Research on Detecting Moving Objects in Football Match Video

Hongyi Xu¹, ²a, Jialing Lu¹, ², Qun He¹, Lei Gu¹, Chi Feng³, *
¹Nanjing University of Posts and Telecommunications, Nanjing, Jiangsu, China
²New York Institute of Technology, NY 10023, New York, USA
³State Grid Anqing Power Supply Company, Anqing, Anhui, China
ahxu18@nyit.edu  *Corresponding author’s e-mail: Fchi@163.com

Abstract. In recent years, detecting moving objects in video has attracted extensive attention in the field of computer vision and pattern recognition. With football becoming one of the most popular sports, analyzing the process of football games is of commercial value and guiding significance. However, the previous research mainly focused on the event detection, and rarely on getting the trajectory of moving objects by directly extracting from the video. This article discussed the method of detecting moving objects in the football match video. We extract the marker lines and build the coordinate system based on LUV color space. And we use the 68 landmarks to recognize the human face and the Kalman filter approach to detect the football. Our research is aimed at helping the audience and football professionals in analyzing the player performance and team strategy.

1. Introduction
With the popularity of the Internet and the development of technology, the audience can watch football matches in different ways. In the past few years, football fans used to stay up late to watch the World Cup in front of the TV. Nowadays, watching football games on the Internet has become the mainstream. In order to meet the increasing demand of the audience for watching games, we add the function of moving objects detection and image identification into the traditional football rebroadcast videos. This function can provide football professionals and fans with a more precise and targeted service for analyzing the game situation and the player performance.

2. Research Status at Home and Abroad
Up to now, a great deal of research has been done on video processing technology, and many effective methods have been proposed. Due to the uniformity and regularity, sports video has received a lot of attention from researchers.

2.1. Playground Inspection
In the football match videos, the presentation of the match is mainly achieved by telephoto shots, so the main semantic content is included in it. Ekin A. mainly used color thresholds for discriminating between field and off-field pixels [1]. However, the appearance of field pixels often changes with the environmental factors including illumination and so on. Hence, it is more difficult to use uniform color thresholds for football field detection. Barnard M. used GMM (Gaussian Mixture Model) to study the pattern of field pixels from the labelled samples [2], and then used the trained GMM to predict the
probability of each pixel in the field area. But it required labelling a large number of field pixels as training samples, which increased labour cost and decreased the practicability. Besides, there was a method using the pixels in the primary color region as training samples for GMM based on the coincidence [3]. In addition, a clustering-based field detection method was mentioned in Hung M. H.’s article, which differentiated whether a pixel belongs to the field region or not [4]. Summing up the above analyses, it can be seen that a single-color feature to distinguish between field pixels and outside. When outside pixels had similar color distributions to field, they are also detected as field pixels, resulting in detection errors. Therefore, there is a need to explore other features that can be applied for field detection to further improve the performance.

2.2. Players Detection
The difficulty in player detection mainly comes from the variable appearance patterns. In a football video, players are in active, so their appearance patterns will change continuously [5]. Besides, they will change when players are blocked by objects. The challenge in player detection is also to overcome the effects of these specific conditions [6]. Zhu G. designs a player detection method based on Support Vector Machine (SVM), which uses manually calibrated player samples to construct a training sample set and then uses color histograms as features to train an SVM-based player detector [7]. Dalal N. referred to this approach, using gradient histogram features to train SVM-based player detectors [8]. Due to the generalization performance of SVM, the detection performance of it greatly improved for detecting players. Additionally, the first “Adaboost” object detection method is proposed by Liu J., which was to automatically select the most discriminating features from a large number of Haar-like features. And it ensured the correctness and detected speed of the detection results by cascading the weak classifiers [9].

2.3. Locating the Football
The detection and tracking of the ball are foundations to football video analysis. Many researchers and institutions explored many aspects of it and proposed effective methods. For example, an improved Hoff transform of a circle could detect circular objects for football segmentation [10], but this method was only applicable in monochromatic solids. Besides, Liang D. first roughly got the target balls in consecutive frames and constructed a weight map [11]. Then he extracted the paths corresponding to the positions of the real balls using the Viterbi algorithm, and finally uses Kalman filter with template matching to track the target balls.

2.4. Event Detection
Denman, H. discussed a method for detecting goal events in snooker. The event of a goal corresponds to the disappearance of the ball in a particular area in the table [12]. When the ball disappears near the pocket, it is considered that a goal event has occurred. Bebie T. showed a system that reconstructs the scenario of a football match by the color information of players and field [13]. Chen Shu-ching also proposed a combination of visual and audio features to detect the occurrence of goal events in the football match [14]. For the visual features, the author used pixel-level comparison, histogram comparison, segmented graph comparison and target tracking to detect the boundary between two shots. When a goal event occurs, the transition of shots and the change in audio have a certain regularity. Similarly, in the analysis of baseball games, Angela Yao used the detection and extraction of audience applause and the sound of the bat hitting the ball to achieve the shot segmentation [15]. In addition, global motion characteristics or camera motion parameters are also commonly used features in the analysis. In Zhang N.’s article, the motion trajectories and mutual motion relationships of video objects were applied to segment the football match video [16].

3. Materials and Methods
According to the above research, this paper will be based on three aspects: marker lines extraction, players recognition and football localization. After locking the objects successfully, we use open source
deep learning databases such as face recognition library (MIT license) to eliminate environmental factors obstruct which can improve the quality of the extraction. Finally, we set up a real-time system to realize the extended application for coordinates positioning to get the player’s running distance, football trajectory and so on.

4. Maker Lines Extraction
Since the size of a standard football playground is a constant, we can determine the range of match activities from the target objects. These marker lines are important parts in the field and implying coordinate information. Their positions are fixed on the field, and the target objects can be obtained by their relative positions.

4.1. LUV Color Space
Before extract the marker lines, we need a more efficiency data type to the next research. LUV color space is CIE 1976(L*, u*, v*) (also called CIELUV) color space. L* represents the brightness of the object, and u* and v* are the chromaticity. It was proposed by the International Commission on Illumination (CIE) in 1976, and it is obtained by a simple transformation of the CIE XYZ space, which has visual uniformity. The similar color space is CIELAB, where the values of u* and v* are from -100 to +100 and the luminance L* is from 0 to 100 for general images. This color space can help us to complete more complicated calculations such as neighbor-pixel differential [17].

4.2. Neighbor-Pixel Differential
The most commonly used algorithm in edge extraction is the Canny edge detect operators. It can identify as many edges of the image as possible, and has the advantage that the edges are not repeatedly identified. But in this paper, the main purpose is to extract marker lines, and discontinuities appear in the marker lines extracted by the Canny edge detect operators due to the luminance. Therefore, we use the neighbor-pixel differential which is convenient for extracting the rectangular border.

According to the characteristics of the rectangular border of the field marker line, we adopt the adjacent pixel differential method for the extraction of marker lines. The edges of the field marker lines are approximately straight lines in both horizontal and vertical directions, so the adjacent pixel difference in both horizontal and vertical directions can be used to extract the field areas easily and efficiently. Neighbor-pixel differential is defined as the difference between two adjacent rows and the difference between two adjacent columns. The formula for the difference between columns is:

\[ \text{DiffC}(x, y) = |CL_{i+1} - CL_i| + |CU_{i+1} - CU_i| + |CV_{i+1} - CV_i|, i \in \{0, H - 1\} \]  

In (1), \( CL_{i+1}, CU_{i+1}, CV_{i+1} \) are components of columns \( i+1 \) in LUV color space, while \( CL_i, CU_i, CV_i \) are components of columns \( i \) in LUV color space. \( H \) is the height of the image.

Similarly, the inter-row differential is:

\[ \text{DiffR}(x, y) = |RL_{i+1} - RL_i| + |RU_{i+1} - RU_i| + |RV_{i+1} - RV_i|, i \in \{0, W - 1\} \]  

In (2), \( RL_{i+1}, RU_{i+1}, RV_{i+1} \) are components of rows \( i+1 \) in LUV color space, while \( RL_i, RU_i, RV_i \) are components of rows \( i \) in LUV color space. \( W \) is the weight of the image.

In order to reduce the interference of these different cases on the marker line extraction, the automatic threshold image binarization based on the neighboring pixel values is used for processing. Based on the characteristics of the neighboring pixel differential method, the expectation of the row differential value multiplied by the percentage of pixels with small differential values is chosen as the threshold value, which is denoted as \( T_p \).

\[ T_p = \frac{\Sigma \text{DiffC}(x, y) + \Sigma \text{DiffR}(x, y) \times \max(n_p)}{WL(W-1)(L-1)} \]

(3)
In (3), \( n_p \) is the number of pixels whose values are smaller than the inter-row and inter-column differences. If the rows and columns of the corresponding point in the differential image are all less than \( T_p \), the point is set to white, otherwise it is set to black.

### 4.3. Building the Coordinate System

In the experiment, the court occupies most of the collected image area, and the field area is marked by the maximum connected domain detection algorithm. The binary image connectivity domain tagging method is to tag the pixels of the image according to some connectivity rules, commonly used are four-neighborhood tagging and eight-neighborhood tagging.

In a four-neighborhood region, pixels have four neighborhood relationships between them, as shown in Figure 1, pixel P and its four horizontal and vertical neighbors have a four-neighborhood relationship. An eight-neighborhood is a connected region where pixels have eight neighbors and pixel P has eight neighbors with eight pixels adjacent to it.

![Figure 1. Four neighborhood relationships and eight neighborhood relationships](image)

Based on the fact that the court edges are rectangular and are approximately in the horizontal and vertical directions respectively, we adopt a 4-neighborhood connectivity approach for line scanning the binary single-frame images extracted from soccer videos. According to the characteristics of the largest area of the football match image, the marked area with the greatest number of pixels is the court area.

The coordinate system is established as the basis for subsequent football positioning. After detecting the court area, a linear fit is used to establish the pitch coordinate system based on the nature of the book edge as a straight line. In order to obtain the set of points required for the straight line fit, a scanning method is used to extract the set of pitch edge points from the top and bottom of the pitch area. The formula for the straight-line fit is as follows:

\[
 f(x) = \hat{k}x + \hat{b} \tag{4}
\]

\[
 \begin{align*}
 \hat{k} &= \frac{\sum_{i=1}^{N} x_i y_i - N \sum_{i=1}^{N} x_i \sum_{i=1}^{N} y_i}{\sum_{i=1}^{N} x_i ^2 - N \left(\sum_{i=1}^{N} x_i \right)^2} \\
 \hat{b} &= \frac{1}{N} \sum_{i=1}^{N} y_i - \hat{k} \left(\sum_{i=1}^{N} x_i \right) 
\end{align*} \tag{5}
\]

Equation (5) is famous as generalized least squares, \( N \) is the number of pixels in the edge point set. \((x_i, y_i)\) is the pixel coordinate.

### 5. Players Recognition

For the established coordinate, putting target objects into the actual court and comparing the relative position in the coordinate is a good way to detect or correct the data. As to recognizing players, face recognition based on open-access deep learning database is a practicable method.

For player identification, we adopt face recognition to implement. As shown in Figure 2, there are 68 landmarks labelling in the human face. With these 68 landmarks, it is easy to recognize where the eyes and mouth are, and follow up by rotating, scaling, and misshaping the image so that the eyes and mouth are as close to the center as possible.
Figure 2. 68 landmarks are labelled in this person’s face

OpenFace has trained a deep convolutional neural network. It can be trained to generate 128 measurements for the face, and all we need to do is to find the one in the database that most closely matches the measurements of our test image. The Euclidean distance between two 128D values can be calculated using some readily available mathematical formulas. In this way we get a Euclidean distance value, and the system will give it a threshold of Euclidean distance, beyond which we consider them to be the same person.

\[ d(x, y) = \sqrt{\sum_{i=1}^{n} (x_i - y_i)^2} \]  

(6)

6. Football Localization

The coordinates of the football are located in the rebroadcast video, and Kalman filter can easily find its trajectory. Through further analysis, the motion state, direction, speed, acceleration and other parameters of the football can be calculated [18].

With the development of digital computing technology, Kalman filter has become the subject of popular research and application in digital signal processing, especially for target tracking and navigation problems [19]. As an effective state estimation method, Kalman filter uses the target state model to predict the state of the target at the next moment, combined with the observation model to calculate the a posteriori probability density function of the state. When the state model of the target state and the observation model meet the Gaussian and linear conditions, Kalman filter can obtain the optimal solution in the sense of minimum mean square error.

The Kalman filter can be divided into two parts: the time update equation and the measurement update equation. The former is used to extrapolate error covariance estimates and state variable sums forward to construct a priori estimates of the target state for the next moment. The latter is responsible for feedback, combining the new measurements with the a priori estimates to construct posteriori estimate of the target state [20].

Assuming the football in the position \( p \), let \( v \) denote the velocity and \( a \) denote the acceleration. \( p, v, a \) are called state variables. We can use the state variable matrix \( A_t = \begin{bmatrix} p_t \\ v_t \end{bmatrix} \) to represent the state of the football at the moment \( t \). After a time of \( \Delta t \), the position and velocity at the current moment are:

\[ p_t = p_{t-1} + v_{t-1} \Delta t + \frac{a_{t-1} \Delta t^2}{2} \]  

(7)

\[ v_t = v_{t-1} + a_t \Delta t \]  

(8)

Because Kalman filter can only describe a linear relationship between states, (7) and (8) can be converted to a matrix vector of the form:

\[ \begin{bmatrix} p_t \\ v_t \end{bmatrix} = \begin{bmatrix} 1 & \Delta t \\ 0 & 1 \end{bmatrix} \begin{bmatrix} p_{t-1} \\ v_{t-1} \end{bmatrix} + \begin{bmatrix} \frac{\Delta t^2}{2} \\ \frac{\Delta t^2}{2} \end{bmatrix} u_t \]  

(9)
In (9), there are two state change matrices, $k_t = \begin{bmatrix} 1 & \Delta t \\ 0 & 1 \end{bmatrix}$ and $b_t = \begin{bmatrix} \frac{\Delta t^2}{2} \\ \Delta t \end{bmatrix}$, so (9) can be rewrite as:

$$\hat{A}_t = k_t \hat{A}_{t-1} + b_t a_t$$

Equation (10) is the Kalman filter’s state prediction formula, where $k_t$ is called the state transfer matrix, which indicates how we deduce the current state from the previous state. $b_t$ is called the control matrix, which indicates the effect of the control quantity $u$ on the current state. $\hat{A}$ indicates that it is an estimate, not the true value, and the $\tilde{t}$ just indicates that it is not the best estimate.

The covariance matrix is then used to represent the uncertainty of the prediction. For two-dimensional noise, the x-direction satisfies a Gaussian distribution and the y-direction also satisfies a Gaussian distribution, which can be correlated, so we need the covariance matrix to represent them.

$$C_t = \begin{bmatrix} \sigma_{xx} & \sigma_{xy} \\ \sigma_{yx} & \sigma_{yy} \end{bmatrix}$$

The noise covariance matrix will have an effect on the system, so we have to pass it on to the system, which is multiplied by the state transfer matrix $k$.

$$C_t = k C_{t-1} k^T$$

Finally, we can get the Kalman parameter through a series of transformations with observation matrix $H$:

$$K_t = C_t H^T (H C_t H^T)^{-1}$$

7. Conclusion

This paper has provided a method to detect moving objects such as the football and players in the football match videos. After analyzing the related research on playground inspection, players detection, football localization and event detection, we found that there was little research on getting the trajectory of moving objects directly from the video. Therefore, we solved the above problem in three aspects: marker lines extraction, players recognition and locating the football. First, based on LUV color space, we used the neighbor-pixel differential to build the coordinate system. Then, we recognized the player according to the 68 landmarks labelling on the human face. Finally, we applied the Kalman filter in locating the football. Our project will help the football commentators and coaches analyze team strategy and player performance.

References

[1] Ekin, A., & Tekalp, A. M. (2003) Robust dominant color region detection and color-based applications for sports video. In Proceedings 2003 International Conference on Image Processing (Cat. No. 03CH37429) (Vol. 1, pp. I-21). IEEE.

[2] Barnard, M., & Odobez, J. M. (2004) Robust playfield segmentation using MAP adaptation. In Proceedings of the 17th International Conference on Pattern Recognition, 2004. ICPR 2004. (Vol. 3, pp. 610-613). IEEE.

[3] Liu, Y., Jiang, S., Ye, Q., et al. (2005) Playfield detection using adaptive GMM and its application. In Proceedings. (ICASSP'05). IEEE International Conference on Acoustics, Speech, and Signal Processing, 2005. (Vol. 3, pp. 610-613). IEEE.

[4] Hung, M. H., Hsieh, C. H., & Kuo, C. M. (2009) An unsupervised playfield segmentation for various sport videos. In 2009 Fourth International Conference on Innovative Computing, Information and Control (ICICIC) (pp. 349-352). IEEE.

[5] Peker, K. A., Cabasson, R., & Divakaran, A. (2001) Rapid generation of sports video highlights using the MPEG-7 motion activity descriptor. In Storage and Retrieval for Media Databases 2002 (Vol. 4676, pp. 318-323). International Society for Optics and Photonics.
[6] Assfalg, J., Bertini, M., Del Bimbo, et al. (2002) Soccer highlights detection and recognition using HMMs. In Proceedings. IEEE International Conference on multimedia and expo (Vol. 1, pp. 825-828). IEEE.

[7] Zhu, G., Xu, C., Huang, Q., et al. (2008) Event tactic analysis based on broadcast sports video. IEEE Transactions on Multimedia, 11(1), 49-67.

[8] Dalal, N., & Triggs, B. (2005) Histograms of oriented gradients for human detection. In 2005 IEEE computer society conference on computer vision and pattern recognition (CVPR'05) (Vol. 1, pp. 886-893). IEEE.

[9] Liu, J., Tong, X., Li, W., et al. (2009) Automatic player detection, labeling and tracking in broadcast soccer video. Pattern Recognition Letters, 30(2), 103-113.

[10] D'Orazio, T., Ancona, N., Cicirelli, G., et al. (2002) A ball detection algorithm for real soccer image sequences. In Object recognition supported by user interaction for service robots (Vol. 1, pp. 210-213). IEEE.

[11] Liang, D., Liu, Y., Huang, Q., et al. (2005) A scheme for ball detection and tracking in broadcast soccer video. In Pacific-Rim Conference on Multimedia (pp. 864-875). Springer, Berlin, Heidelberg.

[12] Denman, H., Rea, N., & Kokaram, A. (2003) Content-based analysis for video from snooker broadcasts. Computer Vision and Image Understanding, 92(2-3), 176-195.

[13] Bebie, T., & Bieri, H. (1998) SoccerMan-reconstructing soccer games from video sequences. In Proceedings 1998 International Conference on Image Processing. ICIP98 (Cat. No. 98CB36269) (Vol. 1, pp. 898-902). IEEE.

[14] Chen, S. C., Shyu, M. L., Zhang, C., et al. (2003) Detection of soccer goal shots using joint multimedia features and classification rules. MDM/KDD, 3, 27.

[15] Yao, A., Uebersax, D., Gall, J., et al. (2010) Tracking people in broadcast sports. In Joint Pattern Recognition Symposium (pp. 151-161). Springer, Berlin, Heidelberg.

[16] Zhang, N., Duan, L. Y., Li, L., et al. (2012) A generic approach for systematic analysis of sports videos. ACM Transactions on Intelligent Systems and Technology (TIST), 3(3), 1-29.

[17] Zhou T. (2014) Research on point-reader coordinate positioning method based on video images (Master's thesis, Nanchang Aviation University) (in Chinese).

[18] Lv T. (2012) Research on ball detection and tracking algorithm in football video (Master's thesis, Xidian University).

[19] Liu, Y., Liu, S., Wang, Y., et al. (2020) Video Coding and Processing: A Survey. Neurocomputing.

[20] Zhu, H., Yan, X., Tang, H., et al. (2020) Moving Object Detection with Deep CNNs. IEEE Access, 8, 29729-29741.