An Algorithm for Highlights Identification and Summarization of Broadcast Soccer Videos

Waldez Azevedo Gomes Junior and Díbio Leandro Borges

Department of Computer Science, University of Brasilia,
70910-900 Brasília, DF, Brazil
waldezjr14@gmail.com, dibio@unb.br

Abstract. This paper presents an algorithm that aims to perform automatic summarization in broadcast soccer videos. The summarization considers identifying and keeping only the highlights of the match. A situation that could be considered a highlight is defined as one with high convergence of players to a spot. Our approach considers velocities and positions of the players, and an inferred movement of the TV camera as basic features to be extracted. The movement of the TV cameras are approximated using the movement of all the players in the image. A motion field is computed over the image in order to analyze aspects of the match. The algorithm was tested with real data of a soccer match and results are promising considering the approach uses as input broadcast videos only, and it has no a priori knowledge of cameras positions or other fixed parameters.

Keywords: Motion field analysis, soccer video analysis, video summarization.

1 Introduction

Broadcast soccer matches are probably the sport TV shows mostly viewed on the planet nowadays. Automatic analysis of soccer video sequences and complete matches has captured the interest of the Computer Vision research community since there are challenges in object, scene recognition and classification to be dealt with, and also a handful of needed applications to be developed and deployed. Automatic identification of highlights (e.g. situations leading to a goal or closer) and further summarization of the match based on those highlights are examples these challenges.

There has been a great deal of recent works in the literature considering automatic analysis of soccer videos. In [5] and [3] the authors consider tracking the players and the ball in order to study the detection of a highlight in the match. However, those works are based in a manual (i.e. prefixed) initialization of the position of the ball. The work done in [8] also approached the problem by tracking the players and it showed good results, but it assumes a fixed and controlled camera view .

The work reported in [6] has taken a different approach to this problem of highlights identification and summarization. It considers the relative position of the soccer field and the surrounding stadium in the video as main objects to be identified in the initial

L. Alvarez et al. (Eds.): CIARP 2012, LNCS 7441, pp. 862–869, 2012.

steps. Sequences to be considered as highlights are classified based on a relative classification of those regions in a frame. Results achieved there are interesting since different matches in complete different stadiums and transmission are tested. [4] an analysis of motion fields resulted from the movement of the players throughout a match is realized. The authors suggest that the movement of all the players may be a good indication of the position of the ball. Their work however aimed to have a dynamic scene analysis methodology to be used in situations where they would have complete knowledge and control of camera setups and positions. Contextual flow [9], a motion analysis methodology based on contextual matching is also a promising technique for target tracking that can be used for video summaries based on motion.

Our approach here considers a global analysis of the movement of the players. We consider that the players on the field have the best view of the game, therefore, analyzing their movement should provide a good indication of what is happening during the match. we make the hypothesis that when the movement of the players highly converges to one place on the field and the cameraman moves the camera with a higher speed than usual there is a relevant event of the match going on, and possibly a highlight. For summarizing a highlight we associate these detected frames as more likely sequences to be extracted. The algorithm starts with the identification of the players on the field, then tracking of the players, the construction of a dense velocity field based on the velocities of the players, the identification of the highlights, and the sorting of the highlight flags. In the next sections the details of those steps are presented followed by the results and evaluation of the algorithm.

2 Identification of the Players

In long shots of a soccer broadcast the prevailing color is the green color of the grass. In order to identify such prevailing color the image was divided in three channels and for each channel the histogram is calculated. With the histograms we obtain the most frequent color for each channel $R_{peak}, G_{peak}, B_{peak}$. Before obtaining such values the image is normalized to decrease the effect of light on the processing.

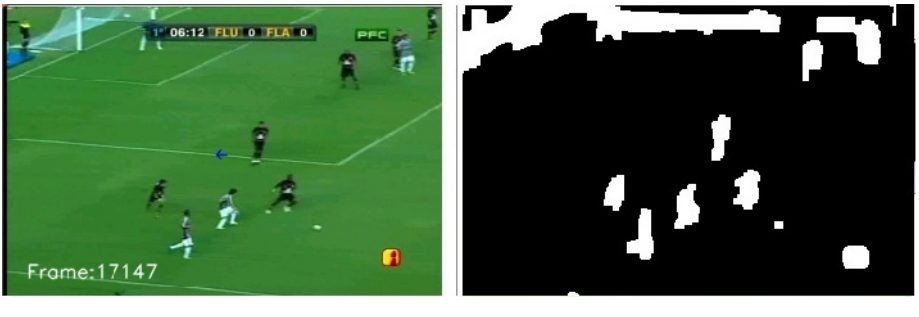

Fig. 1. The input image at left and segmented image at right. The segmented image is white where there is no grass in the pitch, including players and advertisement boards.

To segment the grass regions on the image, a binary image is computed using peak values of the channels and thresholding the grass similar values out. To improve the identification of the players and reduce the noise in I_{NG} some erosions and dilations are applied to it (Fig.1). This step of the algorithm is similar to the work in [5]. At this point, after the normalization and segmentation of the image the algorithm is already able to classify the image as a long/medium distance shot, or a short distance shot.

A shot is classified as long, medium, or short distance according to the following [6]: If it is possible to segment more than 65% of the image as grass for more than 100 frames it is a long distance shot; if it happens for less than 100 frames the shot is classified as of medium distance; and finally, if less than 65% of the image is segmented as grass the shot is classified as of short distance.

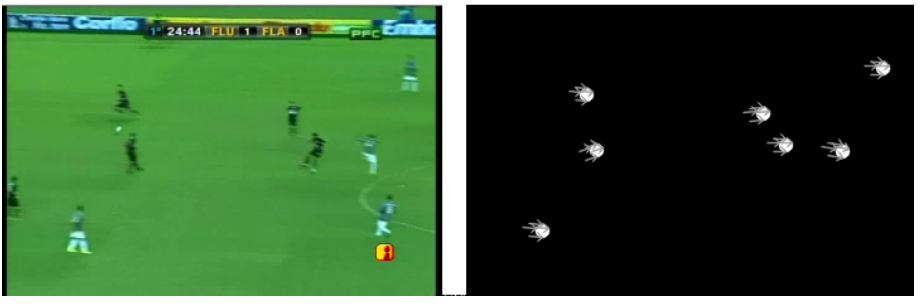

Fig. 2. The input image at left, and a modification of I_{FLRCTR} at right. The image at right also shows the velocities outputted by the optical flow.

The next step is to apply morphological filters to the binary image and then to identify the contours in I_{NG} comparing the area in each contour to two threshold values that are the maximum and minimum values of area for a region of players in the field. Those threshold values already take into account that a contour in I_{NG} possibly covers more than one player. This is done to differentiate the players from any other contour found in I_{NG} such as the audience or advertisement signs. Then, for each contour that is between the thresholds the minimum rectangle possible to contour it is chosen and then filled with white and put into another image I_{FLR} which will act as a mask for the tracking of the players. To increase the rate of success finding the players the box is enlarged a few pixels before being put in I_{FLR} . The tracking of the players is done in a new image, I_{FLRCTR} (Fig. 2), that contains circles centered in the centroids of the boxes at I_{FLR}. To track the players a similar algorithm by Shi and Tomasi [7] is applied, followed by computing the optical flow.

3 Construction of a Dense Velocity Field

To construct the dense velocity field we have to interpolate the velocities of each player, in this case we are interpolating the features extracted and tracked in the previous step. This velocity field should also be smooth so that any inference done by checking it should not be biased by points where the interpolation does not work well.

This is a surface interpolation problem, and it can be solved [4] using a radial basis function. The interpolation function used in this work is:

$$f(x) = c(x) + \sum_{i=0}^{n-1} \lambda_i . \varphi\left(\left\|x - x_i\right\|\right) \tag{1}$$

Where n is the number of points being interpolated, **x** is the position (x,y) of a pixel in an image, λ_i are coefficients, c(**x**) is a polynomial function, here:

$$c(x,y) = c_0 + c_1 x + c_2 y \tag{2}$$

The function $\phi(\|x - x_i\|)$ is a radial basis function, as mentioned in [1], some popular choices of Φ are: $\varphi = r$ (linear); $\varphi = r^2 \log r$ (thin-plate spline); $\varphi = e^{-ar}$ (Gaussian); $\varphi = \sqrt{r^2 + c^2}$ (Multi-quadratic); with

$$r = \left\|x - x_i\right\| \tag{3}$$

As in [4] and [1] we also chose the thin-plate spline. The thin-plate, or 2-D biharmonic spline, models the deflection of an infinite thin plate. While the linear radial basis function will interpolate the data, the thin-plate spline is more attractive since it also provides C^1 continuity and minimizes the energy function:

$$\int \left(\frac{\partial^2 \phi}{\partial x^2}\right)^2 + 2\left(\frac{\partial^2 \phi}{\partial y \partial x}\right)^2 + \left(\frac{\partial^2 \phi}{\partial y^2}\right)^2 dxdy \tag{4}$$

In this sense the thin-plate spline is the smoothest interpolating of $\varphi(x, y)$ [1].

In our implementation it was used a slight modification of the thin plate spline:

$$\varphi = r^2 \log(r + 1) \tag{5}$$

Solving the interpolation means to find the values of the coefficients λ_i, c_0, c_1, c_2. Since there are two components of the velocity of each point there must be done two interpolations in order to have the complete field. It ends to solving the following linear system for each component of the velocities [1]:

$$\begin{bmatrix} A & Q \\ Q^T & 0 \end{bmatrix} + \begin{bmatrix} \lambda \\ c \end{bmatrix} = \begin{bmatrix} U \\ 0 \end{bmatrix} \tag{6}$$

$$A = (a_{ij}) = \varphi\left(\left\|x_i - x_j\right\|\right) \tag{7}$$

$$c = \left(c_0, c_1, c_2\right)^T \tag{8}$$

$$\lambda = (\lambda_0, \ldots, \lambda_{n-1}) \tag{9}$$

$$Q=\begin{pmatrix} 1 & x_0 & y_0 \\ .. & .. & .. \\ 1 & x_{n-1} & y_{n-1} \end{pmatrix} \tag{10}$$

Where **U** is just a matrix with the velocities from each point being interpolated. After interpolating for both components of velocity we have:

$$\phi(x,y)=f(x,y)i+g(x,y)j \tag{11}$$

Where, $\varphi(x,y)$ is the dense velocity field, and $f(x,y), g(x,y)$ are its x and y components. Finally, to create smoother transitions between frames we apply a half-gaussian filter with $\sigma = 1$ to the last four values of φ calculated in the last four frames. The results of the interpolation can be seen in Fig. 3.

4 Identification of the Highlights

In order to identify a frame as a highlight we propose to quantify the convergence of the movements of the players, and the velocity of the camera. If the camera was hold still, the only movement seen would be of the players and the ball, therefore, the motion flow could be represented as $\varphi_1(x,y)$. Since the camera moves while the action is happening inside the pitch, the flow that is seen on the video is:

$$\phi(x,y)=\phi_1(x,y)+V_{CAM}(x,y) \tag{12}$$

The divergence of this dense velocity field is able to quantify the convergence of the movements of the players. Furthermore, $V_{CAM}(x,y)$ does not depend on the variation of x, or the variation of y, so:

$$\nabla.\phi(x,y)=\nabla.\phi_1(x,y)+\nabla.V_{CAM} \tag{13}$$

$$\nabla.\phi(x,y)=\nabla.\phi_1+\frac{\partial V_{CAMX}}{\partial x}+\frac{\partial V_{CAMY}}{\partial y} \tag{14}$$

$$\nabla.\phi(x,y)=\frac{\partial u}{\partial x}+\frac{\partial v}{\partial y} \tag{15}$$

Where, u, v are the components *x* and *y* of the dense velocity field. This shows that the divergence of φ not only quantifies the convergence of the movement of the players but it is also robust to the motion of the camera. The numeric computation was done in Matlab® , where the equations resulted in:

$$r=\sqrt{(x-x_i)^2+(y-y_i)^2} \tag{16}$$

$$\frac{\partial u}{\partial x} = c_1 + \sum_{i=0}^{n-1} \lambda_i . \left(-\log(r+1) - \frac{r}{2(r+1)} \right) . (2x_i - 2x) \tag{17}$$

$$\frac{\partial v}{\partial y} = c_2 + \sum_{i=0}^{n-1} \lambda_i . \left(-\log(r+1) - \frac{r}{2(r+1)} \right) . (2y_i - 2y) \tag{18}$$

The divergence is computed all over the image and then sampled at a grid. The negative values are threshold and then summed so that quantifying the convergence of the movement of the players could be easily done. In Fig.4 the threshold divergent is represented as a circle, the lower the value of the divergent, the redder the circle will appear in that image. The velocity seen in the video is the velocity of the camera plus the velocity of the players. And assuming that the velocities of the players are insignificant when the camera moves very fast, in a highlight for example, the velocity of the camera is approximated as:

$$V_{CAM} \simeq \frac{\sum_{i=0}^{n-1} u_i i + \sum_{i=0}^{n-1} v_i j}{n} \tag{19}$$

In the highlight frame of Fig.4 the blue arrow is V_{CAM}. A frame is defined as a highlight when these two conditions are satisfied:

- $\|V_{CAM}\|$ is large enough;
- The sum of threshold negative values of $\nabla . \varphi$ is low enough.

Each time a frame is identified as a highlight a flag is turned on, and the algorithm assumes that a region of 120 frames before and after that frame should be considered for further analysis.

5 Results and Evaluation

The evaluation of the algorithm was realized with the input of a full broadcast video of the match "Flamengo x Fluminense" played at Maracana Stadium, Brazil (2010). In Fig. 4 it can be observed that the convergence of the movement of the players works properly indicating where the play tends to continue. There were 322 frames considered as highlight frames, of those, 48 were false positives due to wrong classification of the scene, a total of 17.90% of false positive frames only.

One of the huge concerns in the development of the algorithm was to decrease the false positives due to camera movements that were fast but did not indicate any relevant event on the match, such as the movement of the camera after a goal kick for example. In this case the algorithm identified false positives just in 15 frames, a rate of 4.65% of false positives due to movement of the camera that was mainly caused by mistakes

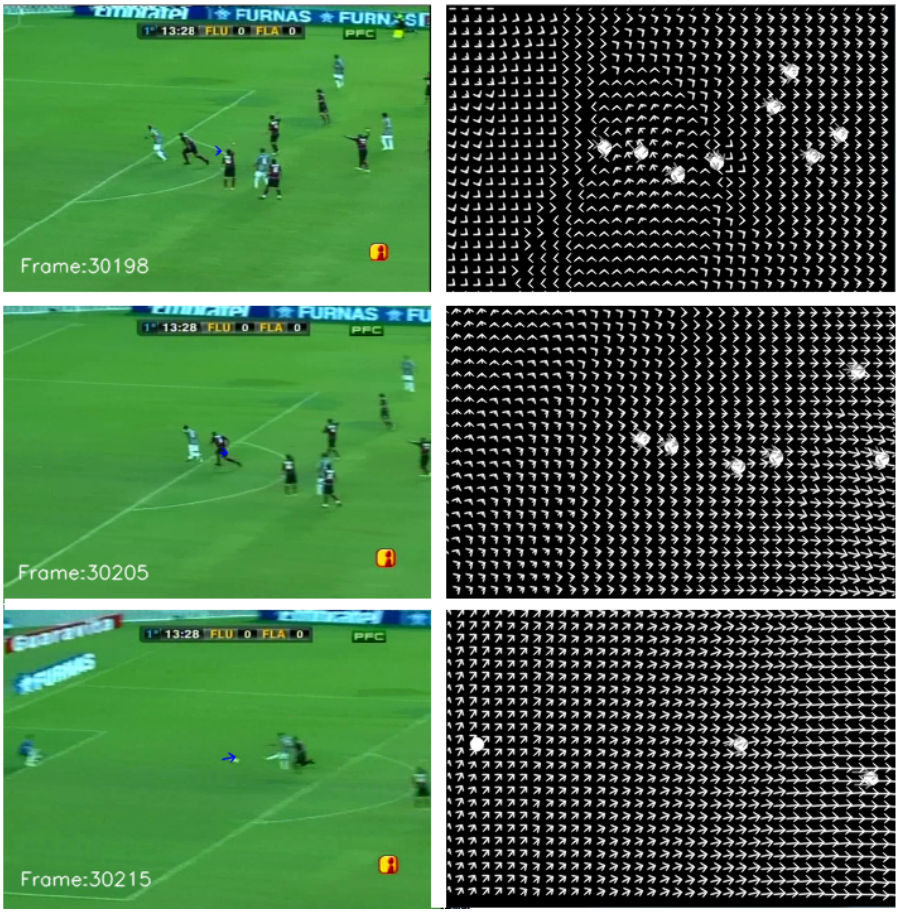

Fig. 3. Input image with the camera's velocity (as a blue arrow) at left, and the motion field at right represented with white arrows

Fig. 4. Frame identified as a highlight. The higher the value of the convergence, the redder is the circle. This image also contains the players boxes and the blue arrow at the center representing the camera velocity. This frame is within the sequence shown in Fig.3.

in the identification of the players. The full video of 109 minutes was summarized into a video of 18 minutes and 30 seconds, therefore a compression rate of 83.02%. The approach of analyzing global movements of the players and the movement of the camera showed good results, even with some amount of false positives. A significant amount of relevant plays on the match were successfully identified.

6 Conclusion and Future Work

In this work we proposed an algorithm for highlights identification and summarization of broadcast soccer videos. Results have shown that the movement of the players and the movement of the camera may indicate successfully relevant events in a soccer match. Most part of the false positives was due to flaws in the processing of the images. It was possible to obtain a compression rate of 83.02% which is a significant compression, which would help most of the public that would benefit from this summarization. A natural next step to the work in here would be to improve the study on the classification of the scenes and identification of the players in order to increase the compression rate of the algorithm.

Acknowledgments. This work was partially supported by FAPDF, DPP-UnB, and CIC-UnB.

References

1. Carr, J.C., et al.: Surface Interpolation with Radial Basis Functions for Medical Imaging. IEEE Transactions on Medical Imaging 16(1), 96–107 (1997)
2. Choi, K., Seo, Y.: Tracking Soccer Ball in TV Broadcast Video. In: Roli, F., Vitulano, S. (eds.) ICIAP 2005. LNCS, vol. 3617, pp. 661–668. Springer, Heidelberg (2005)
3. Choi, K., Seo, Y.: Probabilistic Tracking of the Soccer Ball. In: Comaniciu, D., Mester, R., Kanatani, K., Suter, D. (eds.) SMVP 2004. LNCS, vol. 3247, pp. 50–60. Springer, Heidelberg (2004)
4. Kim, K., et al.: Motion Fields to Predict Play Evolution in Dynamic Sport Scenes. In: Proceedings of CVPR 2010, pp. 840–847 (2010)
5. Seo, Y., et al.: Where are the Ball and the Players? Soccer Game Analysis with Color-based Tracking and Image Mosaick. In: Del Bimbo, A. (ed.) ICIAP 1997. LNCS, vol. 1311, pp. 196–203. Springer, Heidelberg (1997)
6. Sgarbi, E., Borges, D.L.: Structure in Soccer Videos: Detecting and Classifying Highlights for Automatic Summarization. In: Sanfeliu, A., Cortés, M.L. (eds.) CIARP 2005. LNCS, vol. 3773, pp. 691–700. Springer, Heidelberg (2005)
7. Shi, J., Tomasi, C.: Good Features to Track. In: Proceedings of CVPR, pp. 593–600 (1994)
8. Mountney, P.: Tracking Football Players using Conditional Density Propagation. Master Thesis, University of Bristol UK (2003)
9. Wu, Y., Fan, J.: Contextual flow. In: Proceedings of CVPR, pp. 33–40 (2009)