Event Tactic Analysis Based on Broadcast Sports Video

Guangyu Zhu, Changsheng Xu, *Senior Member, IEEE*, Qingming Huang, *Member, IEEE*, Yong Rui, *Senior Member, IEEE*, Shuqiang Jiang, *Member, IEEE*, Wen Gao, *Fellow, IEEE*, and Hongxun Yao, *Member, IEEE*

***Abstract*—Most existing approaches on sports video analysis have concentrated on semantic event detection. Sports professionals, however, are more interested in tactic analysis to help improve their performance. In this paper, we propose a novel approach to extract tactic information from the attack events in broadcast soccer video and present the events in a tactic mode to the coaches and sports professionals. We extract the attack events with far-view shots using the analysis and alignment of web-casting text and broadcast video. For a detected event, two tactic representations, aggregate trajectory and play region sequence, are constructed based on multi-object trajectories and field locations in the event shots. Based on the multi-object trajectories tracked in the shot, a weighted graph is constructed via the analysis of temporal–spatial interaction among the players and the ball. Using the Viterbi algorithm, the aggregate trajectory is computed based on the weighted graph. The play region sequence is obtained using the identification of the active field locations in the event based on line detection and competition network. The interactive relationship of aggregate trajectory with the information of play region and the hypothesis testing for trajectory temporal–spatial distribution are employed to discover the tactic patterns in a hierarchical coarse-to-fine framework. Extensive experiments on FIFA World Cup 2006 show that the proposed approach is highly effective.**

***Index Terms*—Event detection, object tracking, trajectory analysis, tactic analysis, sports video analysis.**

Manuscript received September 05, 2007; revised February 04, 2008. Current version published January 08, 2009. This work was supported in part by National Natural Science Foundation of China under Grants 60773136 and 60702035, in part by National Hi-Tech Development Program (863 Program) of China under Grant 2006AA01Z117, and in part by "Science100 Program" of Chinese Academy of Sciences under Grant 99T3002T03. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Alan Hanjalic.

G. Zhu and H. Yao are with the School of Computer Science and Technology, Harbin Institute of Technology, Harbin, China, 150001 (e-mail: gyzhu@jdl.ac.cn; yhx@vilab.hit.edu.cn).

C. Xu is with the National Lab of Pattern Recognition, Institute of Automation, Chinese Academy of Sciences, Beijing 100080, China and also with the China-Singapore Institute of Digital Media, Singapore (e-mail: csxu@nlpr.ia.ac.cn).

Q. Huang is with the Graduate School of Chinese Academy of Sciences, Beijing, China, 100039 (e-mail: qmhuang@jdl.ac.cn).

Y. Rui is with the Microsoft China R&D (CRD) Group, Beijing, China, 100080 (e-mail: yongrui@microsoft.com).

S. Jiang is with the Key Lab of Intelligent Information Processing, Chinese Academy of Sciences, Beijing, China, 100190 (e-mail: sqjiang@jdl.ac.cn).

W. Gao is with the School of Computer Science and Technology, Harbin Institute of Technology, Harbin, China, 150001 and also with the Institute of Digital media, Peking University, Beijing 100871, China (e-mail: wgao@jdl.ac.cn).

Color versions of one or more figures in this paper are available online at http://ieeexplore.ieee.org).

Digital Object Identifier 10.1109/TMM.2008.2008918

I. Introduction

SPORTS content is expected to be a key driver for compelling new infotainment applications and services because of its mass appeal and inherent structures which are amenable for automatic processing. Due to its wide viewership and tremendous commercial value, there has been an explosive growth in the research area of sports video analysis [1]–[21]. From a sports-watcher point of view, only some portions in a sports video are worth viewing. These video segments of interest are the semantic events which have certain high-level concepts, such as goals in soccer games and homeruns in baseball games. The detection and extraction of game events can be achieved by semantic analysis of sports video, to which most of current research efforts have been devoted [1]–[11].

Semantic analysis aims at detecting and extracting information that describes "facts" in a video, e.g., the "goal" events of a soccer match. In contrast, tactic analysis of sports video aims to recognize and discover tactic patterns and match strategies that teams or individual players used in the games. From the coach and sports professional point of view, they are more interested in the tactic strategies in the specific game events. Taking soccer game as an example, there is a great interest from the coaches and players in better understanding the process and patterns of attacks so that he/she is able to improve the team performance during the game and better adapt the training plan. Furthermore, soccer fans, especially the hardcore ones, may also be interested in the results from tactic perspective for enjoying soccer games with the additional information beyond traditional event "facts." Unfortunately, existing semantic approaches on sports video usually only summarize the extracted events and then present to the users directly without any further analysis on the tactics. Today, for sports professionals to obtain the results of tactic analysis, it is common for them to employ people to conduct the analysis manually. This process is labor-intensive, time-consuming and error-prone. Consequently, there exists a compelling case to automate sports tactic analysis. However, to the best of knowledge, immediately related work in the field is very limited.

In this paper, we propose a novel tactic analysis approach on broadcast soccer games. As the most representative genre of the team sports, the soccer game tends to follow the trend of a group cooperation of players from the tactic perspective to compete with the opponent team and to achieve the final goals. Tactic analysis and summarization for the soccer game can potentially offer assistance to the coaches, players and professionals on improving their own skills and studying the opponent strategies.

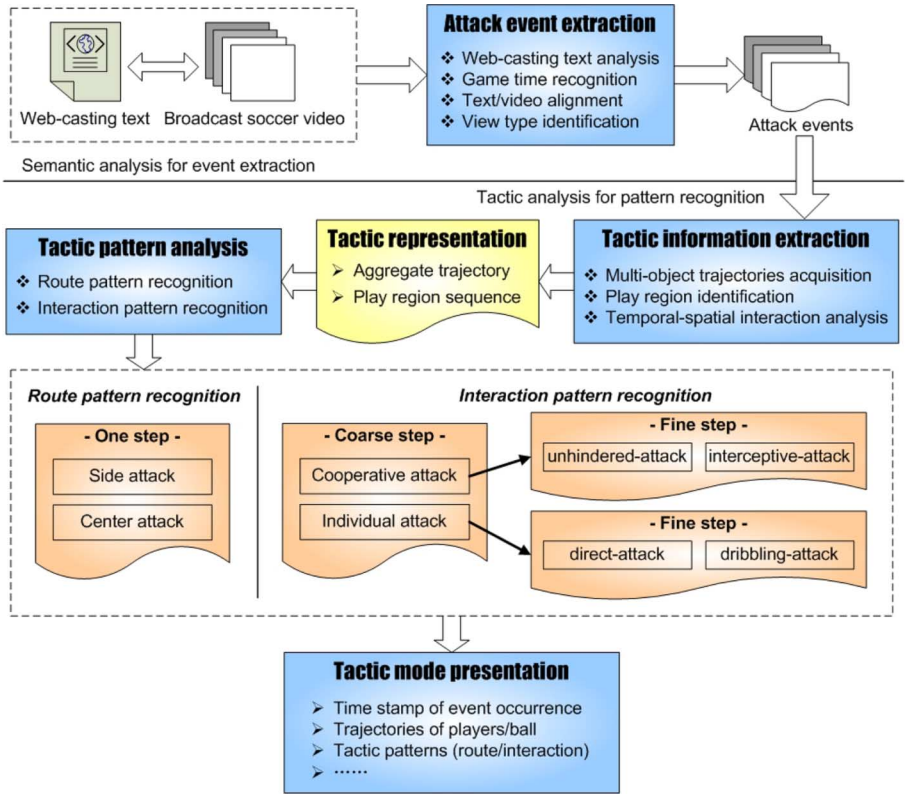

Fig. 1. Flowchart of tactic analysis for attack events in broadcast soccer game.

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 interaction among the players and the ball (e.g., ball passing from one player to others). The trajectories of the players and the ball can reflect such characterization, in which we can locate players and analyze their mutual relationship. In addition, the active field regions of event occurrence provide an indispensable clue for the discovery of tactic process in terms of the spatial distribution. The reason we use broadcast video for tactic analysis is twofold: 1) Because broadcast video has been widely used for sports game broadcasting, it is easy to access and record broadcast video; and 2) Since broadcast video is a post-edited video where the broadcast sequence feed is selected from the composition of multiple camera feeds according to the broadcast director's instruction, multimodal information (e.g., shot transition, visual, audio, and text) can be employed for event detection in broadcast video.

Fig. 1 illustrates the flowchart of our proposed approach which consists of semantic-level and tactic-level soccer game analysis. In the semantic level, the attack events in broadcast soccer video are accurately detected using the multimodal method based on the analysis and alignment of web-casting text and broadcast video. The far-view shots in the events are then identified and aggregated. Far-view shots present the entire process of the attack events and are able to facilitate multi-object detection/tracking and active play region identification. In tactic level, tactic information extraction, representation, and recognition are conducted. Similar to semantic analysis in which the semantic representation is constructed from video content, we need to extract proper clues from the video and construct an effective representation to discover soccer game tactics. Being the salient objects in the games, the movement of the players and the ball is an important piece of information for tactic analysis. In our approach, a multi-object detection and tracking method is employed to obtain the players and the ball trajectories in the attack events. The other fundamental tactic information is extracted from the active field regions in the event based on line detection and competition network. A new temporal–spatial interaction analysis method is proposed to construct the two tactic representations for an event, which are aggregate trajectory and play region sequence, based on multi-object trajectories and field locations. The tactic analysis for the attack events is achieved by recognizing strategic patterns composed of route pattern and interaction pattern. The play region sequence is used to deduce the route pattern, e.g., side-attack and center-attack. The interaction pattern recognition is formatted into a hierarchical coarse-to-fine scheme based on the aggregate trajectory. At the coarse level, the attacks are classified into cooperative pattern where the attack is implemented by multiple players via ball passing, and individual pattern where the attack is implemented by only one player. The coarse patterns are then classified into four elaborated scenarios: unhindered-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 I. Finally, the classified patterns of the attack event with other related information are presented to the professional users in a tactic mode.

TABLE I
DESCRIPTION OF THE INTERACTION PATTERNS

Coarse pattern	Fine pattern	Description
cooperative attack	unhindered-attack	No ball intercepted by defenders in the attack process
	interceptive-attack	Ball intercepted by one or more defenders in the attack process
individual attack	direct-attack	No ball dribbling in the attack event (e.g. penalty kick, free kick)
	dribbling-attack	Ball dribbling by attacker in the attack process

In our previous work [12], preliminary results of tactic analysis of the goal events in broadcast soccer video were reported. Compared with our previous work, a number of significant improvements have been made in this paper. Firstly, the multi-object detection and tracking method used in [12] is enhanced by using probabilistic support vector classification [33] for the construction of proposal distribution in the particle filter to eliminate the mutual occlusion among the players. Secondly, the graph modeling is introduced for trajectory temporal–spatial analysis to increase the robustness of original algorithm, where the temporal relationship between the successive trajectory segments is considered. The Viterbi algorithm is employed to compute the aggregate trajectory based on the constructed weighted graph. Thirdly, with the assistance of more comprehensive keywords definition for web-casting text analysis, the original tactic pattern recognition approach is extended to more general attack events in broadcast soccer video which include not only the scoring (goal) events but also the non-scoring events.

The rest of the paper is organized as follows. Section II introduces the existing work related to semantic and tactic analysis of sports (soccer) video. Section III presents the approach of attack event extraction using multimodal method with the combination of web-casting text analysis and game time recognition. In Section IV, we describe the method to extract the tactic information and construct tactic representations from the attack event in broadcast soccer video. Section V describes the details of tactic pattern analysis for the soccer game based on the constructed tactic representations. The scenario of tactic mode presentation is presented in Section VI. In Section VII, experimental results are reported and discussed. Finally, we conclude the paper in Section VIII.

II. RELATED WORK

In this section, we review the state-of-art in sports video/game analysis in terms of semantic event extraction and tactic strategy analysis corresponding to the two levels in the framework of our proposed approach.

A. Semantic Event Extraction for Sports Videos

Extensive research efforts have been devoted to sports video event detection and semantic extraction. The existing approaches can be classified into two classes: video content based and external sources based.

1) Event Extraction Based on Video Content Only: To identify certain events from lengthy sports video documents, most of existing approaches used audio/visual/textual features directly extracted from video content and built various models to detect event and recognize semantics [1]–[9]. These approaches can be further classified into single-modality based and multimodality based. Single-modality based approaches only use single stream in sports video for event detection in terms of visual [1]–[8], audio [3], and text [4]. The single-modality based approaches have low computational load, but the accuracy of event detection is low because the content of sports video is intrinsically multimodal while only using single modality is not able to fully characterize the events in sports video. In order to improve the robustness of event detection, multimodality based approaches were employed for semantic extraction in sports video. For example, audio/visual features were utilized for highlight extraction [6], [7], and audio/visual/textual features were utilized for event detection [5], [9].

Due to the semantic gap between low-level features and high-level semantics as well as dynamic structures of different sports games, it is difficult to use the above video content based approaches to address following challenges: 1) achieving ideally high event detection accuracy, 2) recognizing the event detail, e.g., who scores the goal and how the goal is scored, and 3) providing a generic event detection framework for different sports games.

2) Event Extraction Based on External Sources: The limitation of event detection approaches based on video content only motivates us to use external sources related to sports video to assist semantic analysis. There are two external sources that can be used for sports event extraction: closed caption and web which are both text sources. Incorporation of text into sports video analysis is able to help bridge the semantic gap between low-level features and high-level events and thus facilitates the sports video semantic analysis.

Closed caption is a manually tagged transcript from speech to text and encoded into video signals. It has been used for sports video semantic analysis [17]. However, closed caption contains a lot of information irrelevant to the games and lacks of a well-defined structure. On the other hand, currently closed caption is only available for certain sports videos and in certain countries. In addition to closed caption, the information in the web was also utilized to assist sports video analysis. Xu and Chua [11] proposed an approach to utilize match report and game log obtained from web to assist event detection in soccer video. In our previous work [10], Xu *et al.* proposed a live soccer event detection system using web-casting text and broadcast video and conducted a live trial on FIFA World Cup 2006. The results are encouraging and comparable to the manually detected events.

B. Tactic Strategy Analysis for Sports Games

Most existing approaches for tactic analysis of sports are focused on the tennis game because its game field and the number of players participating in the match are both relatively small. The key techniques used in these approaches are tracking the trajectories of the players and the ball with the assistance of domain knowledge [13]–[16] and recognizing the actions of the players [9]. 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. Pingali *et*

al. [14] presented a real time tracking approach for the players and the ball in the tennis game to obtain the temporal–spatial trajectories which can provide a wealth of information about the game. This work was based on specific-set camera system. In [15], the tennis games were attempted to be classified into 58 winning patterns for training purpose based on tracking the ball movement from broadcast video. Wang *et al.* [16] presented a novel approach for tennis video indexing by mining the salient tactic patterns in the match process. Unlike trajectory-based algorithm, a novel action-driven tactic analysis approach was proposed in [9] for the tennis game, which is able to discover the insight of the stroke performance of the players. For other sports, a tactic analysis system for American football games was reported in [17] based on the extracted semantic events, e.g., kickoffs and touchdowns. In [18], Han *et al.* proposed a digest system for baseball game, which is able to infer the match strategies. However, these methods are limited for deep tactic analysis because they are based on event-driven indexing of video contents and usually inflexible and intractable in tactic summarization. Furthermore, the nature of the interaction among the players (or players and ball), which is critical for acquiring the match strategies, has not been adequately addressed.

Little work [19]–[21] attempted to conduct tactic analysis for soccer games. In [19], the players' positions were estimated from the soccer game image sequence which was captured by a multiple and fixed cameras system, and then 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 tactic strategy of a soccer team was evaluated. In [20], a study was conducted on the discovery of meaningful pass patterns and sequences from time-series recorded data of soccer games. An evaluation model was proposed in [21] to quantitatively evaluate the performance of soccer players, using the relationship between the trajectories of twenty-two players and a ball as input and having the performance evaluation of several players in a quantitative way as output. However, the existing work was based on the multicamera-recorded [19], human-labeled [20] and computer-simulated [21] trajectory data which has strong limitation and less challenge for object tracking and pattern discovery using broadcast video.

C. Our Contribution

Existing approaches for soccer video analysis mostly focused on event-driven indexing of video content, which cannot provide detailed tactic information used in the game. Little work on tactic analysis used non-broadcast video [20], [21] or had to conduct the camera calibration [19] on broadcast video, which are very difficult or impossible to be adapted to wide applications. In this paper, we propose a novel tactic analysis approach for the attack events in broadcast soccer video.

The main contributions of our work are summarized as follows.

1) Two novel tactic representations, aggregate trajectory and play region sequence, are proposed, which are constructed based on multiple trajectories and active field locations using the analysis of temporal and spatial interaction of the players and the ball.

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

Fig. 2. Example of web-casting text.

2) A hierarchical coarse-to-fine framework is proposed to identify the tactic strategies of the attack events including route pattern and interaction pattern. The inference using play region sequence, the parsing by temporal–spatial object interaction and the hypothesis testing for distribution of aggregate trajectory are employed in the analysis, respectively.
3) An improving tracking strategy based on our previous work [26] is applied to players and ball detection and tracking with the integration of particle filter and support vector machine.
4) Compared with existing work [19]–[21], our approach is implemented on broadcast video, which is widely used in the real applications.

III. Math Attack Event Extraction From Broadcast Soccer Video

We adopt an effective detection method to extract the attack events from broadcast soccer video by combining the analysis and alignment of web-casting text and video content, which has the advantage of low computational load and high detection accuracy. As shown in the semantic analysis module illustrated in Fig. 1, a text event is first detected from the web-casting text and time stamp indicating when the event occurs in the game is obtained. Then, the game time is recognized in the video and the moment when the event occurs in the video is detected by linking the time stamp from text event to the related game time in the video. Based on the event moment, the whole event sequence is detected from video using shot type identification and finite state machine modeling. All the attack events are extracted using the same way and the far-view shots in the detected events are aggregated for tactic analysis.

A. Web-Casting Text Analysis

The web-casting text [22] serves as text broadcasting for sports games. As shown in Fig. 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.

In the web-casting text, each type of event features one or several unique nouns, such as "Goal" and "Headed in" for score event. This is because the web-casting text is tagged by sports professionals and has fixed structures. By detecting these nouns, the sentence relevant to certain event can be identified from the web-casting text. We define these nouns as "event keywords"

TABLE II
KEYWORD DEFINITION FOR ATTACK EVENTS

Event	Keyword
Goal	goal or scored or g-o-a-l or equalize – kick
Shot	shot or header or headed in
Corner	corner kick
Free kick	(take or save or concede or deliver or fire or curl) w/6 (free kick or free-kick or freekick)

Fig. 3. Overlaid video clock.

and use software dtSearch [23] to detect them. dtSearch provides stemming, phonic, fuzzy and Boolean searching options to achieve better performance than simple word matching [24].

Considering the context of event occurrence in the soccer video, we define the attack events of soccer game as the combination of four subevent categories including "goal," "shot," "corner," and "free kick." Therefore, the keywords detection for the attack events is equivalent to the keywords detection of the four subevents. Table II lists the keywords detected using dtSearch's grammar. Note that these four events are not always exclusive, but sometimes rather overlap with each other. For instance, a "goal" event is also a "free kick." However, as most web-casting text sources contain a precise description of each type of event, we can accurately distinguish them. The detected events by keywords searching in the web-casting text are referred to as text events.

B. Game Time Recognition

In broadcast soccer videos, a video clock is usually 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. As shown in Fig. 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 [25]. Compared with the traditional methods, this approach is able to achieve the real time performance and the result is more reliable.

Our algorithm first locates the static overlaid region by static region detection. The region of interest for character is then detected using connected component analysis. Temporal neighboring pattern similarity (TNPS) is used as the most critical feature to locate the digits. Since the clock digits are changing periodically, for each region of interest in a character, we observe its TNPS sequence which is defined as follows:

$$S(n) = \sum_{(x,y)\in I} B_{n-1}(x,y) \otimes B_n(x,y) \tag{1}$$

where $B(x,y)$ is the binarized image pixel value in position (x,y), n is the frame number in the sequence, I is the character region, and $\otimes$ is XOR operation. If the change of TNPS pattern follows the time-changing regulation, the character is considered as a clock digit.

After the clock digits are located, we observe the TEN-SECOND digit pattern change using the TNPS. At the time when the pattern change happens, we extract the pattern templates of digit "0" to "9" from the SECOND digit region of interest automatically. 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. The matching score of numeric character i is calculated as follows.

$$\text{score}(i) = \max_j \left\{ \sum_{(x,y)\in I} T_{i,j}(x,y) \otimes D(x,y) \right\}, \quad i = 0,1,\ldots,9,10 \tag{2}$$

where $T_{i,j}(x,y)$ is the binarized image pixel value in position (x,y) for the j th template of numeric character i, $D(x,y)$ is the binarized image pixel value in position (x,y) for the digit character to be recognized, and I is the region of interest of the digit character. When $i = 10, T_{10,j}(x,y)$ is the template for a flat region without any character. The clock digits on every frame are recognized with a best match. The details of game time recognition can be found in [25].

C. Text/Video Alignment

After recognizing 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.

With the empirical observation, the duration of the most events in broadcast soccer video lasts between 20 to 60 s. Based on the detected event moment f_r in the video, we define a temporal range $[f_r - 60, f_r + 120]$ 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 and 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

$$S = \{s_i | s_i = \langle sbt_i, st_i, ebt_i \rangle, i = 1, \ldots, N\} \quad (3)$$

where $sbt_i \in \{\text{cut, dissolve}\}$ is the start boundary type, $st_i \in \{\text{far} - \text{view}, \text{non}-\text{far}-\text{view}\}$ is the shot type, $ebt_i = \{\text{cut, dissolve}\}$ is the end boundary type, and N is the total number of shots. The view type is identified using the dominant color analysis, where 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 details of event boundary detection can be found in [10].

IV. Tactic Information Extraction and Representation

In this section, we present a method of video representation in the tactic context. We extract the multi-object (players and ball) trajectories and the active field regions in the attack events as the tactic information to construct the tactic representation of video content.

A. Multi-Object Trajectories Acquisition

1) Player Detection and Tracking: In our previous work [26], we proposed a multi-object detection and tracking approach. In this paper, we improved the previous work and applied it to player detection and tracking in the far-view shots of the attack events in soccer videos. Compared with previous work [26], there are two major improvements: 1) an adaptive Gaussian mixture model (GMM) [27] is used in playfield detection to increase the accuracy of object detection, and 2) probabilistic support vector classification (PSVC) [33] is integrated into the framework of support vector regression (SVR) particle filter for the construction of proposal distribution to eliminate the mutual occlusion among the players.

The flowchart of player detection and tracking algorithm is shown in Fig. 4. Playfield detection is first conducted using an adaptive Gaussian mixture color model with the evidence that the playfield pixels are the dominant components in most of the frames of a far-view shot. Compared with traditional GMM, the adaptive GMM can update mixture model parameters by incremental expectation maximization (IEM) algorithm which enables the model to adapt to the playfield variation with time. Moreover, online training is performed which is able to save the buffer space for the samples of GMM training. The effectiveness of adaptive GMM for playfield detection has been demonstrated in [27]. The regions inside the extracted field are considered as the player candidates. Then, recognition module based on traditional deterministic support vector classification (SVC) [32] is employed to eliminate the non-player candidates using the color histograms of the regions. For each of the 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. PSVC

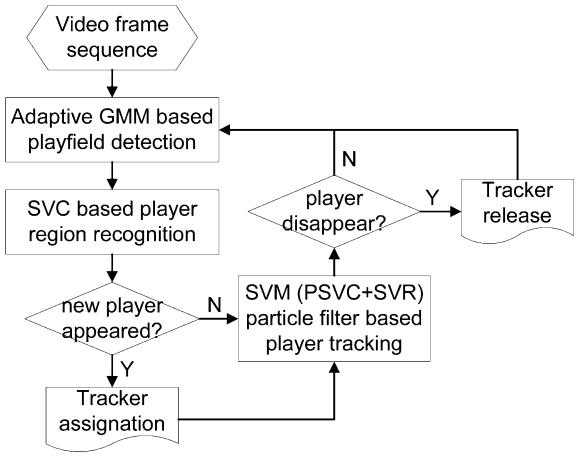

Fig. 4. Diagram of player detection and tracking.

based improvement is exploited to solve the problem of player occlusion. After each tracking interaction, the player disappearance module using SVC recognition model evaluates whether the currently tracked player leaves the scene. If it does, the corresponding tracker is released. During the tracking process, the color histogram of the target region is employed to identify the team affiliation of the tracked player.

The integration of support vector machine and particle filter enhances the power of particle filter based tracker from two perspectives. On the one hand, for the tracking algorithm based on particle filter, the key points are the likelihood computation of the different hypothesis observations 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 [26]. On the other hand, we use PSVC to improve the tracking performance in the case of player occlusion.

It is well known that the choice of proposal distribution is a crucial issue for particle filter tracker design. It has been widely accepted that the proposal distribution which incorporates the recent observation outperforms naïve transition prior proposal $p(\mathbf{x}_t|\mathbf{x}_{t-1})$ considerably [28], where $\mathbf{x}_{t-1}$ and $\mathbf{x}_t$ are the state vectors at time $t-1$ and t respectively. To eliminate the occlusion problem which is motivated by boosted particle filter [29], [30], we construct the proposal distribution using a mixture model that incorporates information from the dynamic models of each player and the detection hypotheses generated by PSVC. PSVC can extract probability distribution from deterministic SVC outputs. The insight of this improvement is to predict the samples to the critical areas of the object distribution. Note that the PSVC detection scenario here is not based on the color information used in player detection but the cascade Haar features proposed in [31]. The representation for the proposal distribution q^* is given by the following mixture.

$$q^*(\mathbf{x}_t|\mathbf{x}_{0:t-1}, \mathbf{y}_{1:t}) = \alpha \cdot p(\mathbf{x}_t|\mathbf{x}_{t-1}) + (1-\alpha) \cdot q_{\text{PSVC}}(\mathbf{x}_t|\mathbf{x}_{t-1}, \mathbf{y}_t) \quad (4)$$

where $q_{\text{PSVC}}(\mathbf{x}_t|\mathbf{x}_{t-1}, \mathbf{y}_t)$ is the probability distribution given by the PSVC detection, $\mathbf{x}_{0:t-1} = \{\mathbf{x}_0, \ldots, \mathbf{x}_{t-1}\}$ is denoted as the state vectors up to time t and $\mathbf{y}_{1:t} = \{\mathbf{y}_1, \ldots, \mathbf{y}_t\}$ is similar

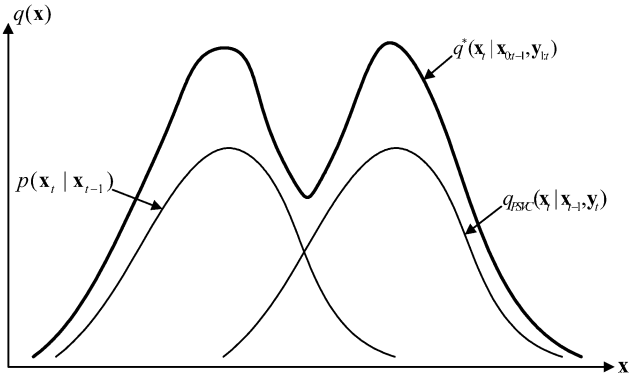

Fig. 5. Mixture of naive transition prior and detection probability for proposal distribution.

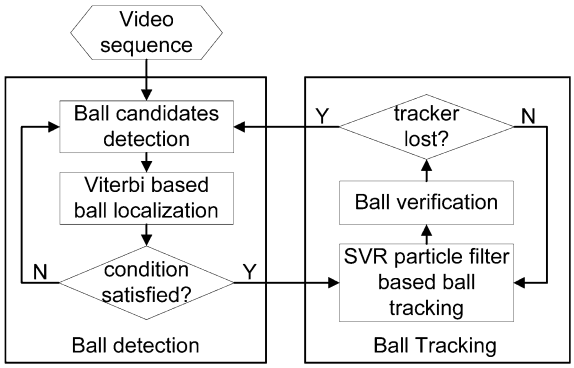

Fig. 6. Diagram of ball detection and tracking.

to the observations. Fig. 5 shows the proposal distribution q^* defined by (4) which is similar to the illustration in [29]. The Haar features are the part-based representation, which are locally extracted from the image of the player region. PSVC detection using Haar features is therefore able to locate the player under the conditions that the player is occluded partly and crossed over shortly. Consequently, the enhanced proposal distribution with the incorporation of PSVC detection is able to propagate the samples to the areas with high probability of player appearance even under the conditions of cross over and occlusion.

In (4), the parameter α can be set dynamically without affecting the convergence of the particle filter. When $\alpha = 1$, our algorithm reduces to the original SVR particle filter. By decreasing α, we place more importance on PSVC detection. We can adapt the value of α depending on tracking situations including cross over, collision, and occlusion. By the tradeoff analysis for three conditions in the experiments, we set α to be 0.7. More detailed information abut the work of mixture proposal distribution for particle filter can be found in [29], [30].

Other minor improvements are exploited to enhance the power of the tracking algorithm: 1) a second order autoregression model is adopted as the dynamics model to replace the first order model in [26], 2) a global nearest neighbor data association technique [30] is used to correctly associate SVC detection with the existing trackers, and 3) PSVC detection is used to maintain the robustness of the observation model to stabilize the trajectories of the targets for reliable movements.

2) Ball Detection and Tracking: The challenges of ball detection and tracking in broadcast video are due to the following issues: 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, 4) many other objects are similar to the ball, such as some regions of the players and some white line segments in the playfield, of which the positions in the frames constitute the multiple hypotheses of ball locations.

To solve the above-mentioned challenges, a new method is proposed by enhancing our previous work [34]. Fig. 6 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. The number of adjacent frames utilized for the graph construction can be set empirically, e.g., five frames in our experiments. The 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 and template matching is started. SVR particle filter and the template are initialized using detection results. In each frame, ball location is verified to update the template to check if the ball is lost. If the ball is lost, the detection runs again.

The trajectory interpolation [35] is employed as the post-processing to solve the situation where the ball is occluded or it is out of the camera view temporarily. Let T_1 and T_2 be two obtained ball trajectories. Let e_1 and s_2 be the indexes of the last frame in T_1 and the first frame in T_2, respectively. We compute the ball locations between e_1 and s_2 using Kalman estimator from two directions if $s_2 - e_1 < Z_{GI}$ where Z_{GI} is set to be the half of the frame rate (frame-per-second) of broadcast video, e.g., $Z_{GI} = 15$.

B. Aggregate Trajectory Computation

Aggregate trajectory is a compact tactic representation for broadcast soccer video, which is built upon multi-object (players and ball) trajectories using mosaic technique and temporal–spatial analysis. In previous approaches [20], [21], the tactic representation was extracted from the trajectories labeled by human or computer simulation. These labeled/simulated data correspond to the locations of the players in the real soccer field and are easy for interaction analysis among the players. However, broadcast video normally contains frequent camera motion and severe object occlusion in the video sequence which leads to more difficulties for interaction analysis from extracted trajectories. Therefore, we need to construct an effective tactic representation for broadcast video to facilitate tactic analysis. Aggregate trajectory is such a representation which can be obtained using video mosaic and temporal–spatial interaction analysis.

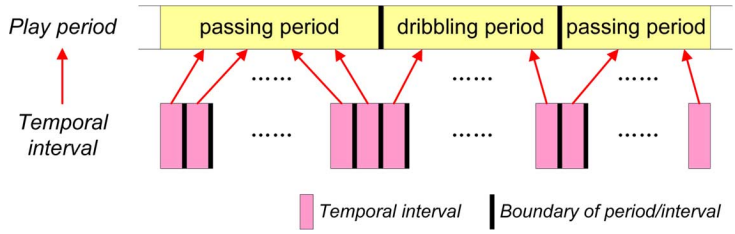

Fig. 7. Period/interval structure for temporal–spatial interaction analysis.

1) Mosaic Trajectory Computation: Mosaic trajectories are generated by transforming the actual trajectories extracted from the broadcast video sequence into a common coordinate space to eliminate camera motions in the video sequence using mosaic technique. The common mosaic technique is based on global motion estimation.

Global motion estimation (GME) [36] is used to establish the mapping between the spatial coordinates in two successive video frames. Using the 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 of an object in frame t and $\mathbf{x}_{t-1}$ denotes the coordinate of the same object in frame $t-1$, the mapping between $\mathbf{x}_t$ and $\mathbf{x}_{t-1}$ is represented as

$$\mathbf{x}_{t-1} = \mathbf{H}_{t,t-1} \cdot \mathbf{x}_t \tag{5}$$

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

Given a video sequence of an event $V = \{f_1, \ldots, f_n\}$ where f_i is the i th frame, $1 \le i \le n$, and n is the total number of the frames, one trajectory in the event is $T = \{p_1, \ldots, p_n\}$ where p_i is the position of the object tracked in frame $f_i, 1 \le i \le n$. Considering (5) and the temporal relation of the frames, we can therefore warp each p_i into the uniform coordinate of frame f_1 as follows:

$$\tilde{\mathbf{p}}_i = \prod_{t=2}^{i} \mathbf{H}_{t,t-1} \cdot \mathbf{p}_i \tag{6}$$

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 event are warped into the coordinate space of frame f_1, which is essentially a common coordinate space.

Once the mosaic trajectories of all the players and the ball are computed, the motions in broadcast video caused by camera behavior can be treated as being removed. The mosaic trajectories correspond to the loci of the players and the ball captured by a fixed camera, which can be used as the input of the following temporal–spatial interaction analysis among multiple objects.

2) Temporal and Spatial Interaction Analysis: The insight of aggregate trajectory is to capture the interaction relationship among the players and the ball in a compact representation. Ball trajectory is the key component in tactic analysis because all the tactic strategies in the soccer game will be finally conducted on the ball. In soccer game, the most two important interactions among the players and the ball are ball-passing and ball-dribbling. The 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.

As shown in Fig. 7, an attack event can be structured into one or more play periods. We can classify the periods into two categories, passing period and dribbling period, which correspond to the two interactions in the soccer game respectively. Each period can be further partitioned into several temporal intervals with smooth and continuous object trajectories. In our previous work [12], we employed a deterministic method based on the similarity metrics in terms of distance and shape between the segments of each player and ball trajectory in the temporal interval to generate the aggregate trajectory. For passing period, the ball trajectory is selected as the component of the aggregate trajectory and the satisfying results were achieved using the previous approach due to the deterministic attribute of the ball trajectory existing in the entire event process. Compared with passing period, the analysis of dribbling period is more challenging because multitrajectory segments corresponding to the multiple players are involved and one segment needs to be selected in each interval. In practice, the deterministic method is sensitive to the noise in the data of multiple trajectories due to the tracking errors thus may result in the false segment selection.

To enhance the robustness of the analysis for dribbling period, we formulate the problem into a probabilistic framework based on multiple hypotheses estimation and temporal structure analysis of trajectory segments. In dribbling period, the multitrajectory segments with the correspondent similarity evaluation values constitute the multiple hypotheses for the components of the aggregate trajectory, where each hypothesis can be viewed as being selected with a probability (similarity value) of occurrence. In our previous method [12], only the spatial relationship (distance and shape) is considered. Considering that trajectory is the time-series data, the temporal relationship among the trajectory segments in the intervals within a dribbling period is investigated and modeled by a weighted graph in this paper. Fig. 8 shows the flowchart of aggregate trajectory generation based on the temporal–spatial analysis.

Let us denote the set of mosaic trajectories for a given attack event is $\mathrm{MT} = \{l_b(t), l_{p_1}(t), \ldots, l_{p_n}(t)\}$ where $l_b(t)$ is the trajectory of the ball and $l_{p_i}(t)$ is the trajectory 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

Input: mosaic trajectories $MT = \{l_b(t), l_{p_1}(t), \ldots, l_{p_n}(t)\}$

Output: aggregate trajectory AT

1. Smooth each trajectory in MT using Gaussian filter.
2. Partition all the trajectories into T subsets $SMT = \{S_1, \ldots, S_T\}$ according to uniformly temporal interval τ, where $S_k = \{l_b^k(t), l_{p_1}^k(t), \ldots, l_{p_{n_k}}^k(t)\}$, and $l_{\#}^k(t)$ is the kth trajectory segment of $l_{\#}(t)$ in MT, $1 \le k \le T$.
3. Use temporal/spatial similarity metrics and graph modeling to select the segment in each temporal interval of passing period and dribbling period to generate the AT.
   - 3.1. Initialize $k = 1$.
   - 3.2. For S_k, evaluate the similarity metric $F_{b,p_j}^k = D_{b,p_j}^k \cdot H_{b,p_j}^k$, where D_{b,p_j}^k and H_{b,p_j}^k, $1 \le j \le n_k$, are the similarity measures computed on $l_b^k(t)$ and $l_{p_j}^k(t)$ in terms of distance and shape using Euclidean measurement. Select $\widehat{F}_{b,p_{j*}}^k$ which is the maximum of $\{F_{b,p_j}^k\}_{j=1}^{n_k}$.
   - 3.3. Passing period process using deterministic method: If $\widehat{F}_{b,p_{j*}}^k < thres$, the ball is identified to be passing and $l_b^k(t)$ is selected as the component of AT in the interval k.
   - 3.4. Dribbling period process using probabilistic graph modeling:
     - 3.4.1. If $\widehat{F}_{b,p_{j*}}^k \ge thres$, the interval k is identified as ball-dribbling. Construct the weighted graph node set NS_k for interval k, where $NS_k = \left\{ \left(l_{p_j}^k(t), v_{p_j}^k \right) \mid l_{p_j}^k(t) \in S_k \wedge F_{b,p_j}^k \ge thres, v_{p_j}^k = F_{b,p_j}^k, 1 \le j \le m_k \right\}$.
     - 3.4.2. Establish the weighted graph edge set $ES_{k-1,k}$ for NS_{k-1} and NS_k in terms of temporal similarity metrics, where $ES_{k-1,k} = \left\{ e_{p_i,p_j}^{k-1,k} \left(l_{p_i}^{k-1}(t), l_{p_j}^k(t) \right) \mid l_{p_i}^{k-1}(t) \in NS_{k-1}, l_{p_j}^k(t) \in NS_k \right\}$. The weighted edge $e_{p_i,p_j}^{k-1,k}$ is calculated as $e_{p_i,p_j}^{k-1,k} = V_{p_i,p_j}^{k-1,k} \cdot C_{p_i,p_j}^{k-1,k}$, where $V_{p_i,p_j}^{k-1,k}$ and $C_{p_i,p_j}^{k-1,k}$ are the similarity measures computed on $l_{p_i}^{k-1}(t)$ and $l_{p_j}^k(t)$ in terms of velocity and linking distance using Euclidean measurement.
     - 3.4.3. Construct the weighted graph and employ the Viterbi algorithm to select the optimal player trajectory segment $l_{p_{j*}}^k(t)$ as the component of AT in the interval k.
   - 3.5. If $k < T$, $k = k + 1$ and goto 3.2. Else goto 4.
4. Concatenate all the selected segments by the temporal index k to generate the AT.

Fig. 8. Flowchart of aggregate trajectory generation based on temporal–spatial interaction analysis.

are uniformly partitioned into the segments using an equally temporal interval τ (e.g., $\tau = 1s$). To generate the aggregate trajectory, two alternative steps are conducted. A deterministic method based on the similarity evaluation in terms of spatial metrics (distance and shape of the trajectory segments) is employed to identify the passing period and dribbling period. For the interval of passing period, the segment of the ball trajectory is selected as the component of the aggregate trajectory. Otherwise, a probabilistic graph modeling method is employed to select the optimal player trajectory segment to generate the aggregate trajectory in the dribbling period.

a) Passing Period Process Using Deterministic Method: As shown in Fig. 9, the distance between the dribbling-player and the ball is nearer in a temporal interval of the ball-dribbling period. Moreover, the shape of trajectories of the player and the ball is similar because the player and the ball are followed the similar route on the field in the interval. However, such observation is not guaranteed in the ball-passing period (including the case that the player passes the ball or the player uncontrolles the ball).

For temporal interval k, we define two similarity measures D_{b,p_j}^k and H_{b,p_j}^k for the trajectory segments of the ball and the j th player in terms of distance and shape respectively, $1 \le j \le n_k$ where n_k is the number of the objects in interval k. Given $l_b^k(t) = \{\mathbf{u}_1, \ldots, \mathbf{u}_r\}$ and $l_{p_j}^k(t) = \{\mathbf{v}_1, \ldots, \mathbf{v}_r\}$ where $\mathbf{u}$ and $\mathbf{v}$ represent the object positions in the trajectory, r is the number of included positions, D_{b,p_j}^k is defined as

$$D_{b,p_j}^k = \exp\left\{ -\frac{1}{c_1 \cdot r} \sum_{q=1}^{r} \|\mathbf{u}_q - \mathbf{v}_q\| \right\} \tag{7}$$

where $\|\mathbf{u} - \mathbf{v}\|$ is the Euclidean distance for $\mathbf{u}$ and $\mathbf{v}$ in $\mathbf{R}^2$ space, and c_1 is a normalization constant ensuring $(1)/(c_1 \cdot r) \sum_{q=1}^{r} \|\mathbf{u}_q - \mathbf{v}_q\| \in [0, 1]$. H_{b,p_j}^k is computed on the spatial curvature of 2-D curve given by

$$c(k) = \frac{x'(k) \cdot y''(k) - y'(k) \cdot x''(k)}{[x'(k)^2 + y'(k)^2]^{3/2}} \tag{8}$$

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 (8), we can calculate the curvature sequences $c_b^k = \{a_1, \ldots, a_r\}$ and $c_{p_j}^k = \{b_1, \ldots, b_r\}$ for $l_b^k(t)$ and $l_{p_j}^k(t)$ where r is the number of positions. The H_{b,p_j}^k is then computed as

$$H_{b,p_j}^k = \exp\left\{ -\frac{1}{c_2 \cdot r} \sum_{q=1}^{r} |a_q - b_q| \right\} \tag{9}$$

where $|x|$ denotes the absolute value of x, c_2 is a constant to normalize the exponential in the range of 0 and 1. Using D_{b,p_j}^k and H_{b,p_j}^k, we define the similarity metric F_{b,p_j}^k as

$$F_{b,p_j}^k = D_{b,p_j}^k \cdot H_{b,p_j}^k \tag{10}$$

and obtain the $\widehat{F}_{b,p_{j*}}^k$ which is the maximum of all the F_{b,p_j}^k, $1 \le j \le n_k$. If $\widehat{F}_{b,p_{j*}}^k < \text{thres}$ where $\text{thres} = 0.2$ is a predefined threshold, the ball is identified to be passing and the trajectory segment $l_b^k(t)$ is selected as the component of the aggregate trajectory in interval k.

b) Dribbling Period Process Using Probabilistic Graph Modeling: According to (10), if $\widehat{F}_{b,p_{j*}}^k \ge \text{thres}$, the interval k is identified belonging to a dribbling period. A weighted graph is constructed and the Viterbi algorithm is used to select the optimal player trajectory segment to generate the component of the aggregate trajectory. Fig. 10 shows the illustration of the graph modeling and optimal segment selection.

In Fig. 10, each graph node represents a player trajectory segment $l_{p_j}^k(t)$ with spatial evaluation metric $F_{b,p_j}^k \ge thres$

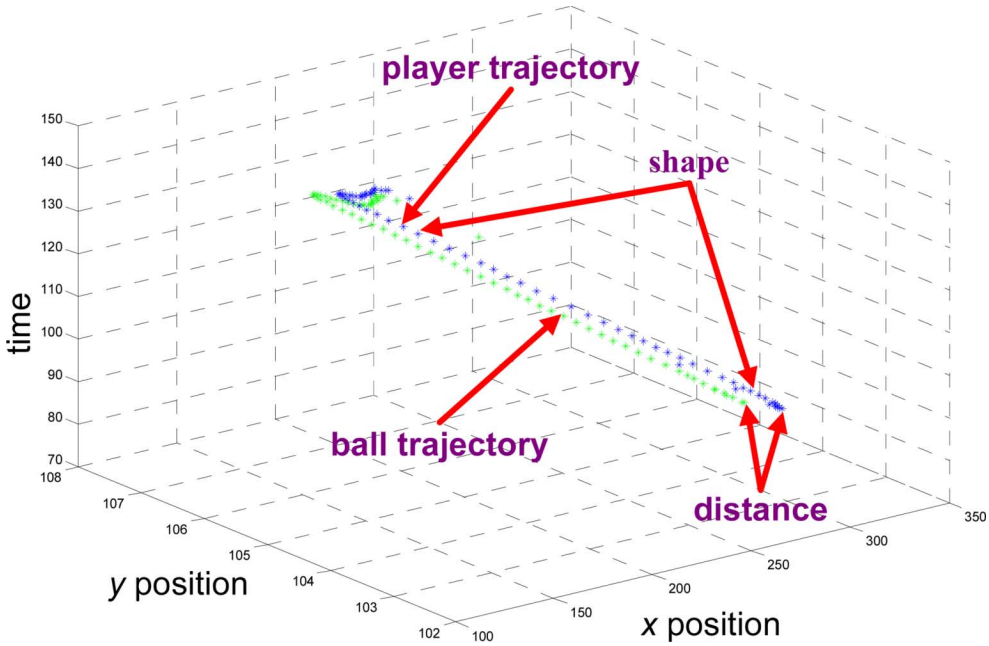

Fig. 9. Illustration of spatial relationship in terms of distance and shape for the trajectory segments of player and ball in dribbling period.

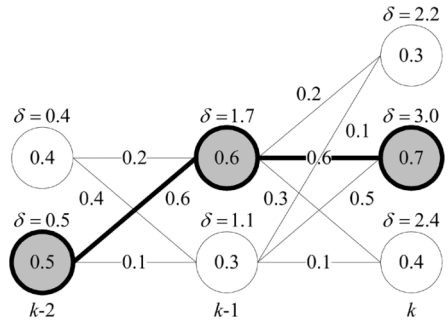

Fig. 10. Illustration of the segment selection based on weighted graph modeling. Three time intervals are presented in the figure to clearly illustrate the detail of the method; the optimal segments selected are represented by the gray circles.

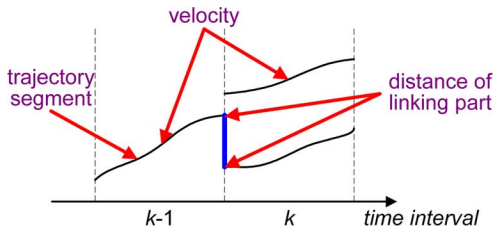

Fig. 11. Illustration of temporal relationship in terms of velocity and linking distance for player trajectory segments in dribbling period.

in interval k. The value F^k_{b,p_j} is set as the node weight $v^k_{p_j}$ of $l^k_{p_j}(t)$, which can be treated as the probability of $l^k_{p_j}(t)$ to be selected. Meanwhile each edge is assigned a weight considering the temporal structure of the dribbling period. With our empirical observation, the two successive dribbling segments have similar velocities because of the motion inertia of the players in the process of ball-dribbling. In addition, the linking distance, which is the length of the linking part between two segments in the successive intervals as shown in Fig. 11, is employed to guarantee the smoothness and continuity of the aggregate trajectory. Thus, two similarity measures $V^{k-1,k}_{p_i,p_j}$ and $C^{k-1,k}_{p_i,p_j}$ are defined in terms of velocity and linking distance among the segments in two successive intervals, respectively. Given $l^{k-1}_{p_i}(t) = \{\mathbf{p}_1, \ldots, \mathbf{p}_r\}$ and $l^k_{p_j}(t) = \{\mathbf{q}_1, \ldots, \mathbf{q}_r\}$ where $\mathbf{p}$ and $\mathbf{q}$ are the player positions in the trajectories, $V^{k-1,k}_{p_i,p_j}$ is defined as

$$V^{k-1,k}_{p_i,p_j} = \exp\left\{-\frac{1}{c_3 \cdot r}\sum_{l=1}^{r}\|\mathbf{p}'_l - \mathbf{q}'_l\|\right\} \tag{11}$$

where $\mathbf{x}'$ represents the velocity of position $\mathbf{x}$ calculated as the first-order derivative of $\mathbf{x}$ and c_3 is a constant to normalize $(1)/(c_3 \cdot r)\sum_{l=1}^{r}\|\mathbf{p}'_l - \mathbf{q}'_l\| \in [0,1]$. To calculate $C^{k-1,k}_{p_i,p_j}$, we use a similar method to the post-processing of ball detection and tracking. For $l^{k-1}_{p_i}(t)$ and $l^k_{p_j}(t)$, $\mathbf{p}_r$ and $\mathbf{q}_1$ are the last and the first positions of the trajectory segments, respectively. $C^{k-1,k}_{p_i,p_j}$ is computed as

$$C^{k-1,k}_{p_i,p_j} = \exp\left\{-\frac{1}{c_4}\|\mathbf{p}_r - \mathbf{q}_1\|\right\} \tag{12}$$

where $\|\mathbf{p}_r - \mathbf{q}_1\|$ is the Euclidean distance of $\mathbf{p}_r$ and $\mathbf{q}_1$, c_4 is a normalization constant which is similar to c_3. Based on $V^{k-1,k}_{p_i,p_j}$ and $C^{k-1,k}_{p_i,p_j}$, the weight of graph edge $e^{k-1,k}_{p_i,p_j}$ is defined as follows:

$$e^{k-1,k}_{p_i,p_j} = V^{k-1,k}_{p_i,p_j} \cdot C^{k-1,k}_{p_i,p_j}. \tag{13}$$

The insight of trajectory segment selection based on graph model is to find the optimal path with the maximum joint probability over the time steps in the graph and select the segments represented by the nodes on the optimal path. Finding the optimal path of a graph is a typical dynamic programming problem. The Viterbi algorithm [37] is employed to extract it based on the constructed graph. The algorithm is described in Fig. 12. Let P^k_j be the optimal path ending at the j th node

1. Initialization: $\delta_i^1 = v_{p_i}^1$, $\psi_1(i) = 0$, $1 \le i \le N_1$ where N_1 is the number of node in interval 1.
2. Recursion: $\delta_j^k = \max_{1 \le i \le N_{k-1}} \left(\delta_i^{k-1} + e_{p_i,p_j}^{k-1,k} + v_{p_j}^k \right)$,
$\psi_k(j) = \arg\max_{1 \le i \le N_{k-1}} \left(\delta_i^{k-1} + e_{p_i,p_j}^{k-1,k} + v_{p_j}^k \right)$, $1 \le j \le N_k$, $2 \le k \le K$,
where N_k is the number of node in interval k, and K is the total number of the intervals involved in a dribbling period.
3. Termination: $q_K = \arg\max_{1 \le i \le N_k} \left(\delta_i^K \right)$.
4. Path backtracking: $q_k = \psi_{k+1}(q_{k+1})$, $k = K-1, K-2, \ldots, 1$.

Fig. 12. Trajectory segment selection using Viterbi algorithm.

in interval k, the notations in Fig. 12 are then explained as follows. δ_j^k is the sum of the nodes and edges weights along P_j^k, $\psi_k(j)$ is the index linking to the node in interval $k-1$ on P_j^k, $\{q_k\}_{k=1}^K$ is the optimal path and q_k is the optimal trajectory segment selected as the component of the aggregate trajectory in interval k of the dribbling period.

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 the aggregate trajectory. This is because the last ball trajectory represents the locus of the ball conducted by the attacker with shot-on-goal action or out-of-field. 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. Fig. 13 shows an example of the aggregate trajectory.

C. Play Region Sequence Generation

Play region is another crucial feature for tactic analysis, especially for route pattern identification. In our implementation [38], the field is divided into 15 areas as shown in Fig. 14(a). Symmetrical regions in the field are given the same labels thus resulting in six labels in Fig. 14(b).

We extract following three features for the identification. 1) Field line location which is represented in polar coordinates (ρ_i, θ_i) $i = 1, \ldots, N$ where ρ_i and θ_i are the i th radial and angular coordinates respectively and N is the total number of lines. 2) Goalmouth 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 circle 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.

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 as shown in Fig. 14. 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 an attack event. The operation manner of CN is shown in the Fig. 15. The input of the identification method is the video sequence and the output is the labeled sequence of the play regions. Finally, the play region sequence is generated by concatenating all the identified field labels according to the frame indexes.

V. Tactic Pattern Analysis

Based on our understanding and discussion with soccer professionals, the tactic patterns used in the attack events in the soccer game can be summarized into two categories: route pattern which is related to the attack route in the soccer field and interaction pattern which is related to the cooperative interaction among the players and the ball. These two categories are independent because the tactic patterns in each category are analyzed and discovered from different tactic perspectives. A pattern in one category can also be a pattern in another category. For example, a route pattern (e.g., side-attack) can also be an interaction pattern (dribbling-attack). Tactic pattern analysis is conducted based on the tactic representations (play region sequence and aggregate trajectory) and the tactic domain knowledge in the soccer game.

A. Route Pattern Recognition

Route pattern can be classified into side attack and central attack by the inference using play region sequence identified from the video sequence of the attack event.

Given the video sequence $V = \{f_1, \ldots, f_n\}$ of an attack event G where f_i is the i th frame, the corresponding play region sequence is $R = \{r_1, \ldots, r_n\}$ where r_i is the identified active field label of frame f_i. The vote that f_i contributes to G for the pattern classification is defined as

$$\text{Vote}(f_i) = \begin{cases} 1, & \text{if } \text{Re}\, g(r_i) = \text{side}-\text{attack} \\ -1, & \text{if } \text{Re} g(r_i) = \text{center}-\text{attack} \end{cases} \tag{14}$$

where $\text{Re}\, g(\cdot)$ is the function of the pattern classification for a play region based on the region label shown in Fig. 14(b)

$$\text{Re}\, g(r_i) = \begin{cases} \text{side}-\text{attack}, & \text{if } r_i = 1, 3, 5 \\ \text{center}-\text{attack}, & \text{if } r_i = 2, 4, 6 \end{cases}. \tag{15}$$

The final route pattern RP of the attack event G is determined by

$$\text{RP}(G) = \begin{cases} \text{side}-\text{attack}, & \text{if } \sum_{f_i \in V} \text{Vote}\,(f_i) \ge 0 \\ \text{center}-\text{attack}, & \text{if } \sum_{f_i \in V} \text{Vote}\,(f_i) < 0. \end{cases} \tag{16}$$

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

B. Interaction Pattern Recognition

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

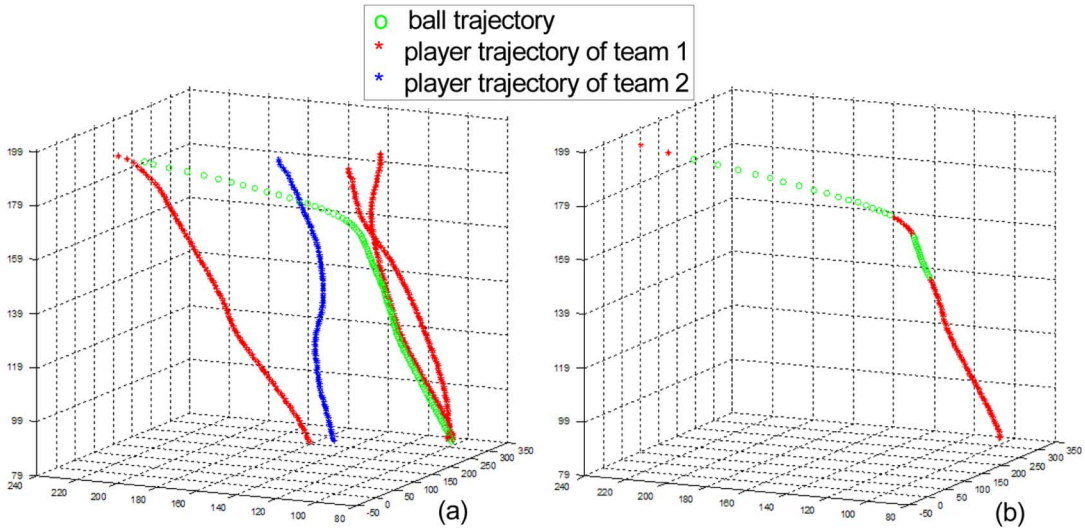

Fig. 13. Aggregate trajectory generation. (a) Mosaic trajectories of players and ball in an attack event. (b) Aggregate trajectory generated by temporal–spatial interaction analysis.

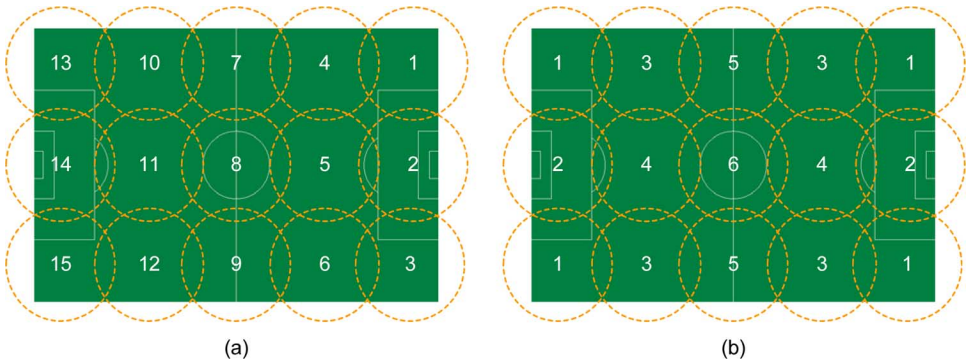

Fig. 14. Play region model: (a) 15 areas and (b) six labels.

1) Coarse Analysis for Pattern Recognition: At coarse step, the interaction pattern is classified into cooperative pattern and individual pattern. The cooperative pattern is defined as the tactics used in the attack event where the attack is carried out by multiple players via ball passing, and the individual pattern is defined as the tactics where the attack is carried out by only one player. The recognition is conducted on the aggregate trajectory of the attack event.

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

$$C_{\mathrm{coarse}}(G) = \sum_{i=1}^{n} \mathrm{ball}\,(s_i) \tag{17}$$

where function $\mathrm{ball}(\cdot)$ is defined as (18), shown at the bottom of the page. Consequently, we can classify the interaction pattern of the attack event at the coarse level as (19), shown at the bottom of the page.

$$\mathrm{ball}(x) = \begin{cases} 1, & \text{if } x \text{ is the segment of ball trajectory} \\ 0, & \text{if } x \text{ is not the segment of ball trajectory.} \end{cases} \tag{18}$$

$$IP_{\mathrm{coarse}}(G) = \begin{cases} \mathrm{cooperative{-}attack}, & \text{if } C_{\mathrm{coarse}}(G) > 0 \\ \mathrm{individual{-}attack}, & \text{if } C_{\mathrm{coarse}}(G) = 0. \end{cases} \tag{19}$$

Input: video sequence of attack event V
Output: label sequence of play region $j\#$

1. At frame t, extract every field line $(\rho_{t,i},\theta_{t,i})$ together with the goalmouth $(x_{t,g},y_{t,g})$ and central circle $(x_{t,e},y_{t,e})$ to form the feature vector $\mathbf{v}_i(t)=(\rho_{t,i},\theta_{t,i},x_{t,g},y_{t,g},x_{t,e},y_{t,e})^T$ where $i=1,\ldots,N$, N is the number of detected lines.
2. The response of each node in CN is calculated as $r_j(t)=\sum_{i=1}^{N}\mathbf{w}_j\cdot\mathbf{v}_i$, where $\mathbf{w}_j=(w_{j1},\ldots,w_{j6})$ is the weight vector associated with the jth node for 15 regions, $j=1,\ldots,15$. The set of wining nodes at frame t is $\{j^*(t)\}=\arg\max_j\{r_j(t)\}_{j=1}^{15}$.
3. The accumulated response is computed by $R_j(t)=R_j(t-1)+r_j(t)-\alpha\cdot Dist(j,j^*(t))-\beta$, where $R_j(t)$ is the accumulated response of node j, α is the scaling positive constant, β is the attenuation constant, and $Dist(j,j^*(t))$ is the Euclidean distance between node j to the nearest instantaneous wining node within the list $\{j^*(t)\}$.
4. The final output of CN at frame t is the node $j\#(t)$ with the maximal accumulated response $j\#(t)=\arg\max_j\{R_j(t)\}_{j=1}^{15}$.

Fig. 15. Flowchart of play region identification.

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 I.

a) Recognition for Cooperative Attack: For cooperative attack, we categorize the patterns into unhindered-attack and interceptive-attack according to whether there is ball-interception during the process of the attack. Given the aggregate trajectory AT of a cooperative attack, the subset $\mathrm{SAT}=\{sp_1,\ldots,sp_m\}$ of AT is extracted which only consists of player trajectories where sp_i is the trajectory segment. The elaborated criteria for cooperative attack recognition at the fine step is defined as

$$C_{\mathrm{fine\_}c}(G)=\sum_{i=2}^{m}\{1-\delta[\mathrm{player}(sp_i)-\mathrm{player}(sp_{i-1})]\} \tag{20}$$

where δ is the Kronecker delta function, $\mathrm{player}(x)$ is the function to identify the team affiliation (team 1 or team 2) for segment x based on the color histogram extracted from the tracked player region in the process of player detection and tracking. Therefore, we can classify the cooperative attack as follows:

$$\mathrm{IP}_{\mathrm{fine\_}c}(G)=\begin{cases}\mathrm{unhindered-attack}, & \text{if } C_{\mathrm{fine\_}c}(G)=0\\ \mathrm{interceptive-attack}, & \text{if } C_{\mathrm{fine\_}c}(G)>0.\end{cases} \tag{21}$$

b) Recognition for Individual Attack: The individual attack is classified into direct-attack and dribbling-attack according to the ball-dribbling occurrence in the attack process. Direct-attack mainly corresponds to the penalty or free kick, while dribbling-attack corresponds to the attack with ball-dribbling. To differentiate two patterns, hypothesis testing is conducted on the spatial distribution of the aggregate trajectory.

By the observation as shown in Fig. 16(a) and (c), the spatial positions of the aggregate trajectory of direct-attack subject to a line distribution compared with dribbling-attack. This observation is further verified by projecting the trajectory into the 2-D space as shown in Fig. 16(b) and (d). Such evidence is easily demonstrated by the process of two patterns in the real game. For penalty/free 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, for dribbling-attack, the player has to dribble the ball to avoid the defensive players which leads to a flexuous trajectory. The individual pattern recognition is conducted using hypothesis testing based approach.

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

$$\begin{cases}H_0: f(X,Y\,|\,k,c)=0\\ H_1: f(X,Y\,|\,k,c)\neq 0\end{cases} \tag{22}$$

where $X=\{x_1,\ldots,x_n\}$ and $Y=\{y_1,\ldots,y_n\}$ are the sets 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 line fitting of the underlying trajectory data. We first use the least square method [39] to estimate the line fitting function $y=L(x\,|\,k,c)$. Then, the average accumulative error (AE) for the given AT is calculated as follows:

$$\mathrm{AE}=\sum_{i=1}^{n}[y_i-L(x_i\,|\,k,c)]/n. \tag{23}$$

According to (23), AE is larger when X and Y are not fitted to a line distribution. Thus, we can classify the individual attack as follows:

$$\mathrm{IP}_{\mathrm{fine\_}i}(G)=\begin{cases}\mathrm{direct-attack}, & \text{if AE } \leq \mathrm{thres}\\ \mathrm{dribbling-attack}, & \text{if AE } > \mathrm{thres}\end{cases} \tag{24}$$

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

VI. Tactic Mode Presentation

With the analyzed results, two issues need to be considered for the information presentation in a tactic mode: 1) the presentation should be provided clearly and concisely so that the users can easily understand the tactic strategies used in the game, and 2) the presentation should provide essentially usable information so that the users can make further strategic analysis according to their personal requirement. The following information extracted from our tactic analysis is selected for the presentation in the tactic mode.

- *Time stamp for event occurrence* which is obtained from the web-casting text analysis.
- *Team labels in terms of offensive and defensive* which is extracted from web-casting text.

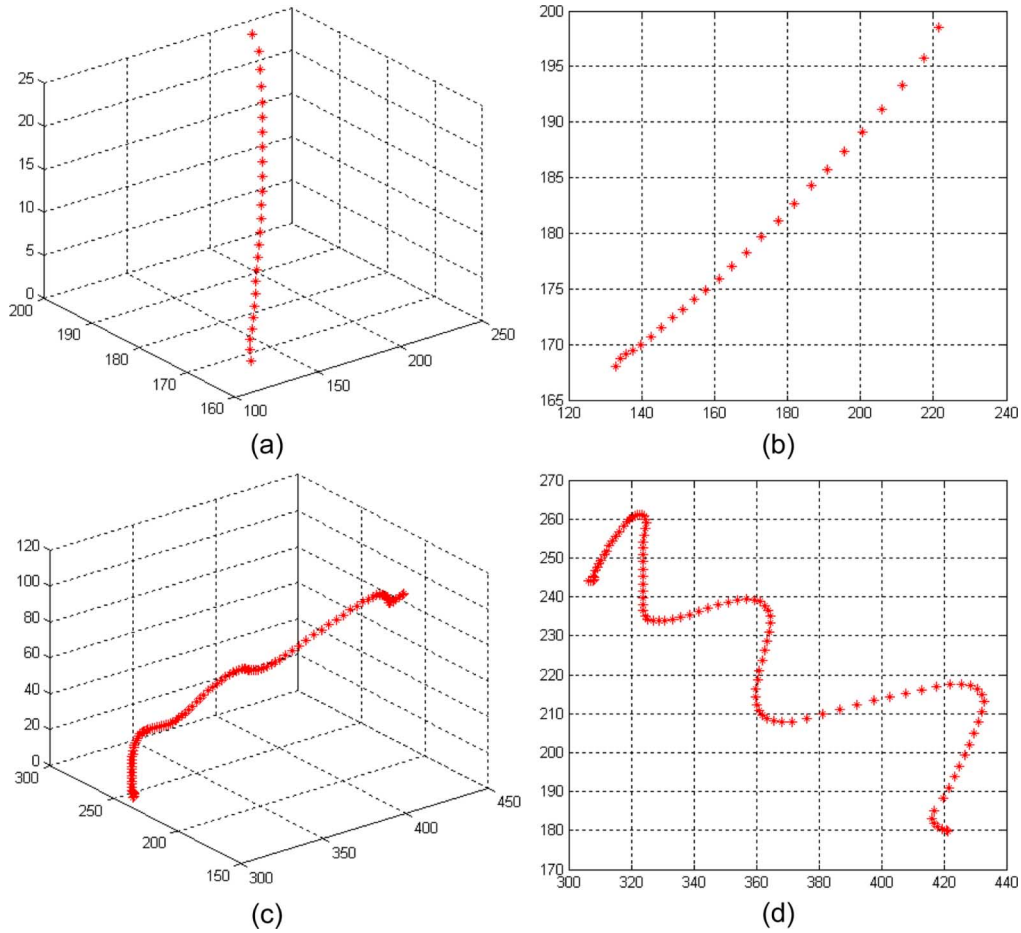

Fig. 16. Spatial distribution of aggregate trajectory. (a) and (c) 3-D distribution of AT of direct-attack and dribbling-attack respectively; (b) and (d) corresponding 2-D projected distribution.

TABLE III
SEMANTIC CATEGORIES ANNOTATION OF SELECTED EVENTS

Event category	# Selected event
Goal	168
Shot	413
Corner	125
Free kick	529
Total	1235

TABLE IV
TACTIC PATTERNS LABELED BY MANUAL ANNOTATION

Tactic pattern			# Selected event
Route pattern	side attack		726
	center attack		509
Interaction pattern	cooperative attack	unhindered-attack	662
		interceptive-attack	134
	individual attack	direct-attack	311
		dribbling-attack	128

- *Trajectories of the ball, offensive players and defensive players respectively* which are extracted by multi-object detection and tracking.
- *Route pattern* (*side- or center-*) which is recognized by route pattern recognition.

TABLE V
SEMANTIC CATEGORIES ANNOTATION OF SELECTED EVENTS

Event category	*BDA* (%)
Goal	92.2
Shot	91.5
Corner	83.9
Free kick	64.6
Average	83.1

- *Interaction pattern* (*two categories at coarse and fine levels respectively*) which is classified by interaction pattern recognition.

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

VII. EXPERIMENTAL RESULTS

To demonstrate the effectiveness of our proposed approaches, we conducted 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 frame resolution of 704×576. All the correspondent web-casting text files were collected from [22].

TABLE VI
RESULTS OF MULTI-OBJECT (PLAYERS AND BALL) DETECTION AND TRACKING

Pattern	Player detection & tracking					Ball detection & tracking		
		Enhanced approach		Previous approach [26]				
	# *SOPs*	Precision (%)	Recall (%)	Precision (%)	Recall (%)	# *SOBs*	Precision (%)	Accuracy (%)
unhindered-attack	969757	79.0	89.0	71.7	80.4	104275	85.5	83.9
interceptive-attack	210437	77.2	82.3	62.4	71.9	23125	83.1	80.3
direct-attack	31500	98.6	95.3	93.5	92.5	21000	94.6	92.6
dribbling-attack	86040	94.3	92.2	90.8	88.6	17925	92.3	89.4
Total	1297734	83.7	88.1	72.0	79.9	166325	88.4	85.1

To obtain the ground truth of the test data in terms of semantic category and tactic pattern, we invited five soccer professionals who are very familiar to the game events and soccer tactics to manually select and annotate the game video by a voting scheme, in which the semantic and tactic categories with the majority votes are labeled to the selected event. Two criteria were considered in the event selection and tactic annotation. Firstly, since generally the longer duration the event persists the more comprehensive scenario the event presents, the segment of the selected event is expected as long as possible to keep the event integrity. Secondly, since various tactic scenarios are applied to World Cup games, the diversity of the tactic scenarios annotated in the selected events is expected to be included as much as possible to cover all the game strategies. Finally, we manually selected 1235 attack events from all the matches. The details about the selected events with respects to semantic categories and tactic patterns are listed in Tables III and IV, respectively. The manual annotation was adopted as the ground truth for the comparison with the results of automatic analysis. Half data of each event category were selected as the training set to construct the SVC/PSVC models for player trajectories acquisition and optimize the parameter values configured in the proposed approach. Then, the experimental evaluation was conducted on all the data.

A. Performance of Attack Event Detection

To assess the suitableness of the automatically extracted event, we used the boundary detection accuracy (BDA) [10] to measure the detected event boundary compared with the manually labeled ground truth, where BDA is defined as

$$\mathrm{BDA} = \frac{\tau_{db} \cap \tau_{mb}}{\max(\tau_{db}, \tau_{mb})} \tag{25}$$

where τ_{db} and τ_{mb} are the automatically detected event boundary and the manually labeled event boundary, respectively. The higher the BDA score, the better the performance. Table V lists the evaluation of our proposed method with the BDA scores for four semantic event categories. It is observed that the performance of free kick events is lower than other events. This is because our selected web-casting text usually includes other event, e.g., foul, before the free kick event, and thus the extracted time stamp is not accurate, which affects the alignment accuracy.

B. Performance of Tactic Information Extraction

1) Results of Multi-Object Detection and Tracking: The performance evaluation of multi-object detection and tracking was conducted on the video sequences of all the selected events according to four interaction patterns. The interaction relationship in the video, such as player-player occlusion and player-ball occlusion, can be employed as the challenging test bed for multi-object detection and tracking to verify the effectiveness of the proposed method. The detailed experimental statistics for object (player and ball) detection and tracking are listed in Table VI. In Table VI, SOPs means sum of players appeared in all the frames, e.g., if three players in the first frame and four players in the second frame, then SOPs is seven. Similarly, SOBs which is the abbreviation of sum of balls is defined for the evaluation of ball detection and tracking.

As shown in Table VI, the object detection and tracking method achieves average precision/recall of 83.7%/88.1% for the player and 88.4%/85.1% for the ball. The incorrect detection and tracking results are due to several reasons: 1) the player region is sometimes so small that it is falsely treated as noise in the background, 2) the player region in the frame is very close to the caption, logo, or mark lines in the field that may lead to its merger with the adjacent region when performing post-processing after playfield detection, 3) the ball is totally occluded by the players or merged with the mark lines which may lead to incapable detection, and 4) the socks of players or the mark line segments are sometimes most likely to be the ball resulting in the false detection and tracking.

A comparison with our previous tracking method [26] was also carried out on the same video data. Table VI summarizes the comparative results. From this comparison, it can be seen that the precision/recall evaluation of the four interaction patterns are increased from 3.5%/2.8% to 14.8%/10.4%, especially for interceptive-attack which the increment of accuracy achieves 14.8%/10.4%. In the scenes of interceptive-attacks, there are many occlusions among the players due to the defensive tackle and body check. Using the enhanced tracking approach, such problem is solved effectively and the remarkable result is achieved. For the other three patterns, the mechanism of proposal distribution construction by integrating PSVC improves the robustness of previous approach and achieves satisfying results.

TABLE VII
RESULTS OF PLAY REGION IDENTIFICATION

Region label	Precision (%)	Recall (%)	Region label	Precision (%)	Recall (%)
1	82.1	86.0	2	98.3	89.6
3	76.6	77.1	4	61.1	68.9
5	87.0	84.3	6	91.4	100

TABLE VIII
RESULTS OF ROUTE PATTERN RECOGNITION

Pattern	n_c	n_m	n_f	R (%)	P (%)
side-attack	641	85	91	88.3	87.6
center-attack	418	91	85	82.1	83.1

2) Results of Play Region Identification: In this experiment, our active play region identification method was used to recognize the play region in each frame as illustrated in Fig. 14(b). The result was then compared with our manually identified region labels and the accuracy is listed in Table VII. It is noted that the detection accuracy for region 3 and 4 are low compared with other labels. This is because that these field regions have fewer cues than the other regions, e.g., it does not have field lines or goalmouth or central circle. The lack of distinct information thus results in poor accuracy. This agrees with our previous work in [38].

C. Performance of Tactic Pattern Recognition

1) Results of Route Pattern Recognition: Using the proposed route pattern recognition, we classified 1235 attack events into two clusters. We calculated Recall (R) and Precision (P) to quantitatively evaluate the performance, which are defined as

$$R = n_c/(n_c + n_m). \tag{26}$$

$$P = n_c/(n_c + n_f). \tag{27}$$

where for each pattern, n_c is the number of attacks correctly recognized, n_m is the number of missed attacks, and n_f is the number of attacks false-alarmed. Table VIII shows the recognition results.

2) Results of Interaction Pattern Recognition: The performance of interaction pattern recognition is evaluated using all the attack events. Multiple trajectories of the players and the ball were extracted to construct the aggregate trajectory. The coarse and fine criteria were computed according to (19), (21), and (24). The metrics R and P defined in (26) and (27) were used to evaluate the performance. The results for the recognition of coarse and fine tactic patterns are listed in Tables IX and X, respectively.

It is observed from Tables IX and X that the performance of the proposed tactic analysis approach is promising. The key issues affecting the recognition results in terms of coarse and fine patterns can be summarized from two perspectives. 1) The robustness of multi-object detection and tracking: Although we enhance the performance of the object detection and tracking approach, there still exist some conditions that the occlusion among the objects is so severe that even the human cannot identify the affiliation of the individual trajectory. 2) The accumulative error of mosaic transform: Mosaic trajectory computation is employed to eliminate the camera motion in 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 when the GME mapping matrices are used in the long-term transform. Consequently, the computed mosaic trajectory does not reflect the insight of the movement of the players and the ball.

TABLE IX
RESULTS OF COARSE INTERACTION PATTERN RECOGNITION

Pattern	n_c	n_m	n_f	R (%)	P (%)
cooperative-attack	712	84	52	89.4	93.2
individual-attack	387	52	84	88.2	82.2

TABLE X
RESULTS OF FINE INTERACTION PATTERN RECOGNITION

Pattern	n_c	n_m	n_f	R (%)	P (%)
unhindered-attack	583	79	71	88.1	89.1
interceptive-attack	100	34	23	74.6	81.3
direct-attack	261	50	41	83.9	86.4
dribbling-attack	109	19	46	85.2	70.3

In addition, the comparison between graph based approach and previous deterministic approach [12] was conducted on the data set of goal events. As shown in Table XI, the results of four patterns using graph based approach are all improved compared with the deterministic approach. This can be explained as follows. 1) As noted in the previous sections, multi-object detection and tracking is the fundamental task to obtain the trajectories of the players and the ball for the tactic representation construction. In this paper, the existing object tracking method is enhanced to eliminate the mutual occlusion among the objects which improves the tracking result. 2) The graph modeling is introduced for the trajectory temporal–spatial analysis to construct the aggregate trajectory. Graph model holds the multiple hypotheses of the candidates of the components of the aggregate trajectory and the Viterbi algorithm considers the temporal structure to select the optimal trajectory segments, both of which increase the robustness of aggregate trajectory construction.

D. Tactic Mode Presentation User Study

The objective of this user study is to evaluate the applicability of the tactic pattern presentation with the selected information. To carry out the evaluation, we employed a subjective user study [40] as there is no objective measure available to evaluate the quality of a presentation fashion.

Five professionals who selected and annotated the attack events in the game videos were invited in the subjective study. Of the five people, two are soccer coaches and three are soccer players who have more than five-year team training and four-year professional game playing experience respectively. All the subjects have rich knowledge of the tactic strategies used in the soccer game. To conveniently facilitate the study, we designed a program for attack event browsing and tactic information presentation. Note that the subjects can choose

TABLE XI
COMPARISON BETWEEN GRAPH BASED AND DETERMINISTIC APPROACHES USING 168 GOAL EVENTS

Pattern	Graph based approach					Deterministic approach				
	n_c	n_m	n_f	R (%)	P (%)	n_c	n_m	n_f	R (%)	P (%)
unhindered-attack	85	10	7	89.5	92.4	81	14	12	85.3	87.1
interceptive-attack	12	4	3	75.0	80.0	11	5	3	68.8	78.6
direct-attack	35	6	4	85.4	89.7	33	8	6	80.5	84.6
dribbling-attack	14	2	5	87.5	73.7	13	3	8	81.3	61.9

TABLE XII
SUBJECTIVE USER STUDY ON TACTIC MODE PRESENTATION

	Conciseness	Clarity	Usability
Subject 1	4.6	4.5	4.8
Subject 2	4.7	4.2	4.6
Subject 3	4.1	4.3	4.7
Subject 4	4.0	4.7	4.3
Subject 5	3.9	4.2	4.4

four different types of trajectories to watch including only the trajectories of offensive players, only the trajectories of defensive players, only the ball trajectory and all the trajectories of players and ball.

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 without tedious content.
- *Clarity*: the information presented is explicit and easy to be understood.
- *Usability*: the information can be used for the further analysis and benefits for the later training and games.

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 XII shows the result of subjective evaluation. It can be seen that the average evaluation results are 4.26, 4.38, and 4.56 for three criteria respectively. This demonstrates that our tactic presentation modal is accepted by the soccer professionals.

VIII. CONCLUSION

Compared with semantic analysis, tactics analysis provides more tactic insight of sports game but so far little work has been devoted to this topic. In this paper, we have presented a novel approach to discover the tactic patterns from the attack events in broadcast soccer video.

As a team sports, the cooperation among the players and the interaction among the players and the ball characterize the tactic patterns used in the soccer game. Accordingly, two tactic representations, which are aggregate trajectory and play region sequence, are constructed based on the multiple trajectories of the players and the ball and active field locations to discover the tactic insight of the game. The tactic clues are extracted from two representations to conduct the pattern analysis of the attack events. The patterns are classified as route pattern and interaction pattern in which more elaborated tactic scenarios are analyzed. We carried out the experiments on the selected attack events from FIFA World Cup 2006. The results demonstrate that our approach is effective.

To the best of our knowledge, our tactic analysis approach is the first solution for the soccer game based on broadcast video. Several issues will be further studied in our future work. 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 general for 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 approach will be extended to other team sports video such as hockey and American football.

REFERENCES

[1] Y. Gong, T. S. Lim, H. C. Chua, H. J. Zhang, and M. Sakauchi, "Automatic parsing of TV soccer programs," in *Proc. Int. Conf. Multimedia Computing and System*, Washington, DC, 1995, pp. 167–174.
[2] A. Ekin, A. M. Tekalp, and R. Mehrotra, "Automatic soccer video analysis and summarization," *IEEE Trans. Image Process.*, vol. 12, no. 7, pp. 796–807, Jul. 2003.
[3] Y. Rui, A. Gupta, and A. Acero, "Automatically extracting highlights for baseball programs," in *Proc. ACM Multimedia*, Los Angeles, CA, 2000, pp. 105–115.
[4] J. Assfalg, M. Bertini, C. Colombo, A. Delbimbo, and W. Nunziati, "Semantic annotation of soccer video: Automatic highlights identification," *Comput. Vis. Image Understand.*, vol. 92, no. 2–3, pp. 285–305, 2003.
[5] 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, Aug. 2004.
[6] A. Hanjalic, "Adaptive extraction of highlights from a sport video based on excitement modeling," *IEEE Trans. Multimedia*, vol. 7, no. 6, pp. 1114–1122, Dec. 2005.
[7] 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*, Berkeley, CA, 2003, pp. 33–44.
[8] 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.
[9] 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*, Santa Barbara, CA, 2006, pp. 431–440.
[10] 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*, Santa Barbara, CA, 2006, pp. 221–230.
[11] H. Xu and T. S. Chua, "Fusion of av features and external information sources for event detection in sports video," *ACM Trans. Multimedia Computing, Communications, and Applications*, vol. 2, no. 1, pp. 44–67, 2006.
[12] G. Zhu, Q. Huang, C. Xu, Y. Rui, S. Jiang, W. Gao, and H. Yao, "Trajectory based event tactics analysis in broadcast sports video," in *Proc. ACM Multimedia*, Augsburg, 2007, pp. 58–67.

[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*, Bombay, India, 1998, pp. 81–90.
[14] G. S. Pingali, Y. Jean, and I. Carlbom, "Real time tracking for enhanced tennis broadcasts," in *Proc. Conf. Computer Vision and Pattern Recognition*, Santa Barbara, CA, 1998, pp. 260–265.
[15] J. R. Wang and N. Parameswaran, "Analyzing tennis tactics from broadcasting tennis video clips," in *Proc. Int. Conf. Multimedia Modeling*, Melbourne, Australia, 2005, pp. 102–106.
[16] 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*, Tokyo, 2004, pp. 49–56.
[17] N. Babaguchi, Y. Kawai, and T. Kitahashi, "Event based indexing of broadcasted sports video by intermodal collaboration," *IEEE Trans. Multimedia*, vol. 4, no. 1, pp. 68–75, Feb. 2002.
[18] M. Han, W. Hua, W. Xu, and Y. Gong, "An integrated baseball digest system using maximum entropy method," in *Proc. ACM Multimedia*, Juan-les-Pins, 2002, pp. 347–350.
[19] 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*, Lausanne, 1996, vol. 3, pp. 815–818.
[20] 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*, Pisa, 2004, vol. 3202, pp. 209–218.
[21] C. H. Kang, J. R. Hwang, and K. J. Li, "Trajectory analysis for soccer players," in *Proc. Int. Conf. Data Mining Workshops*, Hong Kong, 2006, pp. 377–381.
[22] , [Online]. Available: http://soccernet.espn.go.com, [Online]. Available:
[23] , [Online]. Available: http://ww.desearch.com/, dtsearch dtSearch Corp [Online]. Available:, 6.50
[24] J. Wang, E. Chng, C. Xu, H. Lu, and Q. Tian, "Generation of personalized music sports video using multimodal cues," *IEEE Trans. Multimedia*, vol. 9, no. 3, pp. 576–588, Apr. 2007.
[25] Y. Li, C. Xu, K. Wan, X. Yan, and X. Yu, "Reliable video clock time recognition," in *Proc. Int. Conf. Pattern Recognition*, Hong Kong, 2006, vol. 4, pp. 128–131.
[26] 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*, Toronto, ON, Canada, 2006, pp. 1629–1632.
[27] Y. Liu, S. Jiang, Q. Ye, W. Gao, and Q. Huang, "Playfield detection using adaptive GMM and its application," in *Proc. Int. Conf. Acoustics, Speech, and Signal Processing*, Philadelphia, PA, 2005, vol. 2, pp. 421–424.
[28] Y. Rui and Y. Chen, "Better proposal distributions: Object tracking using unscented particle filter," in *Proc. Conf. Computer Vision and Pattern Recognition*, Hawaii, 2001, vol. 2, pp. 786–793.
[29] K. Okuma, A. Taleghani, N. de Freitas, J. J. Little, and D. G. Lowe, "A boosted particle filter: Multitarget detection and tracking," in *Proc. Eur. Conf. Computer Vision*, Prague, Czech Republic, 2004, vol. 1, pp. 28–39.
[30] Y. Cai, N. de Freitas, and J. J. Little, "Robust visual tracking for multiple targets," in *Proc. Eur. Conf. Computer Vision*, Graz, Austria, 2006, vol. 4, pp. 107–118.
[31] P. Viola and M. Jones, "Rapid object detection using a boosted cascade of simple features," in *Proc. Conf. Computer Vision and Pattern Recognition*, Hawaii, 2001, vol. 1, pp. 511–518.
[32] V. Vapnik, *The Nature of Statistical Learning Theory.*. New York: Spinger-Verlag, 1995.
[33] J. C. Platt, "Probabilistic outputs for support vector machines and comparisons to regularized likelihood methods," in *Advances in Large Margin Classifiers*. Cambridge, MA: MIT Press, 1999, pp. 185–208.
[34] 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*, Jeju Island, Korea, 2005, pp. 864–875.
[35] X. Yu, H. W. Leong, C. Xu, and Q. Tian, "Trajectory-based ball detection and tracking in broadcast soccer video," *IEEE Trans. Multimedia*, vol. 8, no. 6, pp. 1164–1178, Dec. 2006.
[36] F. Dufaux and J. Konrad, "Efficient, robust, and fast global motion estimation for video coding," *IEEE Trans. Image Process.*, vol. 9, no. 3, pp. 497–501, Mar. 2000.
[37] L. R. Rabiner, "A tutorial on hidden Markov model and selected applications in speech recognition," *Proceedings of the IEEE*, vol. 77, no. 2, pp. 257–285, 1989.
[38] J. Wang, C. Xu, E. Chng, K. Wan, and Q. Tian, "Automatic replay generation for soccer video broadcasting," in *Proc. ACM Multimedia*, New York, 2004, pp. 32–39.
[39] G. A. Korn and T. M. Korn, *Math Handbook for Scientists and Engineers*. New York: McGraw-Hill, 1968.
[40] J. Chin, V. Diehl, and K. Norman, "Development of an instrument measuring user satisfaction of the human-computer interface," in *Proc. SIGCHI on Human Factors in CS*, Washington, DC, 1998, pp. 213–218.

Guangyu Zhu received the B.S. and M.S. degrees in computer science from the Harbin Institute of Technology, Harbin, China, in 2001 and 2003, respectively, where he is currently pursuing the Ph.D. degree.

His research interests include image/video processing, multimedia content analysis, computer vision and pattern recognition, and machine learning.

Changsheng Xu (M'97–SM'99) received the Ph.D. degree from Tsinghua University, Beijing, China in 1996.

Currently, he is Professor of Institute of Automation, Chinese Academy of Sciences and Executive Director of China-Singapore Institute of Digital Media. He was with Institute for Infocomm Research, Singapore, from 1998 to 2008. He was with the National Lab of Pattern Recognition, Institute of Automation, Chinese Academy of Sciences from 1996 to 1998. His research interests include multimedia content analysis, indexing and retrieval, digital watermarking, computer vision and pattern recognition. He published over 150 papers in those areas.

Dr. Xu is an Associate Editor of ACM/Springer Multimedia Systems Journal. He served as Short Paper Co-Chair of ACM Multimedia 2008, General Co-Chair of 2008 Pacific-Rim Conference on Multimedia (PCM2008) and 2007 Asia-Pacific Workshop on Visual Information Processing (VIP2007), Program Co-Chair of VIP2006, Industry Track Chair and Area Chair of 2007 International Conference on Multimedia Modeling (MMM2007). He also served as Technical Program Committee Member of major international multimedia conferences, including ACM Multimedia Conference, International Conference on Multimedia & Expo, Pacific-Rim Conference on Multimedia, and International Conference on Multimedia Modeling.

Qingming Huang (M'04) received the Ph.D. degree in computer science from Harbin Institute of Technology, Harbin, China in 1994.

He was a Postdoctoral Fellow with the National University of Singapore from 1995 to 1996, and worked in Institute for Infocomm Research, Singapore, as a Member of Research Staff from 1996 to 2002. Currently, he is a Professor with the Graduate University of Chinese Academy of Sciences. He has published over 100 scientific papers and granted/filed more than ten patents in US, Singapore and China. His current research areas are multimedia processing, video analysis, pattern recognition and computer vision.

Dr. Huang was the Organization Co-Chair of 2006 Asia-PacificWorkshop on Visual Information Processing (VIP2006) and served as Technical Program Committee Member of various international conferences, including ACM Multimedia Conference, International Conference on Multimedia and Expo, International Conference on Computer Vision and International Conference on Multimedia Modeling.

Yong Rui (M'99–SM'04) received the Ph.D. degree from University of Illinois at Urbana-Champaign.

he serves as Director of Strategy of Microsoft China R&D (CRD) Group. Before this role, he managed the Multimedia Collaboration team at Microsoft Research, Redmond, WA. He contributes significantly to the research communities in computer vision, signal processing, machine learning, and their applications in communication, collaboration, and multimedia systems. His contribution to relevance feedback in image search created a new research area in multimedia. He has published twelve books and book chapters, and over seventy referred journal and conference papers. Dr. Rui holds 30 issued and pending U.S. patents.

Dr. Rui is a senior member of both ACM and IEEE. He is an associate editor of ACM *Transactions on Multimedia Computing, Communication and Applications* (TOMCCAP), IEEE TRANSACTIONS ON MULTIMEDIA, and IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGIES. He was an editor of ACM/Springer *Multimedia Systems Journal* (2004–2006), *International Journal of Multimedia Tools and Applications* (2004–2006), and IEEE Transactions on Multimedia (2004–2008). He also serves on the advisory board of IEEE Transactions on Automation Science and Engineering. He was on organizing committees and program committees of ACM Multimedia, IEEE CVPR, IEEE ECCV, IEEE ACCV, IEEE ICIP, IEEE ICASSP, IEEE ICME, SPIE ITCom, ICPR, CIVR, among others. He is a general chair of International Conference on Image and Video Retrieval (CIVR) 2006, a program chair of ACM Multimedia 2006, and a program chair of Pacific-Rim Conference on Multimedia (PCM) 2006.

Shuqiang Jiang (M'06) received the M.Sc. degree from College of Information Science and Engineering, Shandong University of Science and Technology, China, in 2000, and the Ph.D. degree from the Institute of Computing Technology, Chinese Academy of Sciences, Beijing, in 2005.

He is currently a faculty member at Digital Media Research Center, Institute of Computing Technology, Chinese Academy of Sciences. He is also with the Key Laboratory of Intelligent Information Processing, Chinese Academy of Sciences. His research interests include multimedia processing and semantic understanding, pattern recognition, and computer vision. He has published over 40 technical papers in the area of multimedia.

Wen Gao (M'92–SM'05–F'09) received M.S. degree in computer science from Harbin Institute of Technology in 1985, and Ph.D. degree in electronics engineering from the University of Tokyo in 1991.

He was a Professor in computer science at Harbin Institute of Technology from 1991 to 1995, a Professor in computer science at Institute of Computing Technology of Chinese Academy of Sciences from 1996 to 2005. He is currently a Professor at the School of Electronics Engineering and Computer Science, Peking University, China. He has been leading research efforts to develop systems and technologies for video coding, face recognition, sign language recognition and synthesis, and multimedia retrieval. He earned many awards include five national awards by research achievements and activities. He has published four books and over 500 technical articles in refereed journals and proceedings in the areas of signal processing, image and video communication, computer vision, multimodal interface, pattern recognition, and bioinformatics.

Dr. gao was general co-chair of IEEE ICME07, and the head of Chinese delegation to the Moving Picture Expert Group (MPEG) of International Standard Organization (ISO) since 1997, he is also the chairman of the working group responsible for setting a national Audio Video coding Standard (AVS) for China. He published four books and over 500 technical articles in refereed journals and proceedings in the areas of signal processing, image and video communication, computer vision, multimodal interface, pattern recognition, and bioinformatics.

Hongxun Yao (M'00) received the B.S. and M.S. degrees in computer science from the Harbin Shipbuilding Engineering Institute, Harbin, China, in 1987 and in 1990, respectively, and the Ph.D. degree in computer science from Harbin Institute of Technology in 2003. Currently, she is a Professor with the School of Computer Science and Technology, Harbin Institute of Technology. Her research interests include pattern recognition, multimedia technology, and human-computer interaction technology. She has published three books and over 100 scientific papers.