Formula 1 Onboard Camera Shot Detector Using Motion Activity Areas

Arnau Raventós, Francesc Tarrés
EETAC, UPC-BARCELONATECH, Spain
arnau.raventos@estudiant.upc.edu
francesc.tarres@upc.edu

Abstract

Shot detection in sports video sequences has been of great interest in the last years. In this paper, a new approach to detect onboard camera shots in compressed Formula 1 video sequences is presented. To that end, and after studying the characteristics of the shot, a technique based in the thresholding comparison between a high motion area and a stationary one has been devised. Efficient computation is achieved by direct decoding of the motion vectors in the MPEG stream. The shot detection process is done through a frame-by-frame hysteresis thresholding analysis. In order to enhance the results, a SVD shot boundary detector is applied. Promising results are presented that show the validity of the approach.

1. Introduction

The production of digital video sequences is increasing everyday and there is a need to provide tools to automate and facilitate its management. In particular, sports video sequences take an important role in multimedia content as a large variety of sports are played getting the interest of many people. However, sports video sequences tend to be lengthy and once the main live event ends people are usually only interested in the most significant parts of it. That is one of the main reasons why automatic summarization has been an important topic of study over the last two decades [1][2][3][4].

The initial step of the summarization process is video shot segmentation. This is done by detecting shot boundaries, which is based on identifying significant changes of the visual content in the neighbor frames of a video sequence. Shot boundary detection is a complex task in sport events [1] due to the resemblance of their shots, the randomness of their motion statistics and the mix of abrupt cuts and gradual transitions. A lot of techniques for shot detection [3][5][6][7] have been presented in the literature and basically can be organized in two categories: genre-independent or genre-specific.

This paper focuses on a specific sport genre, Formula 1 races, and particularly in the detection of onboard camera (OC) shots. This type of shot presents the view of a fixed camera on the Formula 1 car as can be seen in Figure 1. It is a characteristic shot very often used for Formula 1 video sequences summarization.

Figure 1. Onboard camera shot

The paper is organized as follows: Section 2 provides a general overview of related work in this particular problem while section 3 gives a perspective of the approach followed in this paper. The details of the strategy for onboard camera event detection are discussed in section 4 while a refinement method is further presented in the section 5. Results comparing the performance of the system against a groundtruth obtained in real video sequences are presented in section 6. Finally, conclusions and references are given.

2. Related Work

Shot detection using motion analysis [6] by means of the motion compensation vectors encoded in a MPEG stream has been intensively studied due to its appealing low computational cost. In the F1 event detection

978-1-4673-2369-7/12/$31.00 ©2012 IEEE

context, there have been a very few works dealing with the onboard camera shot detection. G. Tardini, et al. [6] have presented a novel shot boundary detector that first splits the video sequence in shots. Then, they characterize the motion of each shot dividing its frames in 4 quadrants and for each one 3 motion characteristics based on the average motion vector are computed. Looking at these characteristics, two types of shots are defined, one of these being the OC shot. They obtained 88.4% recall and 80.9% precision.

Our approach focuses only on the OC shot and thus we define two specific areas selected based on its motion statistics. In addition, as our objective is to measure the amount of motion, the only motion characteristic that we take into account is the magnitude of the motion vectors. We first compute and analyze the motion characteristics, and highlight the possible locations in time of the OC shots. Then a shot boundary detector that refines the exact position of each shot in time is applied.

3. Proposed Approach

The proposed approach is divided in two phases: The onboard camera shot detection phase and the refinement phase. It is assumed that, in this type of shot, the statistics of the pixels in the nose of the car are stationary while in the background of the image they change fast. These assumptions match very well the onboard camera characteristics encountered in Formula 1 video sequences. As a result, stationary and high motion regions are defined. The decision process for detecting the shot starts with the frame-by-frame analysis of the stationary areas motion activity through a hysteresis thresholding. Then a threshold is applied to select high motion areas. The outcome of these processes generates OC shot candidates. Then, two refinement methods are applied to produce the final shot annotations. The first refinement method is a shot boundary detector [7] based on a Singular Value Decomposition (SVD) approach, which is implemented to increase the precision of each shot. The second refinement method checks if two adjacent shots are separated in time with an interval less than a specified threshold. If that is the case, the two shots are merged into one shot. The approach is summarized in the following scheme depicted in the Figure 2.

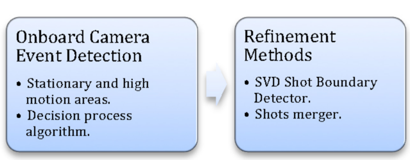

Figure 2. Proposed approach scheme

4. Onboard Camera Event Detection

4.1. Stationary and High Motion Areas

The basic approach underlying our OC shot detection relies in finding the stationary and the high motion areas. In an OC shot the recording camera is fixed on the top of the car, therefore the information about the nose of the car can be assumed constant along the whole shot, so the resulting motion vectors in that area tend to be zero. On the contrary, the background of the scene tends to change fast due to the high speed of the car, thus the motion vectors in that region present high motion activity. Having these two ideas in mind, we chose a rectangular shape at the top of the image to enclose the high motion area and a trapezoid shape for the stationary area. The areas are represented in Figure 3, where the red arrows denote the positive motion vectors and the blue arrows express the zero motion vectors or the macro blocks that does not benefit from motion compensation.

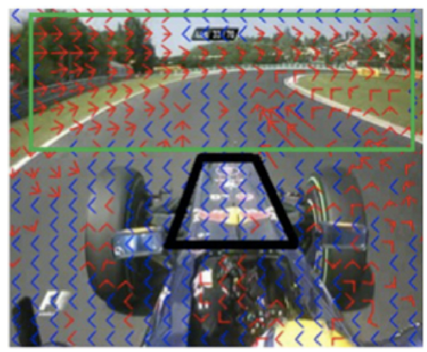

Figure 3. High motion and stationary areas

Since most of the Formula 1 video sequences are MPEG compressed, we take advantage of the available motion vectors of Intra (I) and Predicted (P) frames. As motion vectors are encoded only in P frames, I frames are discarded for the analysis and it is assumed that the video sequence is a continuous stream of P frames. For each region, an average magnitude of its motion vectors is obtained. This is what we denote as motion activity of each region and is calculated as follows:

$$A^{m}_{avg} = \frac{1}{N}\sum_{i=1}^{N}\sqrt{v_{x,i}^{2} + v_{y,i}^{2}} \qquad (1)$$

where N is the number of motion vectors within each area and v_x and v_y are its horizontal and vertical components.

4.2. Decision Process Algorithm

The decision process algorithm finds if the motion activity of the stationary area is below a specified threshold T1. When that occurs a counter starts annotating the number of consecutive frames whose motion activities are below a threshold T2. This thresholding technique is applied to force the algorithm to start with a low value. Thus the hysteresis provides a security margin so that values can oscillate around the first threshold T1. This process is illustrated in Figure 4. Once a set of consecutive frames fulfills the stationary area conditions, another process checks if the motion activity of the high motion area surpasses a predefined threshold T3. This last procedure ensures that the set of frames do not correspond to a full static shot. The final test performed verifies if the length of the set is larger than a threshold T4. At the end, if all the conditions are satisfied, the set of frames is labeled as an OC shot candidate.
All thresholds have been decided empirically. But they have been proven on three randomly chosen F1 races with different luminance environments and the achieved results are very similar.

5. Refinement Methods

In the last section OC shots are annotated. However the precision of its input and output time codes may be improved in some cases. To that end, the following two refinement methods have been applied.

5.1. SVD Shot Boundary Detector

To enhance the time codes of the previously computed OC shots a shot boundary detector has been implemented. As the majority of the boundaries of the OC shots present gradual transitions, a shot boundary detector that performs well in these cases has been used. The chosen detection technique has been proposed by Abd-Almageed, W. [7] and is based on the SVD analysis of a frame sequence. The technique consists of evaluating the rank of a matrix $\mathbf{X}$. This matrix is defined as a window of L frames filled with the Hue Saturation Value (HSV) histograms. The SVD is applied to diagonalize the matrix in $\mathbf{X} = \mathbf{U}\Sigma\mathbf{V}^H$.

Then the eigenvalues are extracted from Σ and each eigenvalue is normalized with respect to the largest one. Finally, the rank of the matrix $\mathbf{X}$ is computed annotating the number of eigenvalues that surpass a predefined threshold.

$$\mathbf{X} = \left.\begin{pmatrix} \mathbf{h}_H^1 & \mathbf{h}_S^1 & \mathbf{h}_V^1 \\ \mathbf{h}_H^i & \mathbf{h}_S^i & \mathbf{h}_V^i \\ \mathbf{h}_H^L & \mathbf{h}_S^L & \mathbf{h}_V^L \end{pmatrix}\right\} L \qquad (2)$$

The main idea of this shot boundary detector is that ideally the rank of the matrix $\mathbf{X}$ will remain 1 within the same shot, will form a pulse shape along time when there is an abrupt transition and will from a triangle shape along time when there is a gradual transition.
This shot boundary detector has been proved with soccer and Formula 1 video sequences providing very good results. See [7] for details.

5.2. Shots Merger

This method checks if two shots are separated a short period of time and may be interpreted as a filter of false OC shots detection. If two OC shots are very close in time this filter will merge them. This simple technique is applied to join shots that have been split due to an abrupt and fast change in the video sequence, as, for example, when the F1 car goes under a bridge.

6. Results

Three complete Formula 1 GP races have been analyzed with the proposed approach and the results have been compared with a manually annotated groundtruth in Table 1. The F1 2010 Melbourne race is an example of a rainy weather event. The letter F is an abbreviation for the number of frames.

Table 1. Results of experiments

	Duration	True positives	True negatives	False positives
F1 2010 Budapest	2h 20m 0s (210001 F)	45 (24361 F)	3 (556 F)	6 (2709 F)
F1 2011 Abu Dhabi	2h 50m 50s (256251 F)	43 (21604 F)	2 (127 F)	3 (1935 F)
F1 2010 Melbourne	2h 42m 56s (244400 F)	44 (21981 F)	8 (2725 F)	7 (4325 F)

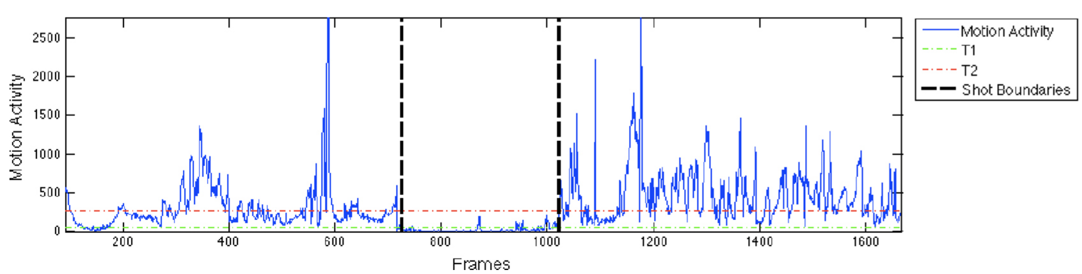

Figure 4. Motion activity in the stationary area

The results obtained have an average recall percentage of 91.9% of correctly detected OC shots. The detector has produced a number of false positives and a few number of shots have not been detected. On one hand the not detected shots can be classified in three categories: Very short shots that do not surpass the T4 threshold of the decision process algorithm, shots that due to unusual reflections in the nose of the car the algorithm does not detect an stationary area where it should be and rainy shots where the OC is covered by drops of water. On the other hand, false positives are usually OC shots, however not the desired OC shot where the camera is at the top-center of the car. In these examples, the majority of false positives correspond to OC shots where the camera is on the left side of the car or in the rear of the car. Examples of these shots are given in Figure 5.

Figure 5. False positives

We believe false positives could be eliminated with an algorithm comparing the edges of the images with a set of predefined patterns of each type of OC shot. This way the non-desired OC shots would be discarded. In the following Table 2, we compute the precision and recall with respect to the number of shots but also with respect to the number of frames.

Table 2. Precision and Recall

	Num. Shots		Num. Frames	
	P	R	P	R
F1 2010 Budapest	88.46	93.88	90.00	97.77
F1 2011 Abu Dhabi	93.48	95.55	91.78	99.41
F1 2010 Melbourne	84.61	86.27	83.55	88.97

As can be seen the recall increased significantly with respect to the number of frames, showing the fact that the not detected shots are short shots.

7. Conclusions

In this paper we propose a new method for detecting the onboard camera shots in Formula 1 Grand Prix video productions. The onboard camera event detection starts with the definition of stationary and high motion regions. Then, their motion activities are compared with predefined thresholds. These processes end up with the annotation of shot candidates, which afterwards are post-analyzed by two refinement methods. These methods seek to increase the precision of the shot boundaries in time.

It has been analyzed and evaluated that the motion statistics of the stationary area and the high motion area are very specific of this type of shots; therefore its assessment can provide promising results.

8. Acknowledgements

The Formula 1 Grand Prix video sequences used in this paper have been kindly provided by Televisió de Catalunya (TVC).

The paper has been partially supported by the Spanish administration agency CDTI under the project CENIT-Buscamedia (CEN-20091026) and the CICYT project TEC2011-22989 'PBInt'.

References

[1] A. Ekin, A. M. Tekalp, and R. Mehrotra, "Automatic soccer video analysis and summarization", *IEEE Transactions on Image Processing*, vol. 12, pp. 796-807, July 2003.

[2] R. Leonardi and P. Migliorati, "Semantic indexing of multimedia documents", *IEEE Multimedia,* vol. 9, no. 2, pp. 44–51, Apr.–June 2002.

[3] Yue Gao and Qiong-Hai Dai,"Shot-based similarity measure for content-based video summarization", *15th IEEE International Conference on Image Processing*, *ICIP 2008,* pp.2512-2515, October 12-15, 2008.

[4] Almeida, J., Torres, R.D.S., Leite, N.J. , "Rapid Video Summarization on Compressed Video," *IEEE International Symposium on Multimedia, ISM*, pp.113-120, December 13-15, 2010.

[5] A. Hanjalic, "Shot-boundary detection: Unraveled and resolved?" *IEEE Trans. Circuits Syst. Video Technol.,* vol. 12, no. 4, pp. 90–105, Apr. 2002.

[6] Tardini, G. Grana C., Marchi, R., Cucchiara, R., "Shot Detection and Motion Analysis for Automatic MPEG-7 Annotation of Sports Videos", *13th International Conference on Image Analysis and Processing, ICIAP 2005,* Cagliari, Italy, pp. 653-660, Sept 2005.

[7] Abd-Almageed, W., "Online, simultaneous shot boundary detection and key frame extraction for sports videos using rank tracing,", *15th IEEE International Conference on Image Processing*, *ICIP 2008*, San Diego, California, USA, pp.3200-3203, October 12-15, 2008.