SEMANTIC VIDEO INDEXING USING MPEG MOTION VECTORS

A. Bonzanini, R. Leonardi, P. Migliorati
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

With the diffusion of large video databases and "electronic program guides", the problem of semantic video indexing is of great interest. In literature we can found many video indexing algorithms, based on various types of low-level features, but the problem of semantic video indexing is less studied and surely it is a great challenging one. In this paper we present a particular semantic video indexing algorithm based on the motion information extracted from MPEG compressed bit-stream. This algorithm is an example of solution to the problem of finding a semantic event (scoring of a goal) in case of specific type of sequences (soccer video).

1 INTRODUCTION

With modern technologies we are able to store and transmit large quantities of audiovisual material, 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 [2] [1] useful for various applications, such as, for example, electronic program guides.

In Figure 1 is represented the fact that, while a man uses its cognitive skills to face the semantic indexing problem, an automatic system can face it in two steps [3]: in the first step some "low-level indexes" 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 low-level indexes.

While the problem of low-level indexes extraction is widely discussed in literature, the problem of finding decision-making algorithms is less studied; furthermore it is quite clear that this problem can be partially solved only for particular types of sequences.

In this work we have addressed the problem of semantic video indexing using the associated motion information. We have faced this problem trying to find a correlation between semantic features and motion indexes associated to a video sequence.

In literature many indexes suitable to be used for any type of video sequences can be found, no matter what the semantic content of the sequence is; on the contrary the problem of finding a correlation between these indexes and semantic content of video sequences can be solved only for a specific type of sequences. For this reason in this work we have considered the problem of semantic indexing of soccer video sequences; the standard format chosen for their representation is MPEG, which is widely used for storage and transmission of TV signals.

For soccer video sequences the semantic content can be identified with 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 motion indexes which represent the following characteristics: lack of motion and camera operations, represented by pan and zoom parameters. We have then studied the correlation between these indexes and the semantic events defined above, and we have found that camera motion is very meaningful. Exploiting this correlation, the proposed algorithm detects the presence of goals and relevant events.

The document is organized as follow. In Section 2 we will discuss the indexes chosen to represent these low-level motion information. In Section 3 the proposed algorithm is presented, while in Section 4 we will report some experimental results. Final conclusions are drawn in Section 5.

2 LOW-LEVEL MOTION INDEXES

We have used three motion indexes related to the following infomation: "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

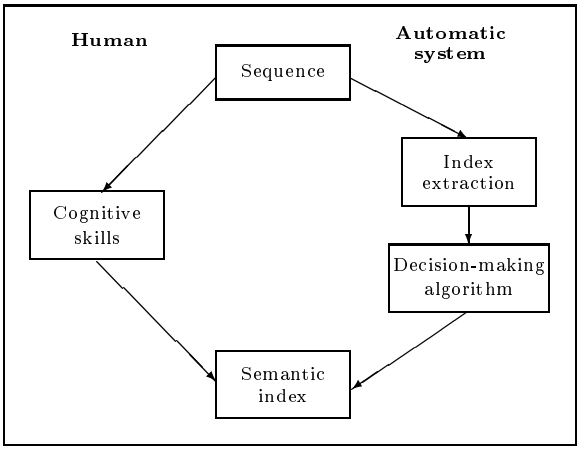

Figure 1: Semantic video indexing problem solution.

each P-frame by:

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

$$\mu < S_{\mathrm{no-motion}} \qquad (2)$$

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 $S_{\mathrm{no-motion}}$ is the threshold value. We have evaluated the correlation between this parameter and semantic events for various threshold values, and we have found that with the value of $S_{\mathrm{no-motion}} = 4$ we are able to detect 65 over 92 semantic events with 60 false detections (asking the presence of lack of motion for at least 3 P-frames). 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 "pan" and "zoom" factors [5], have been evaluated using a least-mean square method applied to P-frame motion fields [6]. We have detected fast pan (or fast zoom) by thresholding the pan factor (or the zoom factor), using the threshold value S_{pan} (or S_{zoom}):

$$\mathrm{pan} > S_{\mathrm{pan}} \qquad (3)$$

$$\mathrm{zoom} > S_{\mathrm{zoom}} \qquad (4)$$

Exploiting the correlation between pan and zoom factors and semantic events, we have noticed that we can find fast pan in correspondence with goal shots or fast passes, and fast zoom in correspondence with interesting situations underlined by the camera operator. Asking 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.

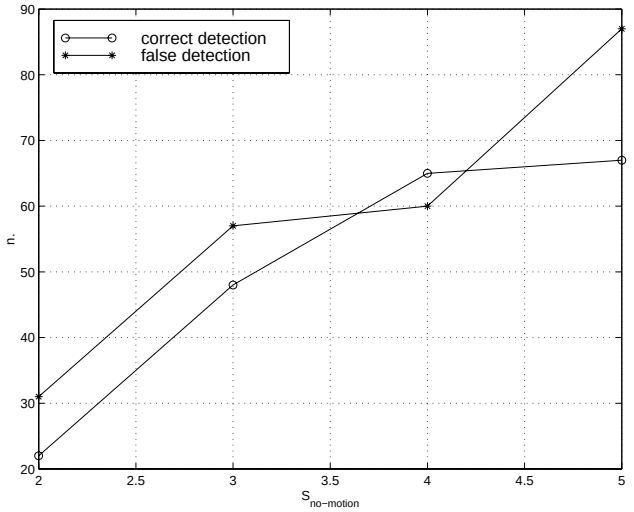

Figure 2: Number of significant events detected (total number: 92) and false detections, with various values of $S_{\mathrm{no-motion}}$.

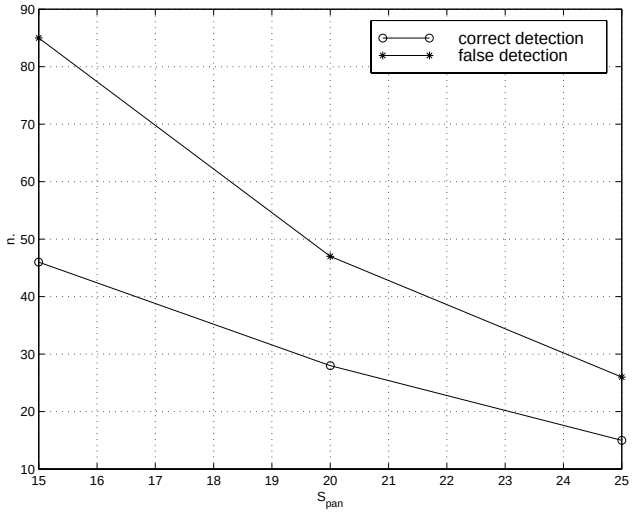

Figure 3: Number of significant events detected (total number: 103) and false detections, with various values of S_{pan}.

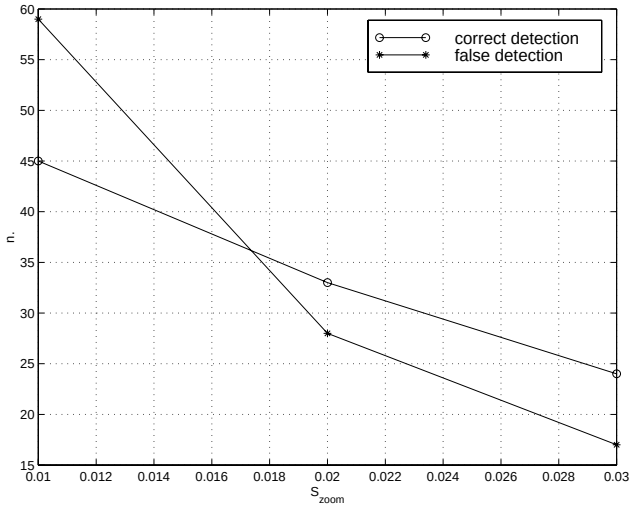

Figure 4: Number of significant events detected (total number: 103) and false detections, with various values of S_{zoom}.

3 THE GOAL-FINDING ALGORITHM

The motion indexes above mentioned are not sufficient, individually, to reach satisfying results (all semantic events detected with few false detection). To find particular events, such as, for example, goals or shot to goals, we have tried to exploit the temporal evolution of motion indexes 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 sequence of this low level events have therefore been detected and exploited.

Shot cuts have been detected using only motion information too; in particular, we have used the sharp variation of the above mentioned motion indexes and the number of intra-coded macroblocks of P-frames [7] [4].

4 EXPERIMENTAL RESULTS

We have tested the performance of the proposed goal-finding algorithm on 2 hours of MPEG2 sequences containing the semantic events reported on table 1, where is reported the number of detected events too. Almost all live goals are detected, and the algorithm is able to detect some shots to goal too, while it gives poor results on penalties. 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 information (typically audio information).

5 CONCLUSIONS

In this paper we have presented a semantic indexing algorithm based on motion information extracted directly from MPEG bitstream. We have presented a particular algorithm which exploits the temporal evolution of some low-level motion indexes, together with its performance when applied to some MPEG2 test sequences. The obtained results are encouraging, particularly if we take into account the fact that the problem is very difficult and that we have used motion information only.

Current research are devoted to the extension of the proposed idea to other types of sequences.

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.

References

[1] C. Saraceno and R. Leonardi, "Indexing audio-visual databases through a joint audio and video processing", International Journal of Imaging Systems and Technology, 9(5):320-331, Oct. 1998.

[2] Yao Wang, Zhu Liu, and Jin-cheng Huang, "Multimedia Content Analysis Using Audio and Visual Information", To appear in Signal Processing Magazine.

[3] Reginald Lagendijk, "A Position Statement for Panel 1: Image Retrieval", Proc. of the VLBV99, Kyoto, Japan, pp. 14-15, October 29-30, 1999.

[4] Thomas Sikora, "MPEG Digital Video-Coding Standards", IEEE Signal Processing Magazine, September 1997, vol. 14, n. 5.

[5] G Avid, "Determining tree-dimensional motion and structure from optical flow generated by several moving object", IEEE Trans. on PAMI, PAMI-7:384-401, July 1985.

[6] P. Migliorati and S. Tubaro, "Multistage motion estimation for image interpolation", Image Communication, July 1995, pp. 187-199.

[7] Yining Deng and B. S. Manjunath, "Content-based search of video using color, texture, and motion", Proc. of IEEE International Conference on Image Processing ICIP-97, vol. 2, Santa Barbara, California, pp. 534-536, October 26-29, 1997.