Spiking and Blocking Events Detection and Analysis in Volleyball Videos

Chun-Chieh Hsu\(^1\), Hua-Tsung Chen\(^2\), Chien-Li Chou\(^1\), Suh-Yin Lee\(^1\)

\(^1\) Department of Computer Science, National Chiao Tung University, Hsinchu, Taiwan
\(^2\) Information & Communications Technology Lab, National Chiao Tung University, Hsinchu, Taiwan

\{hsuch, huatsung, fallwind, sylee\}@cs.nctu.edu.tw

Abstract—In volleyball matches, spiking is the most effective way to gain points, while blocking is the action to prevent the opponents from getting scores by spiking. In this paper, we propose an intelligent system for automatic spiking events detection and blocking pattern classification in real volleyball videos. First, the entire videos are segmented into clips of rallies by whistle detection. Then, we find the court region based on proper camera calibration, and detect the location of the net for judging the positions of spiking and blocking. Via analyzing the changes of moving pixels along the net, we make a bounding box around the blocking location, so as to classify the blocking patterns into two main categories based on the width of bounding box. Finally, two important tactic patterns, delayed spiking and alternate position spiking, are recognized. With the information of spiking events and blocking locations, we can collect the statistical data and make tactics inference easily. To the best of our knowledge, no previous work is focused on spiking or blocking event detection. The experimental results on the videos recorded by a university volleyball team are promising and demonstrate the effectiveness of our proposed scheme.

Keywords-component: sports video analysis; volleyball video; tactic analysis; image processing; event detection

I. INTRODUCTION

With the rapid development of multimedia capturing devices, it becomes easier for people to record their lives. A large amount of multimedia data, such as images and videos, are produced and uploaded to the internet. Therefore, automatically analyzing and understanding the complex and compound multimedia data becomes an important issue. Sports video is one of the most popular multimedia data since sports games hold a lot of audiences worldwide. There has been an explosive growth of researches focusing on analyzing sports videos due to the potential commercial benefits and entertainment demands.

For semantic analysis, we need to find the correspondence between real-world coordinates and coordinates in the video frame. Farin [1] [2] propose a camera calibration algorithm for sports videos based on planar reference objects as well as the court model. Feature points on a plane appearing in different views are required for the plane-based calibration techniques. In reality, significant events are mainly resulted from ball-player and player-player interaction. Yu [3] present a trajectory-based algorithm for ball detection and tracking in soccer video. Potential trajectories are generated from ball candidates by a Kalman filter based verification procedure. The true ball trajectories are finally selected from the potential trajectories according to a confidence index. Chen [4] [5] propose a physics-based algorithm for ball tracking. The characteristic that ball trajectory presents in a near parabolic curve in video frames is exploited in ball position prediction and trajectory extraction. In [6], ball tracking and 3D trajectory reconstruction in basketball videos and shooting location statistics can be obtained. The proposed scheme incorporates domain knowledge and physical characteristics of ball motion into object tracking to overcome the problem of 2D-to-3D inference. Soudeh [7] present a novel approach for tracking the ball and players for indoor soccer games. To track players and the ball, a fast level set contour is used. Pallavi [8] propose a graph-based approach for detecting and tracking multiple players in broadcast soccer videos. A directed weighted graph is constructed, where the nodes represent probable player candidates and each edge links two candidates in two consecutive frames. Finally, dynamic programming is applied to find the trajectory of each player. Liu [9] present a scheme to perform automatic multiple player detection, unsupervised labeling and efficient tracking in broadcast soccer videos. Player detection is achieved by combining the dominant color based background subtraction and a boosting detector with Haar features. And multiple players tracking with Markov chain Monte Carlo (MCMC) data association is performed based on detection and labeling.

Increasing researches focusing on analyzing the tactics used in sports videos. Chang [10] design a Wild-Open Warning (WOW) system to help basketball coaches and players in revealing possible tactics of their opponents. Hu [11] develop a quadrangle candidate generation algorithm and refine the model fitting score to ameliorate the court-based camera calibration technique to be applicable to broadcast basketball videos. Implicit/explicit tactics inferring can be made with the player position and trajectory information in the court coordinate. Zhu et al. [12] propose a novel approach to extract tactic information from the attack events in broadcast soccer video. A hierarchical coarse-to-fine framework is provided for discovering the tactic patterns.
Chen [5] present a physics-based scheme for ball detection and trajectory extraction in volleyball videos. In volleyball matches, players are not allowed to hold the ball so that the ball trajectories almost show in parabolic curves. A 2D distribution analysis is proposed using the physical characteristic that the ball moves parabolically in Y-direction and straight in X-direction as time goes on. The basic actions of players can be detected at the transitions of the ball trajectory. And a set type can be recognized by the set curve. Although attack events can be detected based on ball trajectory in [5], it only locate where the ball is hit. The cloak is not detected in this work. In fact, further tactic analysis needs detecting the cloak since executing tactics requires interaction between players.

For game studies/reviews, there are plenty of videos recorded by coaches, players, or even the audiences with a fixed digital camera, as shown in Fig. 1. For the audience, they usually pay their attention to the exciting events, such as splendid spikings and wonderful digs, since it may take a long time to watch the entire matches. It is a trend to design automatic systems for content-based video retrieval and semantic analysis in order to display such selective events and shots. From the coaches’ and the professional players’ points of view, analyzing the tactic patterns executed by the opponent teams can help them work out corresponding strategies in the training process or even in the matches. However, it is time-consuming and labor-intensive to manually recognize the tactic patterns and collect the statistical data from a large amount of sports videos. Consequently, it becomes an indispensable task to establish a system for automatic tactic patterns recognition and semantic events extraction.

In volleyball matches, spiking is the most direct and effective way to get points and it is always the most exciting part in the matches. On the other hand, blocking can prevent opponents from getting scores by spiking, and even get points with successful blocking, which means grounding the spiking ball back on the opponents’ court. The blocking patterns can be classified into four types based on the number of people included: none, single, double, and triple blockings. The more people participate in blocking, the higher probabilities of successful blocking would make. In addition, marvelous spiking such as delayed spiking and alternate position spiking always results from brilliant set which leads to none or only single blocking. On the other hand, the set type plays an important role in tactics analysis. For this reason, classifying the blocking patterns can help the coaches analyze the successful tactics used by the setters. As for spiking or blocking event detection, there is no previous work, at least to the best of our investigation.

In this paper, we propose a volleyball video analysis system capable of automatic spiking event detection and blocking pattern classification. The schematic flowchart of the proposed framework is illustrated in Fig. 2. Since not all significant semantic events can be detected by only using visual features, audio features should also be considered for assisting video content analysis. Whistle is one of the most indicative audio events for segmenting volleyball matches, so the proposed system starts with whistle detection for the determination of the play boundaries of each rally. Here we adopt the whistle detection method in [5]. For the segmented video clips of rallies, we need to find the correspondence between the court in the video and the real world court. The camera calibration is performed to find the court region. Actually, spiking and blocking events occurred above the top of the net. Therefore, we further detect the location of the net. Then we analyze the changes of moving pixels along the horizontal net, to mine the patterns of spiking and blocking. With the spiking events and blocking locations, we can analyze the patterns of two important tactics in volleyball matches.

II. Camera Calibration

Camera calibration is performed to find the correspondence between the image court model and the real world court model [1] [2]. We start with detecting the white line pixels by the constraints of color and local texture. To extract the court lines, the Hough transform is applied to the detected white line pixels. Thus, the straight lines such as
court lines can be obtained. Then, the intersection points of court lines are calculated. With the corresponding points, the transformation can be obtained and the camera parameters can be then derived. The location of the net can be further recognized by Hough transform. The calibration is only calculated at the beginning of the videos, since the camera is fixed and the background is almost the same in the whole match.

A. Court Line Pixel Detection

The court lines are always in white color for visual clarity. Assuming that court lines are typically not wider than $\tau$ pixels ($\tau = 6$ in our framework), we check whether the brightness at a distance of $\tau$ pixels from four sides of the candidate pixel is considerably darker than the candidate pixel [2]. The white line pixels are classified as court line candidates according to (1):

$$I(x, y) = \begin{cases} 1, & g(x, y) - g(x - \tau, y) > \sigma_d \wedge g(x, y) - g(x + \tau, y) > \sigma_d \\ 1, & g(x, y) - g(x, y - \tau) > \sigma_d \wedge g(x, y) - g(x, y + \tau) > \sigma_d \\ 0, & \text{else} \end{cases} (1)$$

where $I(x, y)$ indicates if a pixel $(x, y)$ is a court line candidate ($I(x, y) = 1$) or not ($I(x, y) = 0$), $g(x, y)$ means the luminance of a pixel at position $(x, y)$, and $\sigma_d$ is the luminance threshold. This process prevents most of the pixels in white regions or white uniforms being detected as white line pixels.

B. Projection Matrix Computation

A standard Hough transform is performed on the detected white line candidate pixels to extract the court lines. Then, we calculate the intersection points [2] and solve the equation system defined as following:

$$\begin{bmatrix} x_1 & y_1 & 1 & 0 & 0 & 0 & -x_1x_1 & -x_1y_1 \\ 0 & 0 & 0 & x_1 & y_1 & 1 & -y_1x_1 & -y_1y_1 \\ x_2 & y_2 & 1 & 0 & 0 & 0 & -x_2x_2 & -x_2y_2 \\ 0 & 0 & 0 & x_2 & y_2 & 1 & -y_2x_2 & -y_2y_2 \\ x_n & y_n & 1 & 0 & 0 & 0 & -x_nx_n & -x_ny_n \\ 0 & 0 & 0 & x_n & y_n & 1 & -y_nx_n & -y_ny_n \end{bmatrix} \begin{bmatrix} h_{00} \\ h_{10} \\ h_{20} \\ h_{11} \\ h_{21} \end{bmatrix} = \begin{bmatrix} x_1 \\ y_1 \\ x_2 \\ y_2 \\ x_n \\ y_n \end{bmatrix} (2)$$

Here we use the four sides of the lower half court to solve the equation system, as shown in Fig. 3. A sample result of court detection is presented in Fig. 4.

C. Net Recognition

In order to detect the spiking events and classify the blocking patterns, the location of the net is essential. We eliminate the pre-detected court line pixels in the result of court line pixel detection. Besides, for simplicity, we draw two vertical segments above the center line as the two vertical markers of the net. Then, Hough transform is performed on the area between two vertical lines above the center line. The detected two nearly horizontal lines are regarded as two horizontal markers of the net. Fig. 5 shows two detection results at slightly different angles of the nets in two matches.

III. SPIKING EVENT DETECTION

Since the spiking and blocking events occurred above the upper horizontal marker of the net, the analyzing area should be higher than the net. For example, we move one third area of the net to the higher location in our experiment. In fact, blocking always accompanies spiking. For this reason, in order to classify the blocking patterns, the spiking events should be extracted first.

Frame difference method is applied to analyze moving pixels, since there is no camera motion in the fixed camera. A Frame Difference Image (FDI) is a binary image formed by comparing every three successive frames, as defined in (3).

$$\text{FDI}(x, y) = \begin{cases} 255, & \text{if } |\text{Intensity}_n(x, y) - \text{Intensity}_{n-1}(x, y)| > T_d \\ 0, & \text{otherwise} \end{cases} (3)$$

Fig. 6 presents the examples of frame difference images.

A. Jump Moment Extraction

Both spiking and blocking are based on the action of jump. The exact moment of jump is indispensable for spiking events detection and blocking patterns classification. In other words, we can detect spiking events by extracting frames of the jump moment. By observation, the sum of moving pixels within the analyzed area has obvious change while the
players are jumping. As shown in Fig. 7(a), the peaks of continuous changes in sum of moving pixels at each rally are discovered.

The frames corresponding to peaks at each rally are analyzed. In frames corresponding to peaks, we take the columns along the analyzed area into account if it contains moving pixels more than a threshold $T_{CMP}$. If the number of columns $N_{col}$ is more than another threshold $T_{NC}$, the frame is regarded as jump frame, as shown in Fig. 7(a). In our experiment, $T_{CMP} = \frac{1}{4} \times \text{(width of analyzed area)}$, and $T_{NC} = \frac{1}{18} \times \text{(length of analyzed area)}$. The parameter $\frac{1}{18}$ is chosen since a net (9 m) can be roughly occupied by 18 men and each man occupies about $\frac{1}{2}$ length of the net. The retrieved frames of jump moment extraction are shown in Fig. 7(b). If the peak of continuous change in sum of moving pixels within the analyzed area satisfies the conditions, we extract the corresponding frame as frame with jump.

IV. BLOCKING LOCATION DETECTION

Based on the extracted frames with jump, further analysis to detect the blocking events at that moment can be made. First, where the spiking and blocking occur should be located. Second, the blocking patterns will be classified into two main categories. Finally, we can make some tactics analysis with spiking events and blocking locations.

For the extraction of the frames with jump, the distribution of the amount of moving pixels along the horizontal restricted area is computed. We first find the column where the maximum of the distribution occurs. From that column, the leftmost and most right columns containing moving pixels are regarded as the blocking boundaries. The search distance $S_{dis}$ should be constrained in case of locating the noises not belonging to blocking. We set the threshold to be length of 3 men since at most 3 men block at the same time. And we make a bounding box to surround the boundary, called blocking region. As shown in Fig. 8, we depict the blocking region in the frames with jump.

V. BLOCKING PATTERN CLASSIFICATION

In volleyball matches, the more people participate in blocking, the higher probabilities of successful blocking would be. Consequently, the blocking patterns can be classified based on the number of people included. In our proposed scheme, the blocking patterns are classified into two main categories based on the width of blocking region: none/single, and twice/triple. None or single blocker involved may result from brilliant tactics or may just a cloak. On the contrary, twice or triple blockers involved means that the success rates of spiking are lower than none or single. We verify which class the frames with jump belong to by their widths of blocking region by (4), where $F$ means the extracted frame with jump. $\text{Class}_{A}$ represents none/single, and $\text{Class}_{B}$ is twice/triple.
VI. TACTIC ANALYSIS

Delayed spiking and alternate position spiking are two basic kinds of important tactics in volleyball matches. For this reason, we try to find the patterns of two kinds of tactics based on domain knowledge. Delayed spiking utilizes close time and positions between the jumps of two spikers to confuse the blockers. Alternate position spiking avoids too many blockers by making a wide set. To discriminate the two kinds of tactics, we can use the occurring time of two spiking events and distance between two blocking locations. In both two tactics, the occurring time of two spiking events is very close. But the distance between two blocking locations in delayed spiking is shorter than alternate position spiking.

VII. EXPERIMENTS

For all experimental steps, we use AVI video sequences and implement the analysis in pixel domain. The resolution of all sequences is 720 × 480. We use three sets of different matches and extract 73 clips with obvious spiking events. There are 25 clips in set1, 25 clips in set2 and 23 clips in set3. The matches are recorded by university volleyball team of NCTU.

A. Results of Spiking Event Detection

Table I illustrates the performance of spiking event detection. The ground truth is the number of actual spiking events by observation. The retrieved candidates mean retrieved frames of spiking events based on our scheme. The number of correct jump frames represents the retrieved frames in ground truth. The precision of spiking event detection is 86.8% and the recall is 89.1%.

Error cases of incorrectly detected spiking events may be caused by ball and jump serve. The detection of ball while it passes through the analyzed area may lead to false alarm since we just use the moving pixels in the analysis. Since there is no depth information, the pattern between spiking and jump serve is hard to discriminate. Error cases of unsuccessfully detected spiking events may result from cloak and far-net-toss spiking. Cloak is a kind of tactics which a player pretends to spike in order to protect other spikers from being blocked by too many blockers. However, sometimes the player executing a cloak may not jump as high as a real spiking. As a result, some cloaks will not be detected because we only focus on the area higher than the net. Far-net-toss spiking is hard to detect since the ball is set to be far from the net.

B. Performance ofBlocking Pattern Classification

Based on the correctly detected spiking events, the experiment of blocking pattern classification can be made. From Table II, the accuracy of classification for ClassA (none/single) is about 81.4%, the accuracy of classification for ClassB (twice/triple) is 91.1% and overall accuracy of classification is about 84.7%. Error cases of blocking pattern classification may result from separated spikers and blockers. If the blockers are not aligned with the spiker, the width of blocking region will be too long to be correctly classified into the right class.

C. Results of Tactic Analysis

We manually select the clips containing two tactics from clips with correctly detected spiking events. In fact, both two kinds of tactics contain a cloak. But a cloak will not be detected if the height of jump is not enough. We check whether the clips with one of two tactics are correctly recognized. In our experiment, if the cloak is detected, the delayed spiking and alternate position spiking can be identified more precisely, as shown in Table III.

VIII. CONCLUSION

Little work on content analysis and event detection has been done for volleyball video because it is much complex to track the ball and players in volleyball video due to the high density of players on the court and the frequent ball-player overlaps. We propose a system that can automatically detect spiking and blocking events without ball tracking. Whistle detection is applied to perform video preprocessing. With extracted video clips, camera calibration technique is exploited to detect the court and the net. After locating the net, moving pixel analysis is applied to detect spiking events. And we analyze the spiking events for the purpose of finding blocking patterns. According to occurring time of spiking events and blocking locations, we can recognize two
important tactics, delayed spiking and alternate position spiking in volleyball matches.

ACKNOWLEDGMENT

This work is supported in part by "Aim for the Top University Plan" of the National Chiao Tung University and Ministry of Education, Taiwan, R.O.C., and in part by National Science Council of R.O.C. under the grant no. NSC 98-2221-E-009-091-MY3 and NSC 101-2218-E-009-004–.

REFERENCES