DETECTION OF PAN AND ZOOM IN SOCCER SEQUENCES BASED ON H.264/AVC MOTION INFORMATION

Luca Superiori, Markus Rupp

Institute of Communications and Radio-Frequency Engineering
Vienna University of Technology
Gusshausstrasse 25/389, A-1040 Vienna, Austria
Email: {lsuper,mrupp}@nt.tuwien.ac.at

ABSTRACT

Unsupervised detection of pan and zoom in soccer sequences allows automatic classification of shots and match analysis. In this work we propose a pan and zoom (both in and out) detector specifically designed for low resolution soccer sequences. Our implementation is based on the analysis of the distribution of the motion vectors, already available in the encoded sequence, among a specific subset of reliable MBs, selected by means of inexpensive image preprocessing.

1. INTRODUCTION

In the last years new digital transmission systems, such as DVB-T and IPTV for domestic use and DVB-H for nomadic use, increased considerably the amount of video contents offered to the customers. New internet services, video portals such as YouTube, allow the users to upload and share their home made contents.

The scarcity of the available bandwidth calls for increasing efficiency of the video encoding. In 2004, the ITU (International Telecommunication Unit) and the MPEG (Moving Picture Expert Group) jointly standardized the H.264/AVC [1] (Advanced Video Coding). As its precursors, both from ITU-T and MPEG, the H.264/AVC is a hybrid block-based video encoder. This family of codecs exploits the correlation of small blocks, called macroblocks (MB) of a picture both in space, considering the neighbouring blocks belonging to the same picture, and in time, considering blocks belonging to the previous pictures.

Due to the variety of the offer, automatic video indexing and analysis tools are increasing in importance. This work deals with the automatic detection of pan and zoom in low resolution soccer video sequences, exploiting the information already available after the encoding.

Pan refers to the rotation of the camera on its horizontal plane. In soccer videos the camera usually shoots the area of the field where the ball is present. In wide angle shots, being the shot objects distant from the camera, this camera movement is appreciated as a rigid camera translation. The zoom is the camera feature of simulating movement away from (zoom out) or toward (zoom in) a subject. In professional cameras, this action is performed varying the distance between the lenses. In soccer videos, *zoom in* is used to highlight the field region where the action is taking place. *Zoom out* is used in case the ball moves rapidly and a wider region of interest has to be considered.

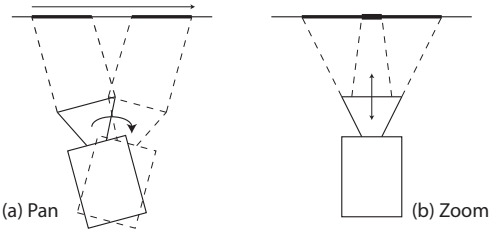


Fig. 1. Camera pan and zoom

Soccer one of the most preferred content, it has, therefore, become the focus of different scientific researches. In this article we discuss the possibility of inferring the mentioned camera information for this specific content. It will be shown how, differently from other scenarios such as panorama sequences, the MBs of a soccer frame do not share the same direction. A simple video processing algorithm is proposed to discriminate the MBs that contribute to the definition of the camera pan and zoom. Our approach relies on the information already stored in the encoded sequence, in particular on the *motion vectors*, indicating the offset between the current block and its best prediction in the previous pictures.

This article is structured as follows. Section 2 presents similar works in literature and highlights the contributions of the present work. Section 3 describes the implementation of proposed approach. Qualitative results are discussed in Section 4. The conclusion are drawn in Section 5.

The authors thank mobilkom austria AG for technical and financial support of this work. The views expressed in this paper are those of the authors and do not necessarily reflect the views within mobilkom austria AG.

2. CHARACTERISTICS OF SOCCER SEQUENCES

Several works in literature address the detection of pan and zoom in video sequences. Most of them exploit the concept of global motion, as defined in [2]. The contributions in [3] and [4] determine the global motion by accurate motion models, involving complex processing such as gradient descent and iterative least square estimation. In this work we exploit the motion information already contained in the motion vectors. This approach was already used in [5] and [6]. In [6] a probabilistic model is designed to determine the averaged probability of zoom in a frame. Dumitras and Haskell [5] first address the necessity to discriminate the MBs contributing to the estimation of the global motion. In their approach, they weight the contribution from the two biggest regions of the frame, where the motion vectors share similar orientation.

We propose a specific model suitable for soccer sequences. The global motion and the zoom are estimated exploiting the direction of the motion vectors, considering the contribution of a set of reliable MBs.

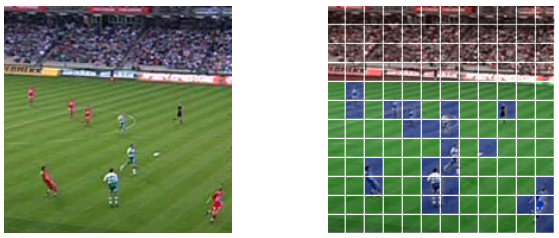
Firstly, we subdivide a soccer frame into three regions:

Field (R0) It mostly consists of low-frequency green pattern. The video encoder chooses the best temporal prediction minimizing a cost function, weighting the size of the encoded MB with respect to the introduced distortion. Because of the lack of details, the best temporal prediction may be offered by a similar block placed in the vicinity of the considered one. The cost is often minimized by a zero motion compensation, where the block is copied from the previous picture.

Player and ball (R1) The players and the ball do not move consistently with the camera movement. The motion vectors associated to their MBs are almost uncorrelated. They do not offer a reliable source of global motion estimation.

Grandstands (R2) The audience represents an almost static background of the soccer videos. In low resolution soccer videos, the MBs containing the audience consist mainly of high frequency patterns. Due to the property of the audience to remain static, the motion vectors selected by the encoder share almost the same horizontal and vertical direction.

In order to detect the MBs belonging to the audience, we proposed in [8] a fully unsupervised segmentation algorithm able to subdivide the soccer frame into the three before mentioned regions. After transforming each frame in the HSV components, the color characteristics of the soccer frame are exploited. A region growing algorithm has been designed to detect the MBs containing the audience. After placing the seeds on the four corners of the frames, new MBs are added to the R2 if their green pixels quota does not exceed a given threshold. The remaining MBs contain the field, the player and the ball. The region R0 and R1 are discriminated with respect to the green pixels quota of the MBs.



(a) Original frame (b) Segmented frame

Fig. 2. Segmentation algorithm

3. DETECTION OF PAN AND ZOOM

Soccer video sequences are characterized by wide angle camera shots, pointing the area of the field where the action is taking place. Usually the camera follows the action, rotating on its horizontal plane. As discussed in Section 2, non static objects do not move consistently with the camera. The MBs with lack of details are subject to a motion compensation that is not necessarily in accordance with the camera movement.

It has been noted that the MBs containing the grandstands offer a reliable source of prediction for the global motion characterizing the shot. In order to demonstrate this, the distribution of the motion vectors in the three regions defined in Section 2 has been analyzed. The *absolute* motion vector applied to each MB of the picture has been saved during the encoding and examined offline. In Fig. 3 the direction of each

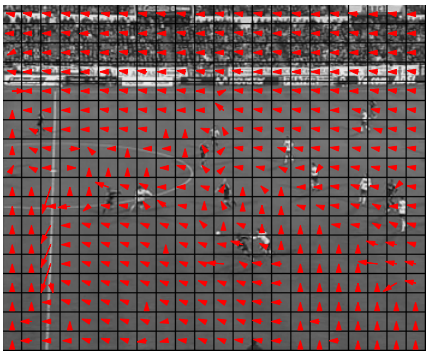


Fig. 3. Distribution of the Horizontal MV

motion vector applied to the picture has been drawn. It can be noticed that the MBs containing the players, the ball and the field lines does not help discriminating a dominant global direction. Most of the MBs containing the field are pointing in two different main directions. Only the MBs containing the audience are pointing a single dominant direction.

The histograms in Fig. 4 show the distribution of the horizontal motion vector in the range $[-16, 16]\times$ quarter of a pixel. Indeed, H.264/AVC apply motion compensation with the resolution of a quarter of a pixel, by means of weighted interpolation. The two dominant directions in R0 as well as the absence of a dominant one for the R1 can be noticed in

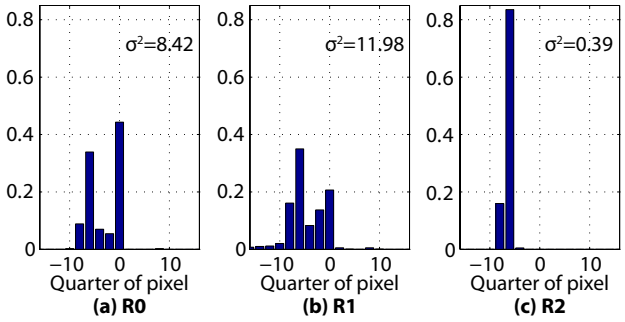


Fig. 4. Distribution of the Horizontal MV

the histograms 4(a) and 4(b), respectively. A global motion component can be recognized for the R2 (Fig. 4(c)): 99.5% of the motion vectors are comprised within four quarters of a pixel.

To calculate the distributions variance and mean, the outliers of the motion vectors distribution has been removed by means of statistical data pruning. In a distribution, the outliers are elements lying at *abnormal* distance from the rest of the data. In the considered case, outliers can be caused by motion compensation of the MBs at the border of the picture. New elements appearing at the border, because of the camera movement, do not have any correspondent block in the previous pictures. Their associated motion vectors might not be consistent with the global motion. To discriminate outliers, an implementation of the Grubbs' test [7] has been used.

Once identified the global motion, the same set of audience motion vectors distribution has been used to perform the zoom detection. It has been noticed that the variance of the audience motion vectors was strongly varying over time. In frames where the variance was particularly high, the motion vectors were distributed as in Figure 5. For each picture row,

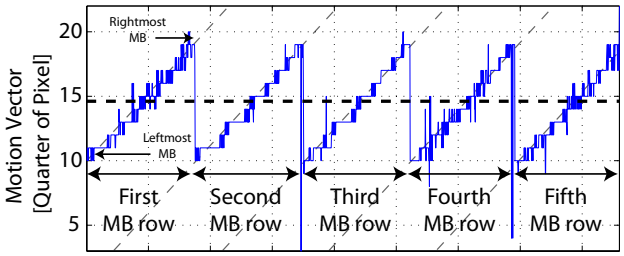


Fig. 5. Amplitude of the horizontal motion vectors

the motion vectors associated to each MB from left to right has been drawn. Since a MB consists of 16 subMBs, each MB is represented by 16 components.

In the considered frame, the audience occupied the whole first five rows of the frame. Being CIF (Common Intermediate Format) the resolution of the frame, we count 22 MBs per row. H.264/AVC allows the subdivision of a MB up to 4×4 pixels block. This results on a maximum number of 16 motion vectors per MB. Independently from the selected subdivision, we decided to consider the motion compensation applied to each 4×4 pixel block, to avoid weighting issues. The motion vectors in Fig. 5 vary linearly their amplitude from the left-most to the right-most MB of each row. The slope of the line remains constant in each row, as indicated by the gray broken lines. Moreover, the mean of each row corresponds to the global motion, indicated by the bold horizontal broken line.

The amplitude of the motion vectors on the left-most MBs are slightly smaller than the average, whereas the one on the right-most MBs are slightly bigger. Assuming a global motion equal to zero, then the motion vectors in the left side of the picture would be negative (pointing to the left) and the one on the right side positive (pointing to the right). Equivalently, after removing the global motion component, the motion vectors of the audience assume the described directions shown in Figure 6.

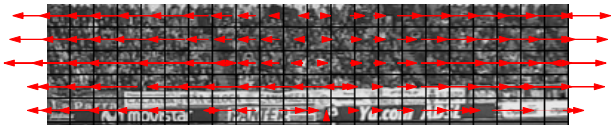


Fig. 6. AC component of the audience MVs

Observing its time evolution, such a behavior corresponds to a *zoom out* action: objects in the border of the picture tend to move to the center of the picture. The reciprocal effect has been observed as well. For negative slopes, the MBs in the border of the picture, point inner MBs of the previous picture. This origins *zoom in*. A comparison between the two effects can be observed in Fig. 7

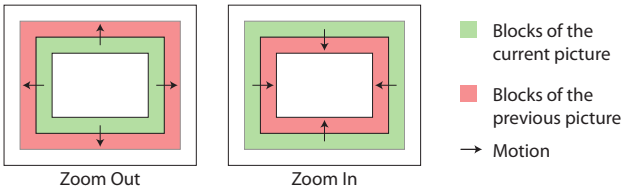


Fig. 7. Zoom operations

4. RESULTS

In order to prove the effectiveness of the method, we tested our algorithm on 50 soccer sequences extracted from a match of the first Spanish league. The original high resolution MPEG2 video has been deinterlaced and spatially downsampled to the CIF resolution (352×288 pixel). The sequences have been encoded using a modified version of the standard developers codec JM (Joint Model) [9] version 14.0 [9]. As additional output, the motion vectors of all 4×4 picture subblocks have been stored.

Concerning the pan, the speed of the movement has been measured observing the detected global motion, as shown in Fig. 8. Positive horizontal motion vectors correspond to camera movement to the right direction, while negative to the left. In a similar manner, positive vertical motion vectors indicate downwards camera movement while negative upwards movement. No relevant correlation was measured between the two

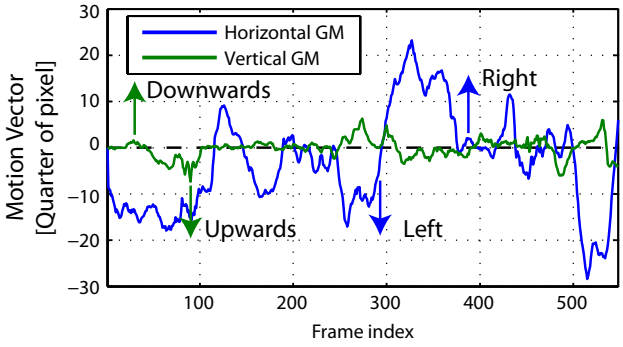


Fig. 8. Horizontal and vertical pan

considered movements. As expected, the amplitude of the horizontal movement is much more relevant than the vertical one.

In order to measure the presence of zoom, only the horizontal movement of the camera has been considered: the width of the considered area (the grandstands) is much wider than its height and, as shown before, the horizontal component is much more significant. The trend of the slope of the motion vectors distribution, as depicted in Fig. 5, has been plotted in Fig. 9.

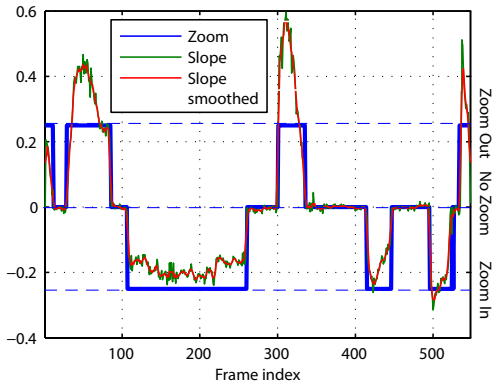


Fig. 9. Zoom Detection

A smoothed slope curve has been drawn to compensate rough slope variations, particularly around the zero, using a moving window of length five. The trend of the graph describes three main behaviors according to Section 3:

1. Slopes around zero: no zoom has been detected in the considered frame.
2. Slopes bigger than zero: zoom out has been detected.
3. Slopes smaller than zero: zoom in has been detected.

The transition between the three regions is abrupt, proofing the robustness of the detector. As a last step, a hard decision block has been implemented. A fixed threshold on the smoothed sloped of 0.02 has been found as a good compromise between detection capabilities and robustness against false detection.

5. CONCLUSION

In this paper we have presented a detector of camera zoom and pan based on the evaluation of the motion characteristics obtained by the video encoder. The method has been specifically designed for soccer video sequences. In order to detect the sequence global movement (pan), only the MB belonging to the audience have been considered. The zoom is detected analyzing the relative size of the motion vector of a MB row, with respect to the recognized pan.

As further improvement, the information about pan and zoom can be combined and converted to an absolute movement measure. This would allow methods for automatic shot classification and unsupervised soccer match analysis.

6. REFERENCES

[1] ITU-T Rec. H.264 / ISO/IEC 11496-10, "Advanced Video Coding," Final Committee Draft, Document JVTE022, Sept. 2002.

[2] C. Stiller, J. Konrad, "Estimating motion in image sequences - a tutorial on modeling and computation of 2D motion," IEEE Signal Processing Magazine, pp. 70-91, July 1998.

[3] F. Dufaux, J. Konrad, "Efficient, robust and fast global motion estimation for video coding," IEEE Trans. on Image Processing, vol. 9, no. 3, pp. 497-501, Mar. 2000.

[4] G. B. Rath, A. Makur, "Iterative least squares and compression based estimations for a four-parameter linear global motion model and global motion compensation," IEEE Trans. on Circuits and Systems for Video Tech., vol. 9, no. 7, pp. 1075-1099, Oct. 1999.

[5] A. Dumitras, B.G. Haskell, "A look-ahead method for pan and zoom detection in video sequences using block-based motion vectors in polar coordinates," in Proc. of ISCAS 2004, Vancouver, Canada, May 2004.

[6] R. Jin, Y. Qi, A. Hauptmann, "A probabilistic model for camera zoom detection" in Proc. of 16th ICPR, Quebec, Canada, Aug. 2002.

[7] F. Grubbs, "Procedures for Detecting Outlying Observations in Samples," Technometrics, 11/1, 1-21, 1969.

[8] L. Superiori and M. Rupp, "Encoding Optimization of Low Resolution Soccer Video Sequences," International Conference on Multimedia and Expo 2008, Hannover, Jun. 2008.

[9] H.264/AVC Software Coordination, "Joint Model Software," ver.14.0, available in `http://iphome.hhi.de/suehring/tml/`.