Acta Technica *62* No. 4B/2017, 1–6

Soccer player detection and tracking based on image processing[1]

Somayeh Lotfi[2,4], Baharak Shakeri Aski[3], Safoora Zakeri[2]

Abstract. Recent advances in soccer player detection and tracking have led to many new methods specifically designed for player tracking. Most of the attention, however, has been given to soccer video objects such as ball and players' detection and tracking. Though there is a number of existing literature to soccer player detection and tracking problems, we try to provide a higher detailed picture for a complete review. Within the broad survey and complex system, we p Table suggest the classification to describe modern soccer player detection and tracking. In addition, tables are utilized in the content to provide an overall view of soccer player detection and tracking methodologies.

Key words. Sports analytics, soccer player detection, soccer player tracking, blob based method, feature based method, occlusion detection.

1. Introduction

Nowadays, we could find much applications for video analysis in sports, such as statistics collection and video archiving, pattern analysis and motion replay. Because of development of new recording and processing devices, and the growing need for the media industry to extract meaningful information and analysis games, sports video analysis has attracted so attention. [1–3]. Sports video analysis consists of extraction of the spatiotemporal data with the corresponding athlete as well as a higher level abstraction of the events. In ball games, annotating video indexing becomes a more difficult task which requires the position of the ball as well as the position of the players [4, 5].

[1]This paper is extracted from research projects for Islamic Azad University, Bandar Abbas branch, Iran.

[2]Department of Computer Engineering, Bandar Abbas Branch, Islamic Azad University, Bandar Abbas, Iran

[3]Department of Computer Engineering, Ramsar Branch, Islamic Azad University, Ramsar, Iran

[4]Corresponding author; e-mail: `s.lotfi@iauba.ac.ir`

2. Methodology

We have implemented a vision system which automatically detects the ball in an acquired soccer video sequence. The proposed system is composed of three main blocks: in the first one, a background subtraction technique is combined with a circle detection approach to extract ball candidate regions. Then a feature extraction scheme is used to represent image patterns, and finally data classification is performed by using a supervised learning scheme. Figure 1 schematizes the proposed approach.

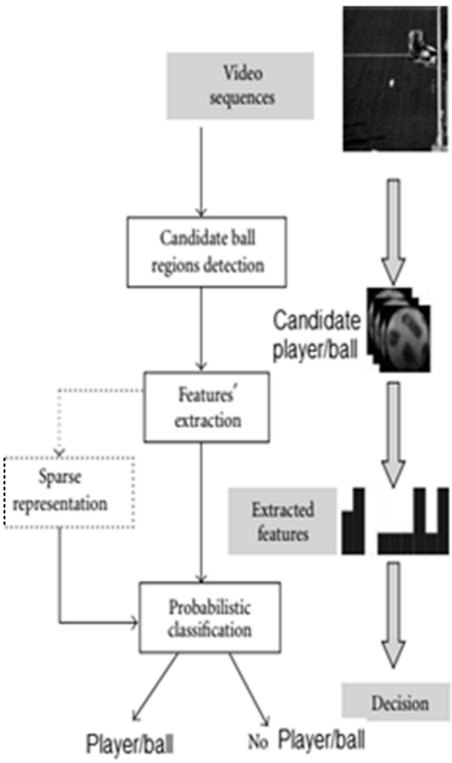

Fig. 1. Graphical overview of the proposed approach

The scale space of an image is defined as a function $L(x, y, \sigma)$, that is produced from the convolution of a variable-scale Gaussian function, $G(x, y, \sigma)$, with an input image, $I(x, y)$, that is

$$L(x, y, \sigma) = G(x, y, \sigma) * I(x, y)\,, \tag{1}$$

where $*$ is the convolution operator and $G(x, y, \sigma)$ is defined as

$$G(x, y, \sigma) = \frac{1}{2\pi\sigma^2} \mathrm{e}^{-\frac{X^2+y^2}{2\sigma^2}}\,, \tag{2}$$

The key points are detected using scale-space extreme in the difference of Gaussian (DoG) function D in convolution with the image $I(x,y)$:

$$D(x,y,\sigma) = [G(x,y,k\sigma) - [G(x,y,\sigma)] * I(x,y) = L(x,y,k\sigma) - L(x,y,\sigma)\,, \qquad (3)$$

where k is the multiplicative constant factor which separates two nearby scales. In order to detect the local maxima and minima of $D(x,y,\sigma)$, each sample point is compared to its eight neighbors in the current image and to its nine neighbors in the scales above and below. It is selected only if it is larger than all of these neighbors or smaller than all of them. Once a keypoint candidate has been found by comparing a pixel to its neighbors, the next step is to perform a detailed fit to the nearby data for location, scale, and ratio of principal curvatures. This information allows points to be rejected if they have low contrast or are poorly localized along an edge. A 3D quadratic function is fitted to the local sample points. The approach starts with the Taylor expansion (up to the quadratic terms) with sample point at the origin

$$D(X) = D + \frac{\partial D^{\mathrm{T}}}{\partial X} X + 0.5 X^{\mathrm{T}} \frac{\partial^2 D^{\mathrm{T}}}{\partial X^2} X\,, \qquad (4)$$

where D and its derivatives are evaluated at the sample point $X = (x,y,\sigma)$. The location of the extremum is obtained taking the derivative with respect to X and setting it to 0, giving

$$\widehat{X} = -\frac{\partial^2 D^{-1}}{\partial X^2} \frac{\partial D}{\partial X}\,. \qquad (5)$$

That is, a 3×3 linear system easily solvable. The function value at the extremum

$$D(\widehat{X}) = D + 0.5 \frac{\partial D^{\mathrm{T}}}{\partial X}\, \widehat{X} \qquad (6)$$

is useful for rejecting unstable extrema with a low contrast. A this point, the algorithm rejects also keypoints with poorly defined peaks, that is, those points having, in the difference of Gaussian function, a large principal curvature across the edge but a small one in the perpendicular direction. By assigning a consistent orientation, based on local image properties, the keypoint descriptor can be represented relative to this orientation and therefore achieve invariance to image rotation.

3. Results

3.1. Player detection

The rule G>R>B, where G, R, B are the Green, Red and Blue components in the RGB space, respectively, is held for the majority of ground pixels. In our proposed method, we apply this feature to extract the ground at first and construct a binary image where the non-ground pixels detected according to this rule are marked using equation

$$\text{Ground(x, y)} = \begin{cases} 0, & g(x,y) \geq r(x,y) \geq b(x,y)\,, \\ \\ 1, & \text{otherwise}\,. \end{cases} \tag{7}$$

Figure 2 shows these problems. As is shown in this figure, varying number of players in each frame, motion blur, and occlusion of players are the main problems of the soccer analysis. Also, our focus is fixed point implementation which is depicted in Fig. 3

Fig. 2. Challenge in player detection and tracking

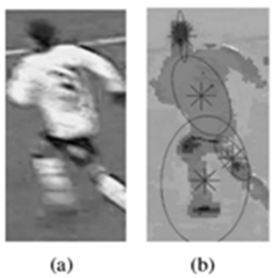

Fig. 3. Sample image of (a) player, (b) 16×4×4 uniform quantized players

As a result, we apply Sobel gradient algorithm to the RGB image to detect the sharp changes to preserve the ball [5]. In Sobel gradient algorithm, the derivative of the intensity values across the image is calculated and the change is found where the derivatives reach their maxima. The gradient is a vector and the components are measured in the x and y direction. The components are found using equations

$$\Delta x = \frac{f(x+\,\mathrm{d}x, y) - f(x,y)}{\mathrm{d}x}\,, \tag{8}$$

$$\Delta y = \frac{f(x, y+\,\mathrm{d}y) - f(x,y)}{\mathrm{d}y}\,, \tag{9}$$

$$M = \sqrt{(\Delta x)^2 + (\Delta y)^2}, \quad \theta = \tan^{-1}\left[\frac{\Delta y}{\Delta x}\right]. \tag{10}$$

Fig. 4. Blobs by removing unreasonable parts

Since the entire image is usually classified and blobbed, many blobs are formed that do not represent objects. These are discarded through a frequently complex system of rules and checks for the likelihood of the blob representing a valid object. Sets of blobs are then checked for mutual consistency. As shown in Fig. 4 the blobs are identified by removing unreasonable parts.

4. Ball detection and tracking

In Fig. 5, the ball and the lines on the ground were discarded. However, Sobel method retains the ball and the lines. Figure 6 is an example of this.

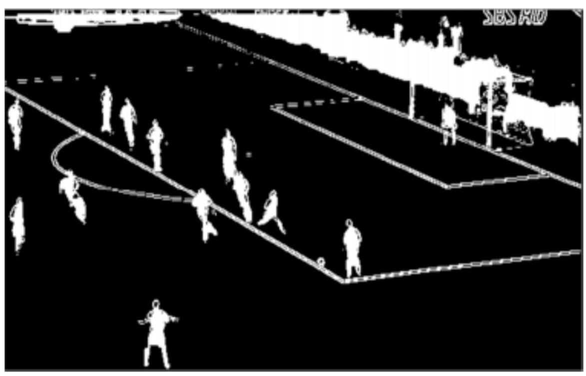

Fig. 5. Ground elimination result

Then the image containing the non-ground pixels (Fig. 5) and the Sobel gradient image (Fig. 6) are added, which yields an image with the players, ball, lines, goal post, spectators and scoreboard. Finally, the image is converted to the binary image and ground is eliminated from the resultant image shown in Fig. 5. Dilation and erosion methods are applied to connect the disjoint lines.

Fig. 6. Sobel gradient image

5. Conclusion

We have presented a summary of detection methodologies, procedures, and technologies for a soccer game. We indicate that there are two major methodologies for player detection techniques: feature based classification approaches and blob-based methods relying on background subtraction. Furthermore, heuristic-based methods have the merit of no prior information of athletes but do not operate well in real-time applications as a result of the high computational complexity. Moreover, we divide each of these groups to extra particular clusters. Eventually, we present a short summary for any of these techniques. The results show the capability and robustness of detecting player, ball and track lines.

References

[1] N. Babaguchi, Y. Kawai, T. Kitahashi: *Event based indexing of broadcasted sports video by intermodal collaboration.* IEEE Transactions on Multimedia *4* (2002), No. 1, 68–75.

[2] L. Ballan, G. M. Cortelazzo: *Marker-less motion capture of skinned models in a four camera set-up using optical flow and silhouettes.* International Symposium on 3D Data Processing, Visualization and Transmission, 18–20 June 2008, Atlanta, GA, USA, Proceedings of 3DPVT (2008), paper 37.

[3] L. Y. Duan, M. Xu, T. S. Chua, Q. Tian, C. S. Xu: *A mid-level representation framework for semantic sports video analysis.* ACM International Conference on Multimedia, 2–8 November 2003, Berkeley, CA, USA, Proceedings of Multimedia (2003), 33–44.

[4] H. Sabirin, H. Sankoh, S. Naito: *Automatic soccer player tracking in single camera with robust occlusion handling using attribute matching.* IEICE Transactions on Information and Systems *E98.D* (2015), No. 8, 1580–1588.

[5] R. A. Sharma, V. Gandhi, V. Chari, C. V. Jawahar: *Automatic analysis of broadcast football videos using contextual priors.* Signal, Image and Video Processing *11* (2017), No. 1, 171–178.

Received October 31, 2017