

Trajectory Based Event Tactics Analysis in Broadcast Sports Video

Guangyu Zhu^{1,2}, Qingming Huang^{2,5}, Changsheng Xu³, Yong Rui⁴, Shuqiang Jiang⁵, Wen Gao¹, Hongxun Yao¹

¹ School of Computer Science, Harbin Institute of Technology, P.R. China, 150001

² Graduate School of Chinese Academy of Sciences, P.R. China, 100080

³ Institute for Infocomm Research, Singapore, 119613

⁴ Microsoft Research Asia, P.R. China, 100080

⁵ Key Lab of Intell. Info. Processing, Institute of Computing Technology, Chinese Academy of Sciences, 100080

{gyzhu, qmhuang, sqjiang, wgao}@jdl.ac.cn, xucs@i2r.a-star.edu.sg, yongrui@microsoft.com, yhx@vilab.hit.edu.cn

ABSTRACT

Most of existing approaches on event detection in sports video are general audience oriented. The extracted events are then presented to the audience without further analysis. However, professionals, such as soccer coaches, are more interested in the tactics used in the events. In this paper, we present a novel approach to extract tactic information from the goal event in broadcast soccer video and present the goal event in a tactic mode to the coaches and sports professionals. We first extract goal events with far-view shots based on analysis and alignment of web-casting text and broadcast video. For a detected goal event, we employ a multi-object detection and tracking algorithm to obtain the players and ball trajectories in the shot. Compared with existing work, we proposed an effective tactic representation called aggregate trajectory which is constructed based on multiple trajectories using a novel analysis of temporal-spatial interaction among the players and the ball. The interactive relationship with play region information and hypothesis testing for trajectory temporal-spatial distribution are exploited to analyze the tactic patterns in a hierarchical coarse-to-fine framework. The experimental results on the data of FIFA World Cup 2006 are promising and demonstrate our approach is effective.

Categories and Subject Descriptors

I.4.8 [Scene Analysis]: *Motion, Object Recognition, Tracking.*

I.2.10 [Vision and Scene Understanding]: *Video Analysis.*

General Terms

Algorithms, Measurement, Performance, Experimentation.

Keywords

Event Detection, Broadcast Video, Object Tracking, Trajectory Analysis, Tactics Analysis.

1. INTRODUCTION

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

MM'07, September 23–28, 2007, Augsburg, Bavaria, Germany.

Copyright 2007 ACM 978-1-59593-701-8/07/0009...\$5.00.

With the explosive growth in the amount of video data and rapid advance in computing power, extensive research efforts have been devoted to content-based video analysis. As an important genre of video document, sports video has attracted increasing attention in automatic video analysis [1]-[11] due to its wide viewership and tremendous commercial potential. From a common audience's point of view, only the exciting portion in a sports game, which consists of the highlights or exciting events such as goal in soccer, is worthy for viewing. Therefore, the ability to access events or highlights from lengthy and voluminous sports video programs and to skip less interesting parts of the videos is of great value and highly demanded by the audience. Most technologies available for sports video analysis were focused on audience oriented event detection and highlight extraction [1][4]-[7][9]-[11].

With the extracted events, the existing approaches usually summarize and present them to the audience directly without any further analysis. Nevertheless, from the coach and sports professional point of view, they are more interested in the tactic patterns used by players in the events. For example with soccer, which is a very popular game in the field sports, there is a great interest from the coach in better understanding the process and performance pattern of attacks or goals so that he/she is able to increase the performance of the team during the game and better adapt the training plan. To achieve this purpose, it is common nowadays for technical staff to employ one person to record matches and prepare video sessions. For example, during the Olympic Games 2004, the Spanish basketball team employed one person who analyzed and prepared three matches per day in order to obtain a five-minute summary to show to the players. The system used in this case consisted of two computers that allowed editing and saving important tactic situations and plays. If we apply the similar scenario for soccer games, it is obviously labor-intensive and time-consuming. Consequently, there exists a compelling case for research on automatic tactics analysis of soccer game. However, available related work in the literature is rather short [16][17][18].

In this paper, we propose a novel approach to discover the tactic patterns from the goal events in broadcast soccer videos based on the tactic clues extracted from players and ball trajectories. As the salient objects in soccer games, the movement of players and ball is the type of useful information for tactic analysis. The trajectories of players and ball, as a function of time, can contribute to identify the patterns of tactic scenario during the games and improving the performance of players at different positions [12]. On the other hand, broadcast video is a post-edited video where the broadcast sequence feed is selected from frequent switches among

multiple cameras according to the broadcast director's instruction. Currently, it is widely used for broadcasting of sports games. The analysis using broadcast sports video has the advantages of extensive source and easy acquisition.

The novelty and contribution of this paper are summarized as follows. 1) We exploit an effective goal event extraction method proposed in [7] based on the combination of analysis and alignment of web-casting text and broadcast soccer video. 2) A sophisticated tracking strategy is applied to players and ball detection and tracking with the integration of particle filter and support vector machine. 3) A novel tactic representation called aggregate trajectory is proposed, which is constructed based on the multiple trajectories according to the analysis of temporal and spatial interaction of players and ball. 4) We propose a hierarchical coarse-to-fine framework to identify the tactic strategies of goal attacks including route patterns and interaction patterns. The inference using play region, the parsing by temporal-spatial objects interaction and the hypothesis testing for distribution of aggregate trajectory are employed in the analysis, respectively. 5) Compared with existing work [16][17][18], our approach is implemented on broadcast video, which is more widely used in real applications.

The rest of the paper is organized as follows. Section 2 reviews related work. Section 4 presents the approaches of goal event extraction using the multimodal method with web-casting text analysis and game time recognition. In section 5 we introduce the method to construct tactic representation and extract tactic information from goal event in broadcast soccer video. Section 6 describes the detail of analysis and presentation of tactic patterns for soccer game based on extracted tactic information. Experiments are reported and analyzed in section 8. Finally, we conclude the paper with future work in section 9.

2. RELATED WORK

2.1 Soccer Video Analysis

Extensive research efforts have been devoted to soccer video analysis. The existing approaches can be reviewed according to three hierarchical layers for the current research routine which are low-level, mid-level, and high-level analysis respectively.

Low-level processing based analysis: Low-level features can be extracted from different modalities (e.g. audio and visual) in the video and can be used separately or combinatorially. For example, visual features including dominant color ratio, shot boundary, and aspect ratio of referee region were employed to accomplish various events detection and summarization for soccer games [2]. In [3], a salient feature set selected from compressed domain, dominant color ratio and motion intensity was employed for structure analysis of broadcast soccer video. However, it is difficult to robustly and accurately detect events using low-level features only due to the semantic gap between low-level features and high-level concepts.

Mid-level representation based analysis: Generic low-level features only deal with representing perceived content, but not with semantics. To bridge the semantic gap between low-level features and high-level concepts, one possible solution is to introduce a middle level representation. In [10], the definition of mid-level representation was introduced in the framework of event detection for soccer video. Various approaches have been proposed from the aspects of visual [10], audio [11] and text [6][7], respectively.

High-level analysis based on multimodality: The integrated use of various information sources is the trend in high-level analysis of soccer video. With the enhancement of more information available, the result of video semantic analysis will be improved. Two categories can be differentiated for high-level analysis: rule-based and model-based. For example, heuristic rules were exploited to detect the goal event [2]. Beyond heuristic rules, statistic models are frequently used for multimodal analysis, e.g. hidden Markov model [3], finite state machine [5], and support vector machine [10]. Different from semantic analysis, affective content understanding of semantic events brings another dimension to content-based soccer video browsing [9].

2.2 Tactics Analysis for Sports Game

Tactics analysis aims understanding the tactics that the teams or individual players have used in the games. Coaches and players are interested in such results for improving their performance in later games, and sports professionals are interested in such results for enjoying sports video with additional statistical information.

Most existing approaches for tactics analysis are focused on tennis sports because its game field is relatively small and the number of players is small. The main techniques employed are tracking the trajectories of the players and ball with the assistance of domain knowledge [13]-[15] and recognizing the actions of the players [8]. Sudhir *et al.* [13] exploited the domain knowledge of tennis video to develop a court line detection algorithm and a player tracking algorithm to identify tactics-related events. In [14], the authors attempted to classify tennis games into 58 winning patterns for training purpose based on tracking ball movement from broadcast video. In [15], Wang *et al.* presented a novel approach for tennis video indexing by mining the salient technical patterns in the match process. Unlike trajectory-based algorithm, a novel action-driven tactics analysis approach is proposed in [8] for tennis game which is able to discover the insight of the stroke performance of the players.

Few works [16][17][18] attempted to perform tactics analysis for soccer games. In earlier work [16], the players' positions were estimated from the soccer game image sequence which was captured by multiple static cameras system, and then were transformed to real soccer field space using camera calibration technique. By introducing the notion of minimum moving time pattern and dominant region of a player, the strategic ability of soccer team was evaluated. In [17], a study was presented on the discovery of meaningful pass patterns and sequences from time-series soccer record data. An evaluation model was proposed in [18] to quantitatively express the performance of soccer players, using as input the relationships between the trajectories of 22 players and a ball and having as output the performance evaluation of several players in a quantitative way. However, the existing work was based on the human-labeled [17] and computer-simulated [18] trajectory data which has strong limitation and less challenge for object tracking and pattern discovery using broadcast video.

3. OUR APPROACH

Existing approaches for soccer video analysis were mostly for event-driven indexing of video content, which cannot provide detailed tactic information used in the game. Little work on tactics analysis used non-broadcast video [17][18] and had to make the camera calibration [16], which cannot be adapted to wide applica-

tions. Here, we propose a novel tactic analysis approach for the goal event in broadcast soccer game.

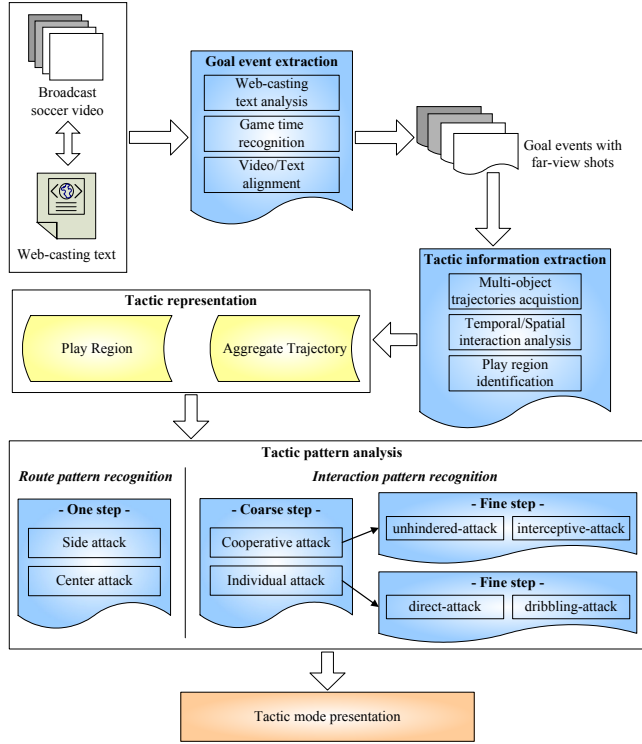


Figure 1. Flow diagram of tactics analysis for the goal event in broadcast soccer game.

Table 1. Description of four interaction patterns

Coarse pattern	Fine pattern	Description
cooperative attack	un hindered-attack	No ball intercepted by defender in goal attack
	interceptive-attack	Ball intercepted by one or more defenders in goal attack
individual attack	direct-attack	No ball dribbling in goal event (e.g. penalty kick)
	dribbling-attack	Ball dribbling by attacker before shot-on-goal

Figure 1 illustrates the flowchart of our approach. 1) Using the multimodal method with web-casting text analysis and game time recognition, we accurately detect the goal events and extract the far-view shots in the broadcast soccer video. Far-view shots present the entire process of goal event and are easily used for the following object detection and tracking. 2) As the fundamental tactic information extraction, multi-object detection and tracking is employed to obtain the players and ball trajectories in the goal events. Mosaic trajectories are computed using global motion estimation based on the positions of players and ball tracked in the frames. 3) For a goal event, two tactic representations are constructed including aggregate trajectory and play region which are based on temporal-spatial interaction among the players and ball and field region viewing by the camera. 4) Tactic analysis for the goal event is achieved by recognizing strategic patterns composed of route pattern and interaction pattern. The play regions are used to deduce the route pattern, e.g. side-attack, center-attack.

Interaction pattern recognition is formatted into a hierarchical coarse-to-fine scheme based on aggregate trajectory. At the coarse level, the goals are classified into cooperative pattern which the goal is scored by multiple players via ball passing, and individual pattern which the goal is scored by only one player. The two coarse patterns are then classified into four elaborated scenarios: un hindered-attack and interceptive-attack for cooperative pattern, direct-attack and dribbling-attack for individual pattern. The detailed description of four tactic patterns is listed in Table 1. 5) The classified patterns of goal events are finally presented to the professional users in a tactic context mode.

4. GOAL EVENT EXTRACTION

Existing methods for event detection in broadcast sports video are mostly based on audio/visual/textual features directly extracted from video content itself [1][5][6][9]-[11]. Due to the semantic gap between low-level features and high-level events, it is difficult to use these methods to identify the event semantics, detect exact event boundaries and perform robustly for various broadcast data resulting in ideal extraction accuracy. Contrast with detection based on video only, external text were used for sports video analysis [6][7]. Here, we adopt an effective detection method to extract the goal events from broadcast soccer video by combining analysis and alignment of web-casting text and video content [7], which has the advantage of high effectiveness and accuracy.

4.1 Web-Casting Text Analysis

The web-casting text [19] serves as text broadcasting for sports games. As shown in Figure 2, the text describes the event happened in a game with a time stamp and brief description such as type and development of event, etc., which are very difficult to be obtained directly from the video using previous approaches.

17:43 Lilian Thuram Clear
 17:48 Fabio Grosso Throw In -Attacking
 18:11 Andrea Pirlo Corner Kick -Outswinger -Shot
 18:42 Marco Materazzi Shot On Goal -Normal -Goal
 18:42 Marco Materazzi Goal - Headed in
 18:54 Andrea Pirlo Assist -Cross
 19:37 Zinedine Zidane Foul -Free Kick

Figure 2. Example for web-casting text.

To detect events from web-casting text, we first observe from our database that each type of the sports event features one or several unique nouns, which are defined as keywords related to this event. This is because the web-casting text is tagged by sports professionals and has fixed structures. Hence by detecting these keywords, the relevant event can be recognized. To extract goal event, we define two sets of keywords for well-structured web text and free-style web text, respectively, as shown in Table 2.

Table 2. Keyword definition for goal event

Event	Keyword
Goal (well-structured)	goal, scored
Goal (free-style)	g-o-a-l or goal or equalize - kick

Once proper keywords are defined, a goal event can be detected by finding the sentences that contain the relevant keywords and analyzing context information before and after the keywords. A simple keyword based text search technique is enough for this

work. Exemplifying with the text shown in Figure 2, we can detect a goal event happened at the game time of 18:42.

4.2 Game Time Recognition

In many broadcast videos of soccer game, a video clock is used to indicate the game lapsed time. Since the time stamp in the text event is associated with the game time, knowing the clock time will help us to locate the event moment in the video. Referring to the event moment, the event boundary can be detected.



Figure 3. Overlaid video clock.

As shown in Figure 3, the digital clock is overlaid on the video with other texts such as the team names and the scores. We exploit a novel approach to read the video clock by recognizing the clock digits using a few techniques related to the transition patterns of the clock [20]. Compared with the traditional methods, this approach is able to achieve the real time performance and the result is more reliable. Temporal neighboring pattern similarity (TNPS) is used as the most critical feature. If TNPS pattern change is in the time-changing regulation, the character is considered as a clock digit character. After the clock digits are located, we observe the TEN-SECOND digit pattern change using the TNPS to automatically obtain the pattern templates of digit “0” to “9”. Since the extracted digits may vary along time due to the low quality of the video, we extract a few templates for the same digit character. For every frame, each clock digit is matched against the templates and recognized with a best match. The detail of game time recognition can be found in [20].

4.3 Video/Text Alignment

After knowing the game time, we can detect the event moment in the video by linking the time stamp in the text event to the game time in the video. In order to extract the entire event from the video, we use video structure analysis and finite state machine (FSM) to detect the event boundaries.

Based on the detected event moment in the video, we define a temporal range containing the event moment and detect event boundary within this range. The basic idea is first to locate the clusters of successive gradual shot transitions in the video as candidate segment boundaries for significant game moments and then decide the accurate event boundaries using finite state machine. In the approach, the mean absolute differences (MAD) of successive frame gray level pixels are computed as the features of abrupt shot changes as well as the multiple pair-wise MAD is used for gradual shot change. With the obtained shot boundary, shot classification is conducted using a majority voting of frame view types identified within a single shot to generate a classification sequence S as

$$S = \{S_i \mid S_i = \langle sbt_i, st_i, ebt_i \rangle, i = 1, \dots, N\} \quad (1)$$

where $sbt_i \in \{\text{cut}, \text{dissolve}\}$ is the start boundary type, $st_i \in \{\text{far-view}, \text{non-far-view}\}$ is the shot type, $ebt_i \in \{\text{cut}, \text{dissolve}\}$ is the

end boundary type, and N is the total number of shots. The view type is identified using the dominant field color, which the field color is dominant in far-view and is not dominant in non-far-view contrastively. Once the shot classification sequence S is generated, finite state machine is employed to detect the event boundaries. FSM has been proved to be robust in modeling temporal transition patterns and has the advantage of without training process. More detail of event boundary detection can be further consulted in [7].

5. TACTIC INFORMATION EXTRACTION AND REPRESENTATION

Similar with semantic analysis in which the semantic representation is constructed from video content, we need to extract proper clues from video and construct an effective representation to discover the deep insight of the soccer game in the tactic context. As the team sports, the tactics used in the soccer game is characterized by the behavior of individual player (e.g. positions of the player in the field) and the interactions among players and ball (e.g. ball passing from one player to other player). The trajectories of players and ball can right reflect such characterization, in which we can locate players and analyze their mutual relationship. Based on extracted multiple trajectories, we construct the novel tactic representations including aggregate trajectory and play region for the soccer game in broadcast video.

5.1 Multi-Object Trajectories Acquisition

5.1.1 Player Detection and Tracking

The acquisition of players’ trajectories in the far-view shots is achieved by object detection and tracking algorithm proposed in [21]. The flowchart is shown in Figure 4. Playfield detection is first performed using Gaussian mixture color models with the evidence that the playfield pixels are the dominant components in most of the frames. The regions inside the extracted field are considered as the player candidates. Then, recognition module based on support vector classification (SVC) is employed to eliminate the non-player candidates. For each of selected player regions, if it is identified as a new appeared player, a tracker is assigned. A filtering based tracker called support vector regression (SVR) particle filter keeps tracking player in the frames. After each tracking interaction, the player disappearance module using SVC recognition model evaluates whether the current tracked player leaves the scene. If it does, the corresponding tracker is released. During the tracking process, the color histogram of target region is employed to identify the team affiliation of tracked player.

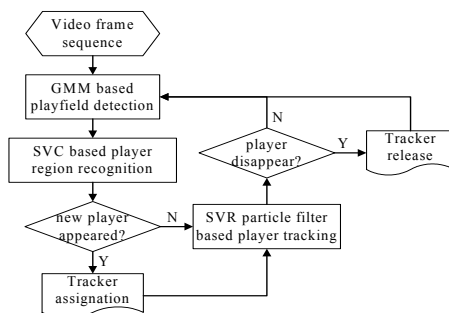


Figure 4. Diagram of player detection and tracking.

For the tracking algorithm based on particle filter, the key points are the likelihood computation of the different hypothesis obser-

vations and the high computational intensity with large number of samples. SVR particle filter integrates support vector regression into sequential Monte Carlo framework to solve these problems and has been demonstrated to be effective [21].

5.1.2 Ball Detection and Tracking

The challenges associated with ball detection and tracking can be attributed to the following factors: 1) the ball's attributes (color, shape, size and velocity etc.) change over frame, 2) the ball becomes a long blurred strip when it moves fast, 3) the ball is sometimes occluded by players, merged with lines, or hidden in the auditorium, and 4) many other objects are similar to the ball.

To solve the above-mentioned challenges, a new method is proposed by enhancing our previous work [22]. Figure 5 illustrates the diagram of our method. It is composed of two alternate procedures including detection and tracking. For ball detection, color, shape and size information are first used to extract candidate regions in each frame. Then, a weighted graph is constructed with each node representing a candidate and each edge linking two candidates in the adjacent frames. Viterbi algorithm is applied to extract the optimal path which is the most likely to be ball path and locations. Such method can enhance the robustness of ball detection because it holds multiple hypotheses of ball locations. Once the ball is detected, the tracking procedure based on SVR particle filter is started. In each frame, ball location is verified to update the template to check whether the ball is lost. If the ball is lost, the detection procedure runs again.

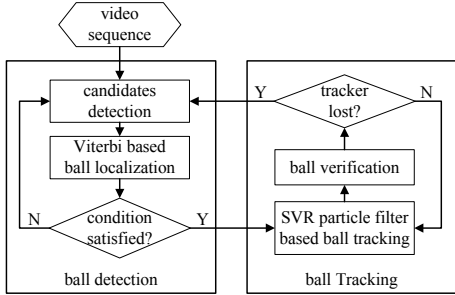


Figure 5. Diagram of ball detection and tracking.

5.2 Aggregate Trajectory Computation

In previous approaches, the tactic clues were extracted from the trajectories labeled by human and computer simulation. These kinds of data are the locations of the players in the real soccer field with the result that interaction analysis can be easily conducted to reflect the cooperation among the players. However, broadcast video normally comprises frequent camera motions in the sequence leading to more challenges for interaction analysis. The objective of aggregate trajectory computation is to construct a compact representation in the tactic context for the broadcast soccer video using mosaic technique and temporal-spatial analysis.

5.2.1 Mosaic Trajectory Computation

Mosaic trajectories are the transform of the original ones being warped into a common coordinate system with the result of elimination of camera motions based on global motion estimation.

Global motion estimation (GME) [23] is used to establish the mapping relationship between the spatial coordinates in two suc-

cessive frames. Using homogeneous representation, $\mathbf{x} = (x, y, w)^T$ represents a point $(x/w, y/w)^T$ in Euclidean \mathbf{R}^2 space. Given two points \mathbf{x}_t and \mathbf{x}_{t-1} , where \mathbf{x}_t denotes the coordinate in frame t and \mathbf{x}_{t-1} denotes the corresponding coordinate in frame $t-1$, the mapping between \mathbf{x}_t and \mathbf{x}_{t-1} is represented as

$$\mathbf{x}_{t-1} = \mathbf{H}_{t,t-1} \cdot \mathbf{x}_t \quad (2)$$

where $\mathbf{H}_{t,t-1}$ is the mapping matrix from frame t to frame $t-1$ obtained by global motion estimation.

Given the video sequence of a goal event $V = \{f_1, \dots, f_n\}$ where f_i is the i th frame and n is the total number of the frames, one trajectory in the goal event is $T = \{p_1, \dots, p_n\}$ where p_i is the position of the object located in frame f_i . Considering Eq. (2) and temporal relation of the frames, we can therefore warp each p_i into the uniform coordinate of frame f_1 as following:

$$\tilde{\mathbf{p}}_i = \prod_{t=2}^i \mathbf{H}_{t,t-1} \cdot \mathbf{p}_i \quad (3)$$

where $\tilde{\mathbf{p}}_i$ is the mapping position of \mathbf{p}_i , both of which are represented in homogenous coordinate for p_i . Consequently, all the trajectories in the goal event are warped into the coordinate space of frame f_1 which is essentially a common coordinate system.

Once the mosaic trajectories are computed, the motion in broadcast video caused by camera behavior can be treated as being removed. The mosaic trajectories correspond to the loci of the players and ball captured by a static camera.

5.2.2 Temporal and Spatial Interaction Analysis

The insight of aggregate trajectory is to capture the interaction relationship among players and ball in a compact representation. Ball trajectory is the major consideration in the analysis because all the tactic strategies in the soccer game will be finally conducted on the ball. The most important interactions among the players and ball are ball-passing and ball-dribbling. Our temporal-spatial interaction analysis is to select the segments of passing (which correspond to the ball trajectories) and dribbling (which correspond to the dribbling-player trajectories) from the mosaic trajectories and then concatenate the selected segments into a new locus representation called aggregate trajectory.

Local temporal-spatial analysis is carried out on the segmented temporal intervals of the whole goal event. In each temporal interval, the inclusive trajectory segments of players are employed as well as the ball trajectory. Figure 6 shows the flowchart of aggregate trajectory generation based on interaction analysis in terms of temporal and spatial aspects.

Let us denote the set of mosaic trajectories for a given goal event is $MT = \{l_b(t), l_{p_1}(t), \dots, l_{p_n}(t)\}$ where $l_b(t)$ is the trajectory of ball and $l_{p_i}(t)$ is the one of the i th player, t represents that each element is the time-series data. For each trajectory $l(t)$ in MT , Gaussian filter is first applied to eliminate the noise in the trajectory. Then, the trajectories are uniformly partitioned into the segments according to the equal temporal interval τ (e.g. $\tau = 2s$). As shown in Figure 7, the distance between the dribbling-player and ball is nearer in a temporal interval of the process of ball-dribbling. Moreover, the shape of trajectories of player and ball is similar because the player and ball are followed the similar route on the field in this interval. Otherwise, such observation does not

guarantee under the condition that the player passed the ball or the player uncontrolled the ball.

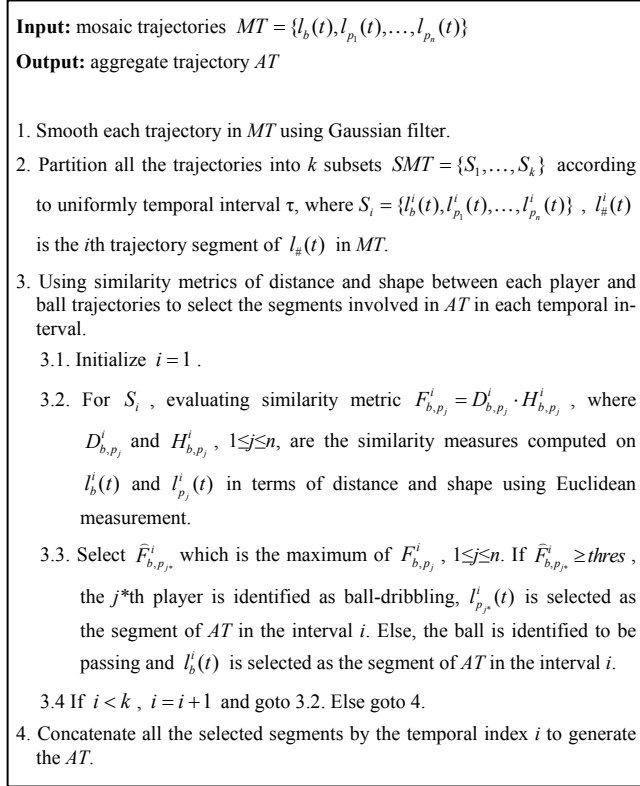


Figure 6. Flowchart of aggregate trajectory generation.

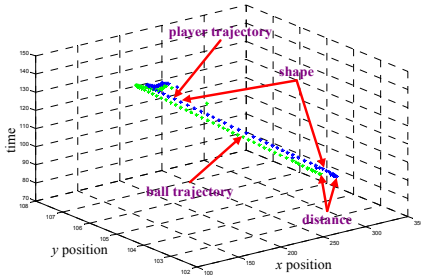


Figure 7. Trajectories of player and ball for ball-dribbling.

For temporal interval i , we define two similarity measures D_{b,p_j}^i and H_{b,p_j}^i for the trajectory segments of the ball and the j th player in terms of distance and shape respectively, $1 \leq j \leq n$ where n is the number of objects. Given $l_b^i = \{\mathbf{u}_1, \dots, \mathbf{u}_m\}$ and $l_{p_j}^i = \{\mathbf{v}_1, \dots, \mathbf{v}_m\}$ where \mathbf{u} and \mathbf{v} represent the object position in the trajectory, D_{b,p_j}^i is defined as

$$D_{b,p_j}^i = \exp \left\{ -\frac{1}{m} \sum_{k=1}^m \|\mathbf{u}_k - \mathbf{v}_k\| \right\} \quad (4)$$

where $\|\mathbf{u} - \mathbf{v}\|$ is the Euclidean distance for \mathbf{u} and \mathbf{v} in \mathbf{R}^2 space. H_{b,p_j}^i is computed on the spatial curvature of 2-D curve given by

$$c^i(k) = \frac{x'(k) \cdot y''(k) - y'(k) \cdot x''(k)}{[x'(k)^2 + y'(k)^2]^{3/2}} \quad (5)$$

where x and y are the X - and Y -axes projections of the point k in the trajectory, x' , x'' , y' , and y'' are the first- and second-order derivatives of x and y by t , respectively. According to Eq. (5), we can calculate the curvature sequences $c_b^i = \{a_1, \dots, a_m\}$ and $c_{p_j}^i = \{b_1, \dots, b_m\}$ for l_b^i and $l_{p_j}^i$. The H_{b,p_j}^i is then computed as

$$H_{b,p_j}^i = \exp \left\{ -\frac{1}{m} \sum_{k=1}^m |a_k - b_k| \right\} \quad (6)$$

where $|x|$ means the absolute value of x .

Using D_{b,p_j}^i and H_{b,p_j}^i , we define the similarity metric F_{b,p_j}^i as

$$F_{b,p_j}^i = D_{b,p_j}^i \cdot H_{b,p_j}^i \quad (7)$$

and obtain the $\hat{F}_{b,p_{j^*}}^i$ which is the maximum of all the F_{b,p_j}^i , $1 \leq j \leq n$. If $\hat{F}_{b,p_{j^*}}^i \geq thres$ where $thres$ is the predefined threshold, the j^* th player is identified as being dribbling the ball and the trajectory segment $l_{p_{j^*}}^i$ is selected as the segment of aggregate trajectory in the temporal interval i . Otherwise, the ball trajectory segment l_b^i is selected. The $thres$ is empirically set to be 0.23.

With all the selected trajectory segments, the aggregate trajectory is generated by concatenating the segments according to the order of corresponding temporal indexes. Note that if the aggregate trajectory is ended by the ball trajectory, we ignore this last ball trajectory and delete it from aggregate trajectory. This is because the last ball trajectory represents the locus of the ball conducted by the attacker with shot-on-goal action. It is different from the ball trajectory segment which has both the sender and receiver. It only has the sender which therefore does not reflect the interaction relationship among the players. Figure 8 shows an example of aggregate trajectory.

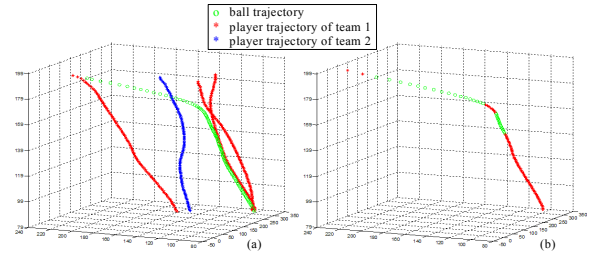


Figure 8. Aggregate trajectory generation. (a) Mosaic trajectories of players and ball in a goal event, (b) Aggregate trajectory generated by temporal-spatial interaction analysis.

5.3 Play Region Identification

Play region is one of the crucial factors for attack pattern identification. In our implementation [24], the field is divided into 15 areas as shown in Figure 9(a). Symmetrical regions in the field are given the same labels resulting in six labels in Figure 9(b).

We extract following three features for the identification. 1) Field line location which is represented in polar coordinates (ρ_i, θ_i) $i = 1, \dots, N$ where ρ_i and θ_i are the i th radial and angular coordinates respectively and N is the total number of lines. 2) Goal-mouth location which is represented by the central point (x_g, y_g) where x_g and y_g are the X - and Y -axes coordinates. 3) Central cir-

cle location which is represented by the central point (x_e, y_e) where x_e and y_e are the X - and Y -axes coordinates.

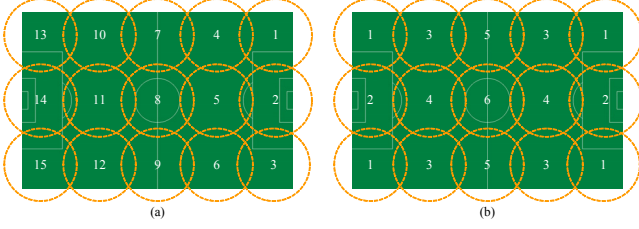


Figure 9. Play region model. (a) 15 areas, (b) 6 labels.

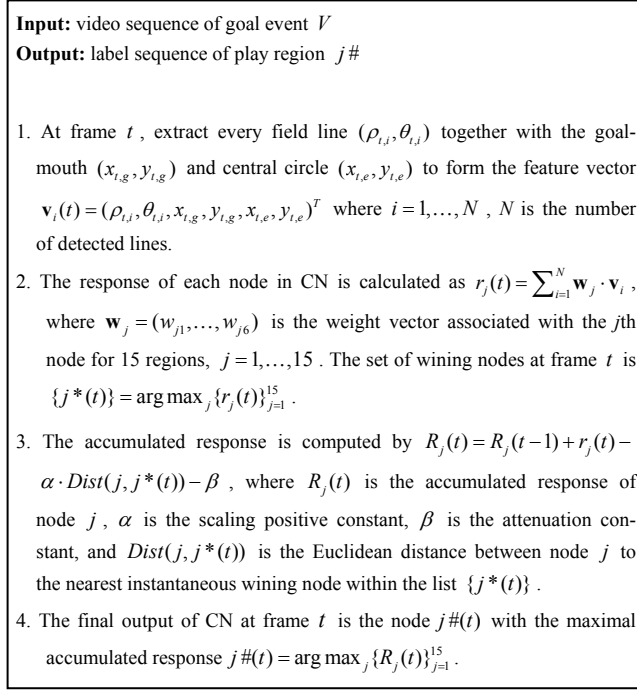


Figure 10. Play region identification using competition network.

To detect the play region, we employ a competition network (CN) using the three shape features described above. The CN consists of 15 dependent classifier nodes, each node representing one area of the field. The 15 nodes compete among each other, and the accumulated winning node is identified as the play region. The input of the CN is the individual frames in the goal event. The operation manner of CN is shown in the Figure 10. The input of the identification method is the video sequence and the output is the labeled sequence of the play regions.

6. TACTIC PATTERN ANALYSIS

The significant contribution of our tactic analysis approach is that the tactic scenarios are categorized into two independent groups: route pattern and interaction pattern. Route pattern describes the attack routine implemented by the players and ball to achieve a goal, while interaction pattern represents the interactive relationship among the players to complete a goal attack. Using the tactic representation including play region and aggregate trajectory, we conduct the tactic pattern analysis.

6.1 Route Pattern Recognition

Route pattern recognition is achieved by the inference using play regions identified from the video sequence of the goal event. The pattern categories consist of side attack and center attack.

Given f_i which is the i th frame in the video sequence V of the goal event G , the corresponding play region is r_i . The vote that f_i contributes to G for the pattern classification is defined as

$$\text{Vote}(f_i) = \begin{cases} 1 & \text{if } \text{Re}g(r_i) = \text{side-attack} \\ -1 & \text{if } \text{Re}g(r_i) = \text{center-attack} \end{cases} \quad (8)$$

where $\text{Re}g(\bullet)$ is the function for the pattern classification for a play region based on region label shown in Figure 9(b)

$$\text{Re}g(r_i) = \begin{cases} \text{side-attack} & \text{if } r_i = 1, 3, 5 \\ \text{center-attack} & \text{if } r_i = 2, 4, 6 \end{cases} \quad (9)$$

The final route pattern RP of the goal event G is determined as

$$RP(G) = \begin{cases} \text{side-attack} & \text{if } \sum_{f_i \in V} \text{Vote}(f_i) \geq 0 \\ \text{center-attack} & \text{if } \sum_{f_i \in V} \text{Vote}(f_i) < 0 \end{cases} \quad (10)$$

Because there are two pattern categories, the equal sign is just assigned for the side attack so as to avoid the occurrence of marginal classification because there are more side attacks than center attacks in the soccer game by our observation.

6.2 Interaction Pattern Recognition

To more effectively capture the insight of a goal attack, the interaction pattern recognition is hierarchized into a coarse-to-fine structure. As shown in Figure 1, two coarse categories are first classified and four patterns are then identified elaborately.

6.2.1 Coarse Analysis for Pattern Recognition

At coarse step, the interaction patterns are first classified into two categories including cooperative pattern and individual pattern. The cooperative pattern is defined as the scenario in the goal event which the goal is scored by multiple players via ball passing, and the individual pattern is defined as the scenario which the goal is scored by only one player. The recognition is conducted on the aggregate trajectory of the goal event.

Given the aggregate trajectory $AT = \{s_1, \dots, s_n\}$ computed for the goal event G where s_i is the trajectory segment, we can define the criteria C_{coarse} for the coarse classification as

$$C_{coarse}(G) = \sum_{i=1}^n \text{ball}(s_i) \quad (11)$$

where function $\text{ball}(\bullet)$ is defined as

$$\text{ball}(x) = \begin{cases} 1 & \text{if } x \text{ is the segment of ball trajectory} \\ 0 & \text{if } x \text{ is not the segment of ball trajectory} \end{cases} \quad (12)$$

Consequently, we can classify the interaction pattern of the goal event at the coarse level as following.

$$IP_{coarse}(G) = \begin{cases} \text{cooperative-attack} & \text{if } C_{coarse}(G) > 0 \\ \text{individual-attack} & \text{if } C_{coarse}(G) = 0 \end{cases} \quad (13)$$

6.2.2 Fine Analysis for Pattern Recognition

More elaborated interaction patterns are discovered at the fine level for cooperative attack and individual attack. The definitions of the fine patterns to be recognized are listed in Table 1.

6.2.2.1 Recognition for Cooperative Attack

For cooperative attack, we categorize the patterns into unhindered-attack and interceptive-attack according that whether there is ball-interception during the process of goal attack. Given the aggregate trajectory AT of a cooperative attack classified at the coarse step, the subset $SAT = \{sp_1, \dots, sp_m\}$ of AT was extracted which only consists of player trajectories where sp_i is the trajectory segments. The elaborated criteria for cooperative attack recognition at the fine step is defined as

$$C_{fine_c}(G) = \sum_{i=2}^m \{1 - \delta[\text{player}(sp_i) - \text{player}(sp_{i-1})]\} \quad (14)$$

where δ is the Kronecker delta function, $\text{player}(x)$ is the function to identify the team affiliation (team 1 or team 2) for segment x . Therefore, we can finely classify the cooperative attack as

$$IP_{fine_c}(G) = \begin{cases} \text{unhindered-attack} & \text{if } C_{fine_c}(G) = 0 \\ \text{interceptive-attack} & \text{if } C_{fine_c}(G) > 0 \end{cases} \quad (15)$$

6.2.2.2 Recognition for Individual Attack

For individual attack, direct-attack and dribbling-attack are discovered in terms of the ball-dribbling occurrence in the attack process. Direct-attack pattern mainly corresponds to the penalty kick, while dribble-attack corresponds to the goal with ball-dribbling. To differentiate two patterns, hypothesis testing is conducted on the spatial distribution of aggregate trajectory.

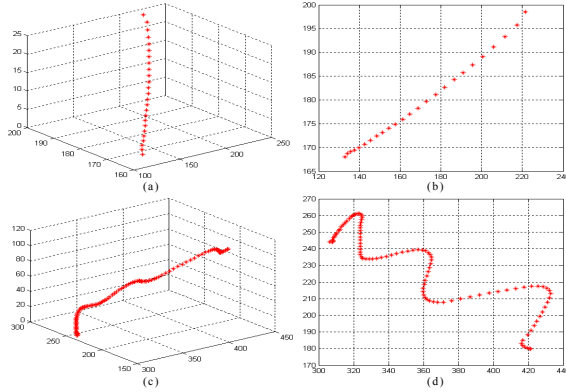


Figure 11. Spatial distribution of aggregate trajectory. (a) and (c) are the 3D distribution of AT of direct-attack and dribbling-attack respectively, (b) and (d) are the corresponding 2D projected distribution.

By the observation as shown in Figure 11(a) and (c), the spatial positions of the trajectory of direct-attack subject to a line distribution compared with dribbling-attack. This observation is further verified by projecting trajectory into the 2D space as shown in Figure 11(b) and (d). Such evidence is easily demonstrated by the process of two patterns in real game. For penalty kick, the player usually runs a short distance directly to the ball and shot-on-goal with the result that the trajectory in the spatial space is approximately a line. However, the player has to dribble the ball to avoid the interception from the defensive players leading to a flexuous trajectory for dribbling-attack. Our hypothesis testing based approach classifies the two patterns.

Given the aggregate trajectory $AT = \{s_1, \dots, s_n\}$, we perform the accumulative error test to determine whether the spatial distribution of AT is similar to a line. Therefore, we have two hypotheses:

$$\begin{aligned} H_0 &: f(X, Y | k, c) = 0 \\ H_1 &: f(X, Y | k, c) \neq 0 \end{aligned} \quad (16)$$

where $X = \{x_1, \dots, x_n\}$ and $Y = \{y_1, \dots, y_n\}$ are the set of X - and Y -axes projections of the points in the trajectory which $s_i = (x_i, y_i)$, k and c are the parameters for the fitting line of the underlying trajectory data. We first use the least square method [25] to estimate the fitting line function $y = L(x | k, c)$. Then, the average accumulative error (AE) for the given AT is calculated as

$$AE = \sum_{i=1}^n [y_i - L(x_i | k, c)] / n \quad (17)$$

According to Eq. (17), AE is larger when X and Y are not fitting to a line distribution. Thus, we can classify the individual attack as

$$IP_{fine_i}(G) = \begin{cases} \text{direct-attack} & \text{if } AE \leq \text{thres} \\ \text{dribbling-attack} & \text{if } AE > \text{thres} \end{cases} \quad (18)$$

where thres is the predefined error threshold which is set to be 5 empirically in the experiments.

7. TACTIC MODE PRESENTATION

With the analyzed results, two problems are considered for tactic pattern presentation which are 1) the presentation should be provided clearly and concisely so that the viewers can easily know about the tactic strategies used in the game, 2) the presentation should provide essentially usable information so that the professionals can make further strategy analysis according to the personalized requirement.

The following information extracted from our analysis approach is selected for the presentation in the tactic mode.

- *Time stamp for goal occurrence* which is obtained from the web-casting text analysis.
- *Team labels in terms of offensive and defensive* which is analyzed from web-casting text.
- *Trajectories of ball, offensive player and defensive player respectively* which are extracted by multiple objects detection and tracking.
- *Route pattern (side- or center-)* which is recognized by route pattern recognition.
- *Interaction pattern (two categories at coarse and fine levels respectively)* classified by interaction pattern recognition.

The basic principle of information selection is to present comprehensive summary for the game in the tactic context.

8. EXPERIMENTAL RESULTS

To demonstrate the effectiveness of the proposed approach, we carried out the experiments on the video data of FIFA World Cup 2006. The test videos including all the 64 matches were recorded from live broadcast television program and compressed in MPEG-2 Video standard with the frame resolution of 704×576. The goal events involved in the video data consist of nearly all the tactic strategies which are adopted in nowadays soccer games. The goals were automatically extracted from the videos using the event detection method based on analysis and alignment of web-casting text and broadcast video. All the 168 goal events were extracted from the videos. Compared with manually labeled ground truth, our approach achieved 92% accuracy for the event boundary detection evaluated by the BDA metric proposed in [7].

All the 168 goals were applied to the experiments. Before conducting the automatic analysis, we invited three soccer professionals who are well familiar with the soccer tactics to annotate the tactic patterns for all the goals by voting scheme. The manual annotation was adopted as the ground truth for the comparison with the result of automatic analysis. Table 3 shows the result of manual annotation.

Table 3. Tactic patterns labeled by manual annotation

Tactic Pattern			Number
Route Pattern	side-attack		96
	center-attack		72
Interaction Pattern	cooperative attack	unhindered-attack	95
		interceptive-attack	16
	individual attack	direct-attack	41
		dribbling-attack	16

8.1 Route Pattern Recognition Results

Using the proposed route pattern recognition, we classified 168 goals into two clusters. We calculated Recall (R) and Precision (P) to quantitatively evaluate the performance, which are defined as

$$R = n_c / (n_c + n_m) \quad (19)$$

$$P = n_c / (n_c + n_f) \quad (20)$$

where for each pattern, n_c is the number of goals correctly recognized, n_m is the number of missed goals, and n_f is the number of goals false-alarmed. Table 4 shows the analysis results.

Table 4. Results of route pattern recognition

Pattern	n_c	n_m	n_f	R (%)	P (%)
side-attack	82	14	11	85.4	88.2
center-attack	61	11	14	84.7	81.3

It can be seen that the results for proposed approach are satisfactory. The false classifications are mainly due to the low detection accuracy for field area 3 and 4 (Figure 9(b)) because there are not enough field lines, goalmouth or central circle in these areas. This lack of distinct information thus results in poor accuracy.

8.2 Interaction Pattern Recognition Results

The interaction pattern recognition performance is evaluated using all the goals. Multiple trajectories of players and ball were extracted to construct the aggregate trajectory. The coarse and fine criteria were computed according to Eq. (13), (15), and (18). The metrics R and P defined in Eq. (19) and (20) were used to evaluate the performance. The results for the classification of four tactic patterns are listed in Table 5.

Table 5. Results of interaction pattern recognition

Pattern	n_c	n_m	n_f	R (%)	P (%)
unhindered-attack	81	14	12	85.3	87.1
interceptive-attack	11	5	3	68.8	78.6
direct-attack	33	8	6	80.5	84.6
dribbling-attack	13	3	8	81.3	61.9

It is observed from Table 5 that the performance of our tactic analysis approach is promising. The key issues affecting the results can be summarized in three ways. 1) Robustness of multiple objects detection and tracking: As noted in the previous section, multiple objects detection and tracking is the fundamental task to obtain the trajectories for the tactic representation construction. From the computer vision point of view, many factors will affect the result of detection and tracking, e.g. the occlusion among the objects. This can be demonstrated by the low recall of interceptive-attack pattern because there are severe occlusions due to the defensive tackle and body check. 2) The accumulative error of mosaic transform: Mosaic trajectory computation is employed to eliminate the camera motion in the broadcast video based on global motion estimation. However, GME is an optimization process which will produce the error at each time step. The error accumulation will be magnified while the GME mapping matrixes are used in the long-term transform. Consequently, the computed mosaic trajectory does not reflect the insight of movement of player and ball. 3) The ability of interaction analysis: The cooperation is the crucial tactic information for team sports. How to comprehensively capture the interaction among the players and ball is the key point to facilitate the tactic analysis. The ability of the current analysis with the distance of objects and the shape of trajectory should be enhanced with more various features in terms of object movements and actions.

8.3 Tactic Mode Presentation User Study

To verify the applicability of the tactic pattern presentation with the selected information proposed in section 7, we employed a subjective user study [26] since there is no objective measure available to evaluate the quality of a presentation fashion.

We invited 4 men aging from 23 to 40 in the subjective study. Of the 4 peoples, 1 is soccer coach and 3 are soccer players who have more than five-year team training and four-year professional game playing experience, respectively. They all have rich concepts of soccer game in terms of tactic strategy. To conveniently facilitate the study, we designed a program for goal event browsing and tactic information presentation. Note that the viewer can optionally choose to see the different kinds of trajectories, such as only the trajectories of offensive players, only the trajectories of defensive players, or all the trajectories.

Table 6. Result of subjective study on tactics presentation

	<i>Conciseness</i>	<i>Clarity</i>	<i>Usability</i>
Subject 1	4.4	3.9	4.8
Subject 2	3.9	4.0	4.7
Subject 3	4.5	4.4	4.8
Subject 4	4.2	4.1	4.6

In the study, the subjects were asked to score the presented tactic information according to the following three criterions.

- *Conciseness*: all the information presented is necessary having no tedious content.
- *Clarity*: the information presented is explicit and easily to be understood.
- *Usability*: the information can be used for the further analysis and benefits for the later training and competition.

Five scales are given for the score corresponding to better (5), good (4), common (3), bad (2), and worse (1). For each criterion, the average value of the scores is the final evaluation.

Table 6 lists the result of subjective evaluation. As shown in Table 6, it can be seen that the average evaluations are 4.3, 4.1, and 4.7 for the three criteria respectively. This demonstrates our tactic mode presentation is adapted to the soccer professionals.

9. CONCLUSION

Tactics analysis provides the users more detailed insight of sports game but until now little work has been devoted to this topic. In this paper, we have proposed a novel approach to discover the tactic patterns from the goal events in broadcast soccer video.

As a team sports, the cooperation among the players and the interaction between player and ball characterize the tactic patterns used in the soccer game. Accordingly, two tactic representations, which are aggregate trajectory and play region, are constructed based on the multiple trajectories of players and ball to discover the tactic insight of the game. Compared with existing trajectory modeling approaches, the aggregate trajectory is a novel compact representation for the video content in the tactic context. The tactic clues are extracted from two representations to conduct the pattern analysis of the goal attack. The patterns are classified as route pattern and interaction pattern in which more elaborated tactic scenarios are analyzed. To our knowledge, this is the first solution for soccer game tactic analysis based on broadcast video. We carried out the experiments on the goal events of FIFA World Cup 2006. The results demonstrate that our approach is effective.

Besides the visual tracking exploited in our approach, the acquisition of object trajectory can be achieved by sensor or infrared based methods. Therefore, object trajectory is one kind of the generic features for the team sports. The tactic representation and information extracted from the trajectory is consequently general for the tactics analysis of team sports. In future work, the proposed tactic representation and temporal-spatial interaction analysis will be applied to mining more tactic patterns in soccer games. In addition, the current analysis approach will be extended to other team sports video such as hockey and American football.

10. ACKNOWLEDGMENTS

This work was supported by National Hi-Tech Development Program (863 Program) of China (No. 2006AA01Z117), "Science100 Program" of Chinese Academy of Sciences (No. 99T3002T03), and sponsored by NEC Laboratories China.

11. REFERENCES

- [1] Y. Gong, T.S. Lim, H.C. Chua, H.J. Zhang, and M. Sakauchi, "Automatic parsing of TV soccer program," in *Proc. Int. Conf. Multimedia Computing and System*, 1995, pp. 167-174.
- [2] A. Ekin, A.M. Tekalp, and R. Mehrotra, "Automatic soccer video analysis and summarization," *IEEE Trans. Image Processing*, vol. 12, no. 7, pp. 796-807, 2003.
- [3] L. Xie, P. Xu, S.F. Chang, A. Divakaran, and H. Sun, "Structure analysis of soccer video with domain knowledge and hidden Markov models," *Pattern Recognit. Lett.* vol. 25, no. 7, pp. 767-775, 2004.
- [4] Y. Rui, A. Gupta, and A. Acero, "Automatically extracting highlights for baseball programs," in *Proc. ACM Multimedia*, 2000, pp. 105-115.
- [5] J. Assfalg, M. Bertini, C. Colombo, A. Delbimbo, and W. Nunziati, "Semantic annotation of soccer video: automatic highlights identification," *Comput. Vis. and Image Understanding*, vol. 92, no. 2-3, pp. 285-305, 2003.
- [6] N. Babaguchi, Y. Kawai, T. Ogura, and T. Kitahashi, "Personalized abstraction of broadcasted American football video by highlight selection," *IEEE Trans. Multimedia*, vol. 6, no. 4, pp. 575-586, 2004.
- [7] C. Xu, J. Wang, K. Wan, Y. Li, and L. Duan, "Live sports event detection based on broadcast video and web-casting text," in *Proc. ACM Multimedia*, 2006, pp. 221-230.
- [8] G. Zhu, C. Xu, Q. Huang, W. Gao, and L. Xing, "Player action recognition in broadcast tennis video with applications to semantic analysis of sports game," in *Proc. ACM Multimedia*, 2006, pp. 431-440.
- [9] A. Hanjalic, "Adaptive extraction of highlights from a sport video based on excitement modeling," *IEEE Trans. Multimedia*, vol. 7, no. 6, pp. 1114-1122, 2005.
- [10] L.Y. Duan, M. Xu, T.S. Chua, Q. Tian, and C.S. Xu, "A mid-level representation framework for semantic sports video analysis," in *Proc. ACM Multimedia*, 2003, pp. 33-44.
- [11] M. Xu, L. Duan, C. Xu, and Q. Tian, "A fusion scheme of visual and auditory modalities for event detection in sports video," in *Proc. Int. Conf. Acoustics, Speech, and Signal Processing*, vol. 3, 2003, pp. 189-192.
- [12] P.J. Figueroa, N.J. Leite, and R.M.L. Barros, "Tracking soccer players aiming their kinematical motion analysis," *Comput. Vis. and Image Understanding*, vol. 101, no. 2, pp. 122-135, 2006.
- [13] G. Sudhir, J.C.M. Lee, and A.K. Jain, "Automatic classification of tennis video for high-level content-based retrieval," in *Proc. Int. Workshop on Content-Based Access of Image and Video Databases*, 1998, pp. 81-90.
- [14] J.R. Wang, and N. Parameswaran, "Analyzing tennis tactics from broadcasting tennis video clips," in *Proc. Int. Conf. Multimedia Modeling*, 2005, pp. 102-106.
- [15] P. Wang, R. Cai, and S.Q. Yang, "A tennis video indexing approach through pattern discovery in interactive process," in *Proc. Pacific-Rim Conf. Multimedia*, 2004, pp. 49-56.
- [16] T. Taki, J. Hasegawa, and T. Fukumura, "Development of motion analysis system for quantitative evaluation of teamwork in soccer games," in *Proc. Int. Conf. Image Processing*, vol. 3, 1996, pp. 815-818.
- [17] S. Hirano, and S. Tsumoto, "Finding interesting pass patterns from soccer game records," in *Proc. Eur. Conf. Principles and Practice of Knowledge Discovery in Databases*, vol. 3202, 2004, pp. 209-218.
- [18] C.H. Kang, J.R. Hwang, and K.J. Li, "Trajectory analysis for soccer players," in *Proc. Int. Conf. Data Mining Workshops*, 2006, pp. 377-381.
- [19] <http://soccer.net.espn.go.com>
- [20] Y. Li, C. Xu, K. Wan, X. Yan, and X. Yu, "Reliable video clock time recognition," in *Proc. Int. Conf. Pattern Recognition*, vol. 4, 2006, pp. 128-131.
- [21] G. Zhu, C. Xu, Q. Huang, and W. Gao, "Automatic multi-player detection and tracking in broadcast sports video using support vector machine and particle filter," in *Proc. Int. Conf. Multimedia & Expo*, 2006, pp. 1629-1632.
- [22] D. Liang, Y. Liu, Q. Huang, and W. Gao, "A scheme for ball detection and tracking in broadcast soccer video," in *Proc. Pacific-Rim Conf. Multimedia*, 2005, pp. 864-875.
- [23] F. Dufaux and J. Konrad, "Efficient, robust, and fast global motion estimation for video coding," *IEEE Trans. Image Processing*, vol. 9, no. 3, pp. 497-501, 2000.
- [24] J. Wang, C. Xu, E. Chng, K. Wan and Q. Tian, "Automatic replay generation for soccer video broadcasting," in *Proc. ACM Multimedia*, 2004, pp. 32-39.
- [25] G.A. Korn and T.M. Korn, *Math. Handbook for Scientists and Engineers*, New York: McGraw-Hill, 1968.
- [26] J. Chin, V. Diehl, and K. Norman, "Development of an instrument measuring user satisfaction of the human-computer interface," in *Proc. SIG-CHI on Human Factors in CS*, 1998, pp. 213-218.