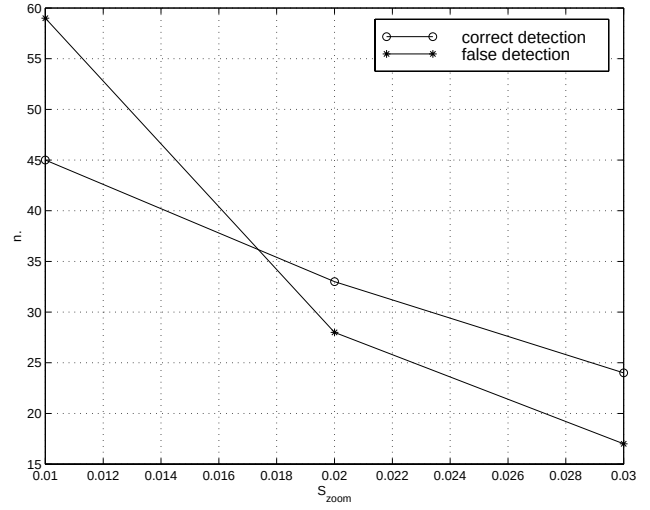


Figure 3: Number of significant events detected (over a total of 103 significant events) and false detections, by setting various values to S_{zoom} .

Events	Present			Detected		
	Live	Replay	Total	Live	Replay	Total
Goals	20	14	34	18	7	25
Shots to goal	21	12	33	8	3	11
Penalties	6	1	7	1	0	1
Total	47	27	74	27	10	37
False	116					

Table 1: Performance of the proposed goal-finding algorithm.

for at least 3 consecutive P-frames (in this case the game action is probably still going on). Two "timeouts" are used to return into initial state SI from states S1 and S2 in case nothing is happening for a certain number of P-frames (corresponding to about 1 minute of sequence).

5. EXPERIMENTAL RESULTS

The performance of the proposed goal-finding algorithm have been tested on 2 hours of MPEG2 sequences containing the semantic events reported in Table 1. Almost all live goals are detected, and the algorithm is able to detect some shots to goal too, while it gives poor results on free kicks. The number of false detection is quite relevant, but we have to take into account that these results are obtained using motion information only, so these false detection will probably be eliminated using other type of media information (e.g., audio information).

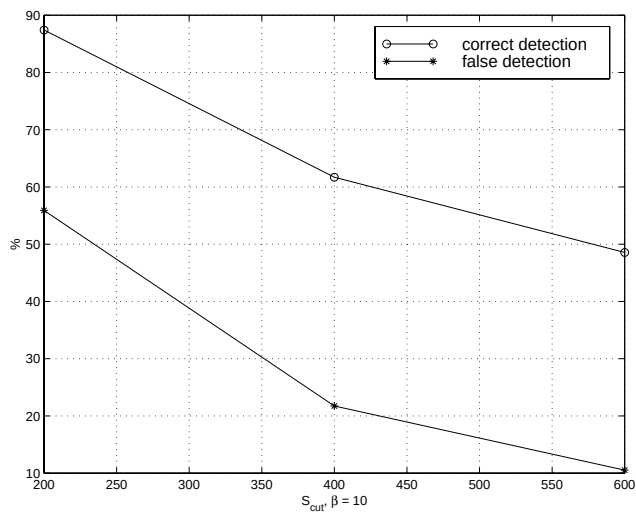


Figure 4: Percentage of shot-cuts detected (over a total of 177 significant events) and false detections, by setting various values to S_{cut} and $\beta = 10$.

6. CONCLUSIONS

In this paper we have presented a semantic indexing algorithm based on finite-state machines and low-level motion indices extracted from MPEG compressed bit-stream. The obtained results are encouraging, particularly if we take into account the fact that we have used motion information only. Current research is devoted to the extension of the proposed algorithm to detect salient semantic events in other categories of video programmes.

7. REFERENCES

- [1] Yao Wang, Zhu Liu, Jin-cheng Huang, "Multimedia Content Analysis Using Audio and Visual Information", To appear in Signal Processing Magazine.
- [2] C. Saraceno, R. Leonardi, "Indexing Audio-Visual Databases Through a Joint Audio and Video Processing", International Journal of Imaging Systems and Technology, Vol. 9, No. 5, pp. 320-331, Oct. 1998.
- [3] R. Zhao, W.I. Grosky, "From Features to Semantics: Some preliminary Results", Proc. of IEEE International Conference ICME2000, New York, NY, USA, 30 July - 2 August 2000.
- [4] A. Bonzanini, R. Leonardi, P. Migliorati, "Semantic Video Indexing Using MPEG Motion Vectors", Proc. EUSIPCO'2000, pp. 147-150, 4-8 Sept. 2000, Tampere, Finland.

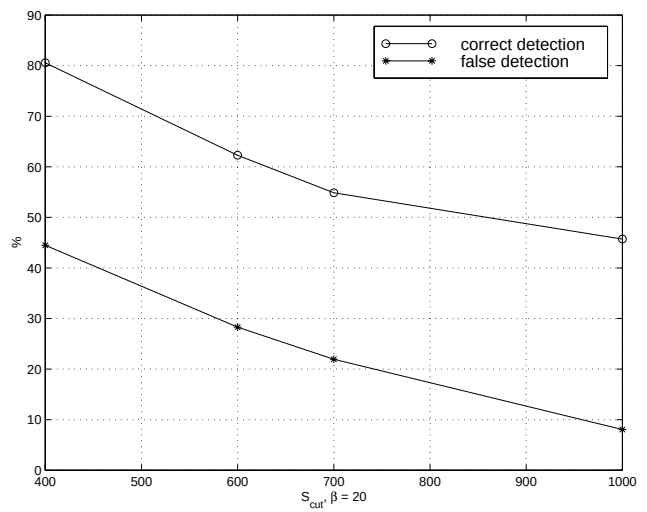


Figure 5: Percentage of shot-cuts detected (over a total of 177 significant events) and false detections, by setting various values to S_{cut} and $\beta = 20$.

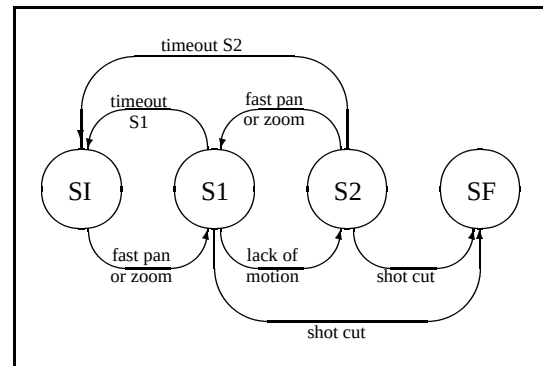


Figure 6: Goal-finding algorithm.

- [5] E. Ardizzone, M. La Cascia, "Video Indexing Using Optical Flow Field", Proc. ICIP 96, pp. 831-834, 16-19 September 1996, Lausanne, Switzerland.
- [6] W. A. C. Fernando, C. N. Canagarajah, D. R. Bull, "Video Segmentation and Classification for Content Based Storage and Retrieval Using Motion Vectors", Proc. SPIE Vol. 3656, pp. 687-698, 1999.
- [7] P. Migliorati, S. Tubaro, "Multistage Motion Estimation for Image Interpolation", Signal Processing: Image Communication, Vol. 7, pp. 187-199, July 1995.
- [8] Yining Deng, B. S. Manjunath, "Content-Based Search of Video Using Color, Texture, and Motion", Proc. of IEEE International Conference ICIP-97, Santa Barbara, California, USA, pp. 534-536, October 26-29, 1997.