GameFlow: Narrative Visualization of NBA Basketball Games

Wei Chen, Tianyi Lao, Jing Xia, Xinxin Huang, Biao Zhu, Wanqi Hu, Huihua Guan

***Abstract*—Although basketball games have received broad attention, the forms of game reports and webcast are purely content-based cross-media: texts, videos, snapshots, and performance figures. Analytical narrations of games that seek to compose a complete game from heterogeneous datasets are challenging for general media producers because such a composition is time-consuming and heavily depends on domain experts. In particular, an appropriate analytical commentary of basketball games requires two factors, namely, rich context and domain knowledge, which includes game events, player locations, player profiles, and team profiles, among others. This type of analytical commentary elicits a timely and effective basketball game data visualization made up of different sources of media. Existing visualizations of basketball games mainly profile a particular aspect of the game. Therefore, this paper presents an expressive visualization scheme that comprehensively illustrates NBA games with three levels of details: a season level, a game level and a session level. We reorganize a basketball game as a sequence of sessions to depict the game states and heated confrontations. We design and implement a live system that integrates multi-media NBA datasets: play-by-play text data, box score data, game video data, and action area data. We demonstrate the effectiveness of this scheme with case studies and user feedbacks.**

***Index Terms*—Game narration, basketball games, visualization, video.**

I. Introduction

Basketball is one of the most popular and widely watched ball games all over the world. It is a fast-paced contact game in which the players are constantly moving in heated confrontations, thus leading to quick transitions from defense to offense or vice versa. Analytical narrations of games aim to compose a complete game from different perspectives. This composition is time-consuming and heavily depends on domain experts. Thus, a timely and effective analysis of the games has been greatly demanded in basketball game narration. For instance, commentators need timely comprehension of the games to quickly gain insight and detect interesting patterns that may be hidden, i.e., the performance of players or the tactics used in the games. Nevertheless, the existing analysis of basketball games is less effective and time consuming. Users still heavily rely on watching lengthy videos and reading statistical tables to examine the fast-changing games. Developing an effective and efficient solution that facilitate the analysis of basketball games is urgently needed.

Wei Chen, Tianyi Lao, Jing Xia, Xinxin Huang, Biao Zhu, Wanqi Hu, Huihua Guan were with the State Key Lab of CAD&CG, Zhejiang University. Being a devotee of basketball, Xinxin Huang came up with the idea of basketball game visualization and implemented the very first version of the system.

Manuscript received April 19, 2005; revised August 26, 2015.

This paper aims to use multimedia datasets, such as play-by-play (PBP) text data, action areas data, basic statistics, and game video data of NBA games to enable the audience, such as basketball devotees and commentators, to conduct game narration. The PBP data contain abundant information on the players, time-outs, and score series, etc. and describe all changes in a game. The PBP data also describe player actions, such as shot attempts, turnovers, possession of the ball, fouls, etc. Action area data, which indicate where actions take place, are utilized to further analyze the attack styles of teams. Basic statistical data, such as field goals and rebounds, provide a numeric overview of the performance of a team or individual players in a game. Game video data record every movement of a game. We use game video data only for review and verification as analyzing a complete game process with a game video takes at least 48 minutes. Overall, multimedia NBA datasets enable a multi-facet analysis in characterizing the entire proceeding of a season and individual games. Users can perform in-depth, comprehensive game narration by evaluating the performance of teams and players and by understanding the unfolding of a game among others. The integration of heterogeneous datasets enables users to explore the game from different perspectives with a rich game context.

Quantitative analysis using data mining methods or statistical methods, has been employed by expert users to analyze the game data. The book *Basketball on Paper* [1] emphasizes the power of quantitative analysis in basketball game analysis. However, these methods have two obvious shortcomings. First, these methods usually depend on well-trained models and predefined features [2], [3] that require domain knowledge. Thus, they cannot be easily interpreted by non-expert users. Second, the methods can successfully detect well-known patterns, but can hardly find unexpected ones. Finally, the results produced by the methods may still require more work to read and understand, not to mention to reach a wide audience. To address these issues, an increasing number of visualization research has been conducted in recent years, such as visual analysis of player trajectories, visualization of field goals of a player [4], and visualization of basic statistics of different players in different games [5]. Nevertheless, these works only focus on a particular aspect of a game. A visual analysis system that enables non-expert users to conduct comprehensive game narration is still absent.

By integrating multimedia data of NBA basketball games, we provide a comprehensive representation of an entire NBA season. Our visualization reveals NBA games at multiple levels of details: the season level, the game level and the session level. The overall season view (the season level) visualizes the win-lose statistics and other statistics, thus enabling the comparison of the strengths and weaknesses of all teams. The progress of a game (the game level) is visualized as an ordered sequence of sessions segmented by the substitution of players. In a session (the session level), actions are designed as glyphs and the performers are directly revealed. Thus, the key player and his contributions in a session can be inferred. PBP data, video data and action areas data together facilitate multi-faceted visualization of a session. Based on the visual designs, our **GameFlow** system makes visual narrations of a complete proceeding of NBA games efficiently: action by action, player by player and session by session. Users can identify the changes in intenseness of a basketball game, and explore the contributing factors of the final score.

Our contributions are summarized as follows:

- An integration scheme of multimedia heterogeneous N-BA game datasets that enables a comprehensive visual narration of a complete basketball game.
- A level-of-detail exploration scheme that reveals the entire NBA season, an NBA game, and a game session progressively and interactively.

II. Related Work

A. Basketball Games Analysis

The quantitative analysis of basketball games can be divided into two categories: modeling methods and non-modeling methods. The Markov chain model is widely applied in modeling transitions in basketball games [2], [3]. In this model, each game is a sequence of transitions between discrete states. A well-fitted model can build a simulation of a basketball game to forecast the final scores of a game. In non-modeling methods, Bashuk [6] predicted the performance of a team and an individual player in basketball games with cumulative win probabilities. Hamilton et al. [7] modeled basketball free throws as a function of angle, speed, and spin at release and calculated the optimal trajectory. They then provided a function of the height of ball release for a higher hit rate. Perše et al. [8] proposed a trajectory-based approach with a probabilistic play model and team activities. The approach could automatically recognize complex multi-player behaviors in basketball games. Other than the trajectories in the trajectory-based approach, the quantitative analysis approaches do not utilize visualization. Thus they are usually lack of illustrative description of their analysis.

Instead of making predictions, visualization methods focus on representations, statistical values, or events on the basketball field. CourtVision [4] is an ensemble method that quantifies and visualizes the shooting range of NBA players and the differences in their shooting abilities. PluMP [5] augmented the traditional plus/minus plot to a measurement of impact through differentials to all box score statistics. The media has released various visualization reports on basketball. The Fathom information design [9] visualized basketball movement throughout the game as well as players' movements with a dataset generated by the latest SportVU tracking system. Unfortunately, the trajectory data are not publicly available. The New York Times [10] depicted basketball players' shooting positions with a shooting heatmap and realized mid-range jumpers are rare in the NBA. The HotShotChart [11] utilized an arc diagram to reveal the assist relationships between players in NBA games. All these great visualizations exhibit a basketball game with one or two perspectives. Nevertheless, because of the complexity of basketball game data, none of them can narrate a complete game with the changing scores and events.

B. Sports Visualization

Other than basketball game analysis, sports visualization has attracted much academic attention. Soccer Matches [12] is an application that provides an overview of soccer championship results with glyphs in a matrix representation. Although the overall championship season is visualized, Soccer Matches does not deal with individual soccer games. Integrating Twitter data, Wongsuphasawat [13] narrated the most interesting moments in the UEFA Champions League. The entire season and all the interesting moments in each game detected from Twitter were visualized. However, inter-player collaboration was not covered in their work. SoccerStories [14] enriches the current soccer analysis by aggregating game phases into a series of connected visualizations of actions such as goal attempts. Janetzko [15] built a prototype of soccer data that covers single-player, multi-player, and event-based aspects of soccer analysis. Note that basketball is a more intense game than soccer. Gathered in a much smaller field, players' movements and actions, such as shooting, assisting, and fouls, can happen in quick successions. This intensity also increases the challenges in visualization.

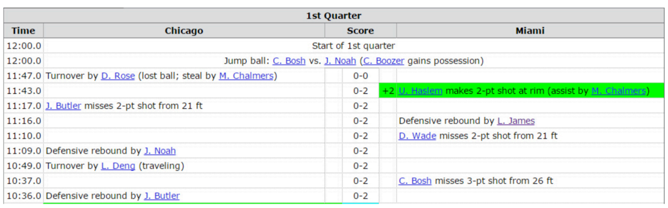
Other than soccer, other sports can also be visualized. Parry et al. [16] proposed a generic scheme for event prioritization and selection in a video storyboard. They demonstrated the scheme using a case study of a snooker video storyboard. However, the idea of visualizing trajectories with a video is not feasible with basketball because the game field is too large to cover all the details in a game. Introducing the radial heat map technique, SnapShot [17] integrates visualization into a hockey intelligence-gathering process. The focus of SnapShot is the shot data, specifically the shot length. Baseball4D [18] reconstructs baseball games and obtains a better baseball analysis by combining visualization and statistical methods. As baseball is not a fast-paced game like basketball, Baseball4D also focuses on ball and fielder trajectories. TenniVis [19] provides two new visualizations and rich interactions that enable tennis coaches and players to obtain insights into match performance. Pie Meter views and Fish Grid views were designed to visualize the proceeding of a game. Rather than a one-to-one sport, basketball is a multi-player sport. Both single-player performance and multi-player collaboration are crucial factors for a team victory.

Regardless of all great sports and analysis visualization, a complete analysis of the proceedings of basketball games is still required.

III. Data Description

Box score data of the entire season, PBP data, game video data, and action area data are the datasets used in this paper. Box score data of the whole season cover all winnings and losings of each team and other statistics, including average scoring, average assists, average turnovers, etc. Game video data contain the videos of games with the video time calibrated to the corresponding game time. Action area data keep track of where the actions take place in the game, and can be used to analyze the attacking styles of the teams.

The format of PBP data is a sequence of event records with timestamps, scores and descriptions of actions (see Figure 1).



1st Quarter				
Time	Chicago	Score		Miami
12:00.0	Start of 1st quarter			
12:00.0	Jump ball: C. Bosh vs. J. Noah (C. Boozer gains possession)			
11:47.0	Turnover by D. Rose (lost ball; steal by M. Chalmers)	0-0		
11:43.0		0-2	+2	U. Haslem makes 2-pt shot at rim (assist by M. Chalmers)
11:17.0	J. Butler misses 2-pt shot from 21 ft	0-2		
11:16.0		0-2		Defensive rebound by L. James
11:10.0		0-2		D. Wade misses 2-pt shot from 21 ft
11:09.0	Defensive rebound by J. Noah	0-2		
10:49.0	Turnover by L. Deng (traveling)	0-2		
10:37.0		0-2		C. Bosh misses 3-pt shot from 26 ft
10:36.0	Defensive rebound by J. Butler	0-2		

Fig. 1. An example of a PBP data.

For each record, we extract the time, the player, the action as well as its attributes. Taking the record in green as an example, the extracted attributes are listed in Table I. Attributes include times of the event, scores for each team, actions, areas of the actions, and players involved in the actions. To be specific, a "made" shot may be 1 point, 2 points, or 3 points. And only the last player who passes the ball to the shooter and directly helps contribute the shot (if ever) is called an "assist" player. If the action involves another action (a "made" action may involve an "assist" action), the attributes of the other action are also extracted. Actions and the number of players involved are listed in Table II.

Time	11:43:0
Team 1 Score	0
Team 2 Score	2
Action Team	Miami
Action Player	U. Haslem
Action Type	Made
Action Attribute 1	2-pt
Action Attribute 2	rim
Action Involvement	M. Chalmers
Involvement Type	Assist

TABLE I
ATTRIBUTES EXTRACTED FROM A PBP EVENT RECORD.

Action Type	Player Involvement
Made (1 or 2 or 3 point(s))	1 or 2 may associate with (Assist)
Turnover	1 or 2, may associate with (Steal)
Miss	1
Rebound	1
Foul	2
Steal	2
Assist	2
Timeout	0

TABLE II
ACTION TYPES AND THE NUMBER OF PLAYERS INVOLVED.

IV. Tasks and Principles

Concerning the particular application of basketball game narration, we propose user tasks that can guide us to design principles. The tasks involve level-of-detail descriptions of an NBA season: the entire season navigation, the game overview and the session level player performance display.

Depict the win-lose situation of the entire season
: The win-lose view provides a good overview of all teams in an entire game season. Users can analyze the strength or weakness of a team from this overview. It can also serve as a game entry through which users select a loss of a strong team or a victory of a weak team or other games of interest.

Identify the change in score difference
: A descriptor of the score difference is the direct indication of the game and is thus the basis of the basketball game analysis. Dramatic changes in scores indicate an intensive period of the game. A great score difference at the end of a game drags the game into trash time. Coaches usually change their tactics when their teams are losing.

Identify how an intensive period or a turning point proceeds
: Users can determine an intensive period or a turning point of the game from a score difference curve. However, users cannot obtain the whole picture of the game without detailed information on player-wise performance or inter-player interactions. In the case of two teams that are good at defending, a score difference curve cannot locate the intensity of a game. Thus, other visualizations are still required to identify the intensity or the turning point of a game.

Evaluate the contribution of a player
: Valuable players are not merely the players who made the most scores. Assists and rebounds are also important contribution metrics. Players who had great performance at the turning points of a game are also considered valuable. A common question to answer at the session level is how to evaluate the role and contribution of a player.

Based on the narration tasks, the design principles are summarized before we start the visualization designs.

- Visualization designs at season, game, and session levels are required.
- Score difference reflects game states and actions reveal game proceedings. To make a comprehensive game narration, both score difference and player actions should be integrated.
- Although it has no explicit data, a good visualization design should reveal intensive periods or turning points of basketball games.

V. Season Overview

As an entry to select a particular game for further analysis, the season overview depicts the win-lose situation at the season level, and meanwhile provides effective indicators to the game of interest. We design two views, namely, the team season view

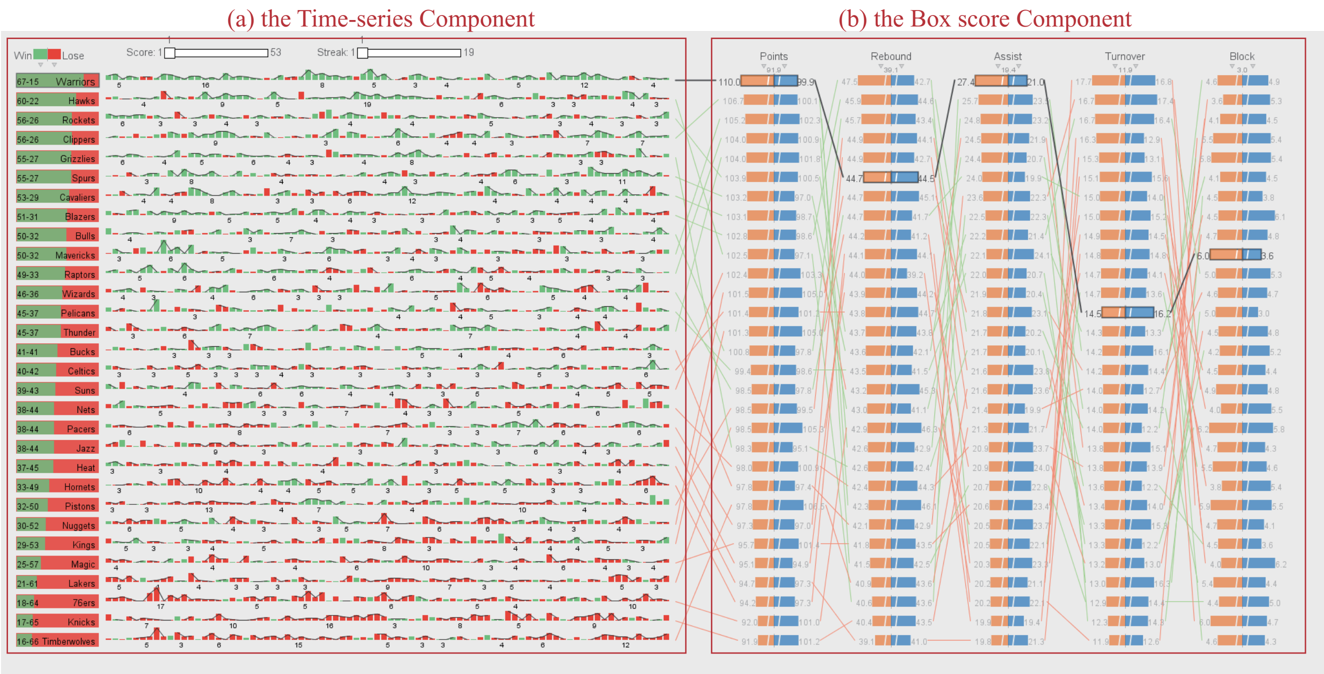


Fig. 2. The team season view. It contains two components: the time series component and the box score component.

and the opponent season view, to provide a season summary from different perspectives.

A. Team Season View

The team season view consists of two components: the time-series component and the box score component, (see Figure 2).

1) Time-series Component: The time-series component encodes each team in a separate line. Within each line, all the games the corresponding team played are shown in a temporal ascending order. This component shows the wins and losses of individual teams throughout the whole season. As shown in Figure 2, the time-series component is decomposed into three levels:

- **Team Game Line:** For each team, all games it played throughout the season are encoded within a line from left to right in a temporal ascending order. At the left end of each line, the name of the team and the number of wins and loses are displayed. The ratio of winnings and losings is encoded in the background rectangle of the team name. The red and green sections represent the percentage of wins and losses, respectively.
- **Single Game:** Each single game is encoded with a rectangle. The color of the rectangle encodes the result of the game: green for win and red for lose. As the final score difference for a game can reflect the gap between both sides, it is encoded by the height of the rectangle.
- **Win-Lose Streak:** A winning streak and a losing streak can reflect the excellent and poor status of a team within a period of time, respectively. In our design, we use a poly-line to connect winning/losing streak rectangles similar to Sparkline[1]. In this way, the poly-line is able to show the trend of score difference within a winning/losing streak. To enable the quantitative analysis of these streaks, we explicitly add the number of each streak and place it beside the corresponding streak.

[1]http://en.wikipedia.org/wiki/Sparkline.

2) Box score Component: The box score component displays the average values of each user-selected attribute for all the teams. It is similar to parallel coordinates and is designed to facilitate the comparison among different teams.

In this component, each column encodes a user-selected attribute. At the top are labels showing the names of the attributes. Below each label are 30 glyphs, one for each team. These glyphs are initially aligned with the teams in the time-series component. Each glyph consists of two parts. The left part encodes the average value of the attribute in that column for the team in that row, and it is calculated from all the games that team played throughout the season. The right part encodes the average value of the same attribute, which is calculated from the same games but for the opponent teams. Values of attributes such as "points" have large bases (usually larger than 90), which make the differences between teams too small to distinguish. To address this problem, we truncate the bars into bases and remainders and explicitly visualize the breaks (as shown in the box score component of Figure 2) to avoid misleading. The base values are labelled on the top of the columns and the breaks are indicated by slash marks. The two parts indicate different abilities of a team. Suppose that the attribute is "Points." The left part shows the average points a team gained, which indicate the attack ability of a team. The right part shows the average points gained by its opponent team, which reflect the defensive ability of the team.

All the attributes are sortable, either by the value of a team or its opponent teams. Thus, we can determine how teams are ranked with respect to different attributes. The adjacent glyphs of the same team are connected by a line, and thus users are able to trace the data of a specific team, which further enable users to compare the different teams.

3) Interactions: We have carefully designed several interactions for this view to help users better view and compare the data.

a) Attribute Selection: Users are free to decide which attribute to show and which one to hide in the box score component. As different compositions of attributes indicate different characters of a team, users are able to compare teams from different aspects in this way.

b) Game and Team Highlight: When users hover over a game rectangle or an attribute glyph, the corresponding team is highlighted as shown in Figure 2. If a game rectangle is hovered, basic information, such as team names, final score and date, of that game will be displayed beside that rectangle. Moreover, the average values of the selected attributes for this specific game are displayed above the corresponding attribute glyph as shown in Figure 2.

If users click the left-most team name, the corresponding team will remain highlighted until a second click. With this interaction, users can highlight multiple teams at the same time.

The highlighting of a game or team makes viewing the information of a single game or team easier for users, and the highlighting of multiple teams at the same time simplifies the comparison among different teams.

c) Filtering: We enable users to filter games in the time-series component in three ways:

- Win/lose filtering can filter out all win games or lose games.
- Score difference filtering can filter out games with a smaller score difference than the user-selected value.
- Streak length filtering can filter out games within the streaks of which the length is smaller than the user-selected streak length.

The filtered out game rectangles and streak lines are shown with high transparency.

d) Sorting: Aside from sorting by attribute as mentioned in section V-A2, the team game lines in the time-series component can also be reordered by the number of either wins or losses. Therefore, users can determine which team plays well and which team plays poorly.

B. Inter-Team View

Different NBA teams have different characteristics, advantages, and disadvantages. The analysis of games played between each two teams may be meaningful and interesting. For example, a weak team may defeat a specific strong team because its tactical characteristics can restrain the advantage of the strong team.

Therefore, we design the inter-team view (see Figure 3) to visualize games played between pairwise teams. We employ the adjacent matrix [20] because of its effectiveness in displaying relational data. To the left of the matrix, 30 teams are encoded and laid out in the same way as in the team season view. We denote these teams as $teamA_i$ ($i \in [1, 30]$). At the top of the matrix, we simply display the names of the 30 teams, which are denoted as $teamB_j$ ($j \in [1, 30]$). Within each matrix cell $cell_{ij}$, we encode the games played by $teamA_i$ and $teamB_j$ with horizontal bars. As two to four games are played between each two teams within a season, each $cell_{ij}$ has two or four bars. The bars are arranged from top to bottom in a temporal ascending order. The colors of the bars are used to encode the win-lose information: the green bar indicates that $TeamA_i$ wins the game, and the red bar indicates that $teamA_i$ loses the game. Thus, $cell_{ij}$ is referred to as the win-lose cell. The length of the bars encodes a user-selected attribute, e.g., "Rebound," "Assist," etc. The background color of $cell_{ij}$ encodes the aggregated outcome: light red indicates that $TeamA_i$ wins more games than $teamB_j$, and light green means that $teamB_j$ wins more games than $teamA_i$. The teams are on a draw when the color is gray. Although the entries in the upper-right triangle of the matrix are indeed the inverse of that in the lower-left triangle, we keep both triangles to maintain consistent visualization of a team's records.

VI. Game Flow View

After investigating the season overviews introduced in Section V, users may find several interesting games and want to further analyze the details of each game. The game flow view is designed to depict the games at the game level and at the session level. In this section, we introduce several design considerations and then describe visual encodings and interactions.

A. Game Session

According to Dean Oliver in his book *Basketball on Paper* [1], "The box score summarizes one game, the unit of basketball time that is most important to coaches and players.". We define the unit of basketball time as the duration between two substitutions and refer to it as a **session**. A session is a confrontation period of five players with five players without substitution. As every player has his position in the team, the tactic is usually maintained in a session. A basketball game is a fast-paced game that leads to a quick transition from defense to offense. We segment a full basketball game with the unit of basketball time to reveal possible game transitions.

B. Visual Encodings and Interactions

As shown in Figure 4, the game flow view is divided into two parts: the upper half that encodes information on the home team and the lower half that encodes information on the away team. Team names and logos are displayed in the left. In the middle of the view, a score difference curve (see Figure 4 (a)) is presented. If the curve is above (below) the middle axis, then the home (away) team is leading the game. If the curve is on the middle axis, the two teams are on a draw at that time point. Score scale is shown at both the left and the right sides of the curve. The score difference curve can be utilized to indicate intensive periods or turning points of the game. Upon mouse hover, time and score difference at the corresponding

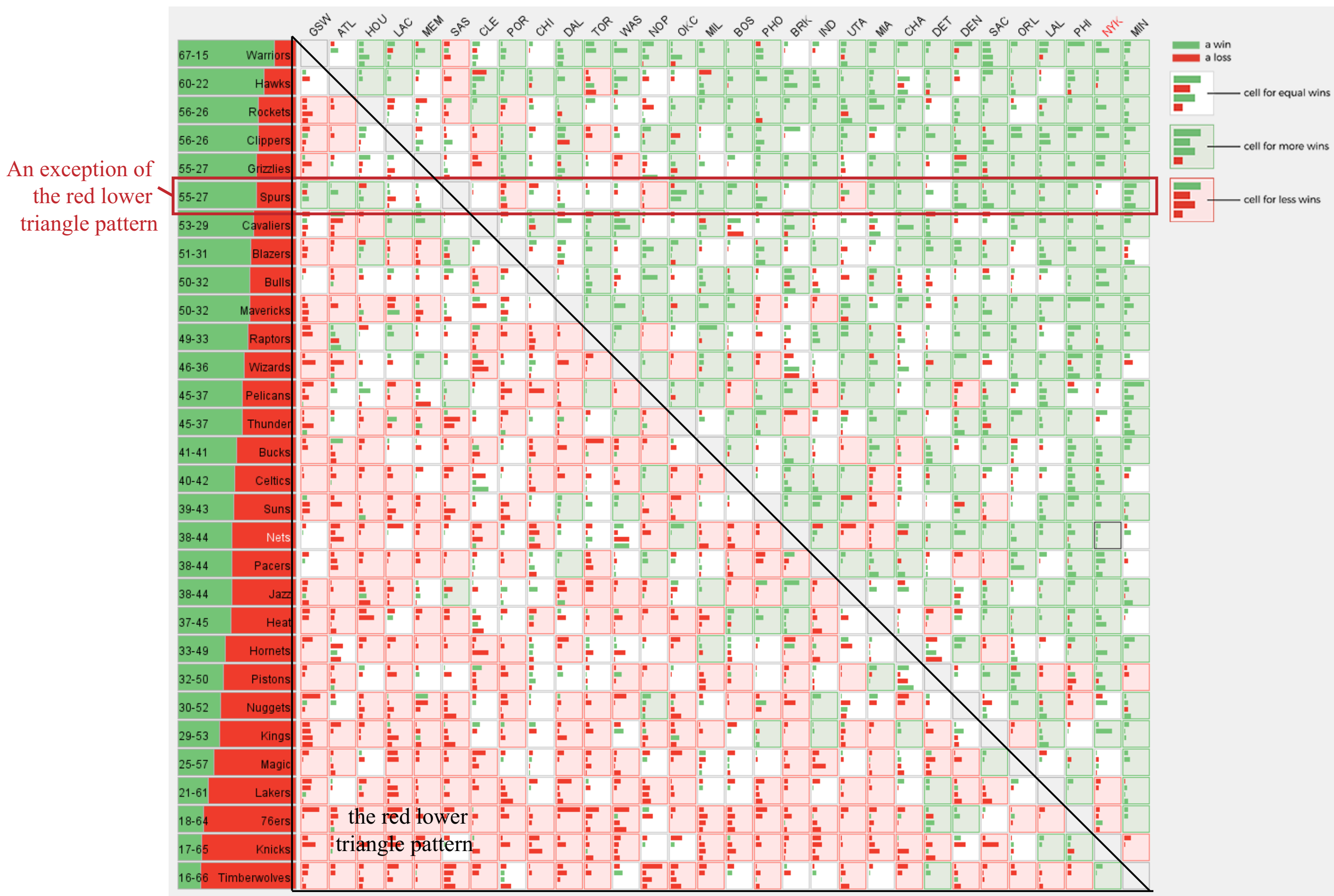


Fig. 3. The inter team view. The red lower triangle pattern indicates that strong teams often had advantages against teams weaker than them. The Spurs is an obvious exception of this pattern: among the five teams who got higher ranks, they won over three teams and made draws with the rest two teams.

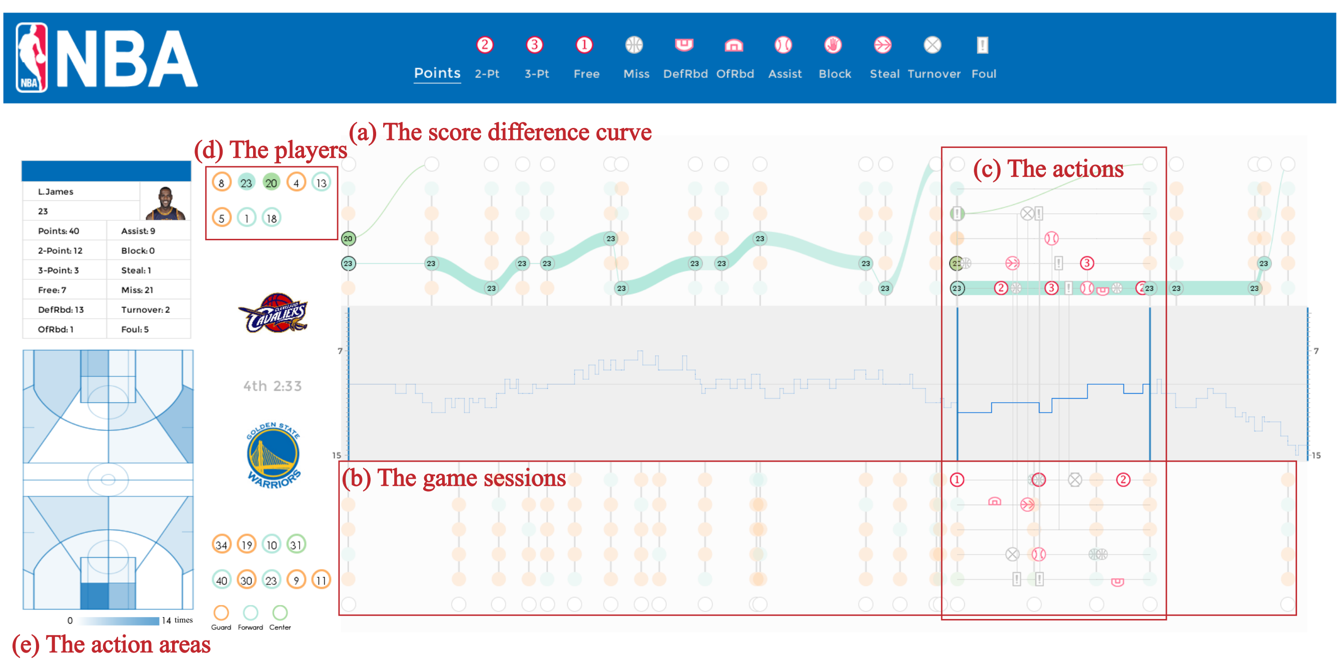


Fig. 4. The game flow view. (a) The score difference curve. (b) The game sessions. (c) The actions. (d) The players. (e) The action areas.

time point is shown to the left of the curve. A"play" icon is shown at the corresponding time axis position, and users can click it to watch the video at that time (Figure 7). The game sessions are displayed above and below the score difference curve (see Figure 4 (b)). Here, we take home team as an example. The game sessions of the away team are encoded and laid out in the same way. We use circles to encode players. Each session consists of six circles in a column: five connected circles denote the court players and the remaining ones denote the off-court players. The fill color of each circles encode the position of the corresponding player. Specifically, the off-court players' circle is filled with white. The color of the circle border encodes the substitution of a player. The color is black if a player is substituted into the court; otherwise, the color of the circle border is gray. When users hover over any on-court player's circle, all the circles representing the player are highlighted. Moreover, the adjacent ones are connected by a band, the transparency of which is calculated on the basis of the scoring rate (i.e., the points gained per second) within the time period between the two sessions. The lower the scoring rate is, the more transparent the band, and vice versa.

If users click between two adjacent session columns, all the actions taking place in that time period will be displayed with action icons. If the space between two adjacent sessions is too narrow, users can keep clicking to enlarge the space, so that all the actions can be displayed without visual clutter. The legend of the actions is shown at the top of the view. By clicking a specific action icon in the legend, the corresponding action icons will show or hide in the view. The actions are also indicators of intensive periods of the game. By integrating the players' performance bands and their actions, users are able to evaluate the contribution of the players.

Our data contain position information for each action, as shown in Figure 4 (e). As the whole court is divided into 28 regions, and we can determine which region a specific action takes place. At the beginning, all the regions are colored according to the numbers of goals made within them. When users hover on an action icon, the corresponding region is highlighted. To collapse the actions within the session, users need to right-click between the two adjacent session columns.

A list of all the players (see Figure 4 (d)) of the home team who played in the game is located on the left side of the session panel. The list is filled with white background and the jersey number of the corresponding player. The border color encodes the position of the player. When clicked, the circles are highlighted by filling with their border color. The corresponding player is highlighted in the session panel, and the information of that player is displayed on the top left corner. This interaction enables users to compare among different players in the game.

VII. Case Study

A. 2014-2015 NBA Season

We collect the standings and other box score statistics of all teams in the 2014-2015 regular season.

1) Team season view: We begin the analysis from the team season view. By ordering the teams by their winning games

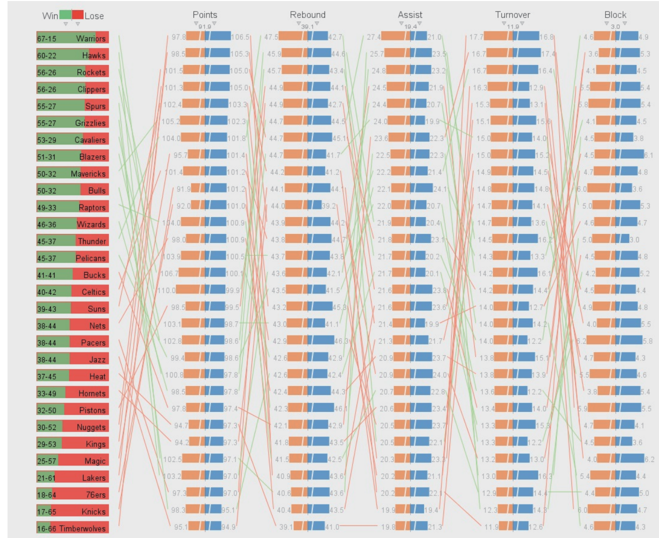


Fig. 5. Sorting teams by points made by the opponents. A clear negative correlation pattern between the strength of teams and points made their opponents is revealed.

(Figure 2 (a)), we distinguish the strong teams, such as the Warriors, from the weak teams, such as the Timberwolves. Strong teams usually have large score differences and long winning streaks. Taking the Warriors as an example, the teams' longest winning streak is 16. Thus, we wonder whether good offense or good defense makes a team strong.

An initial assumption is made that strong teams must make more points and have better collaboration among players. On the basis of this assumption, we sort the teams by points and assists in the slope graph of box scores (Figure 2 (b)). With strong teams indicated by green lines and weak teams by red lines, we find that strong teams (in green lines) usually have higher ranks in both points and assists. However, green lines of a great downwards slope and red lines of a great upwards slope indicate exceptional cases such as the Grizzlies and the Nuggets, respectively. The Grizzlies is a strong team that made relatively less points and less assists. Conversely, the Nuggets are a weak team that made passable number of points and assists. We sort the teams by other metrics and then realize that it is the points made by the opponents, rather than the points made by the team, that determines the strength of a team. Figure 5 shows the teams sorted by points made by their opponents. A clear negative correlation pattern between the strength of teams and points made by their opponents is revealed, as most green lines have a downwards slope. The stronger the team is, the sharper the slope, and vice versa. Strong teams tend to lose less points than their opponents. The pattern of assists is the opposite, but it indicates a similar situation that stronger teams tend to have better assists.

From this observation, we make another assumption that defense and teamwork are better indicators of the strength of a team than offense. Nevertheless, exceptions can be found based on the new assumption: the Thunders and the Mavericks. The Thunders ranked fourth in average points, but they performed poorly in assists and had many turnovers. The three technical statistics reflect the style that the team had

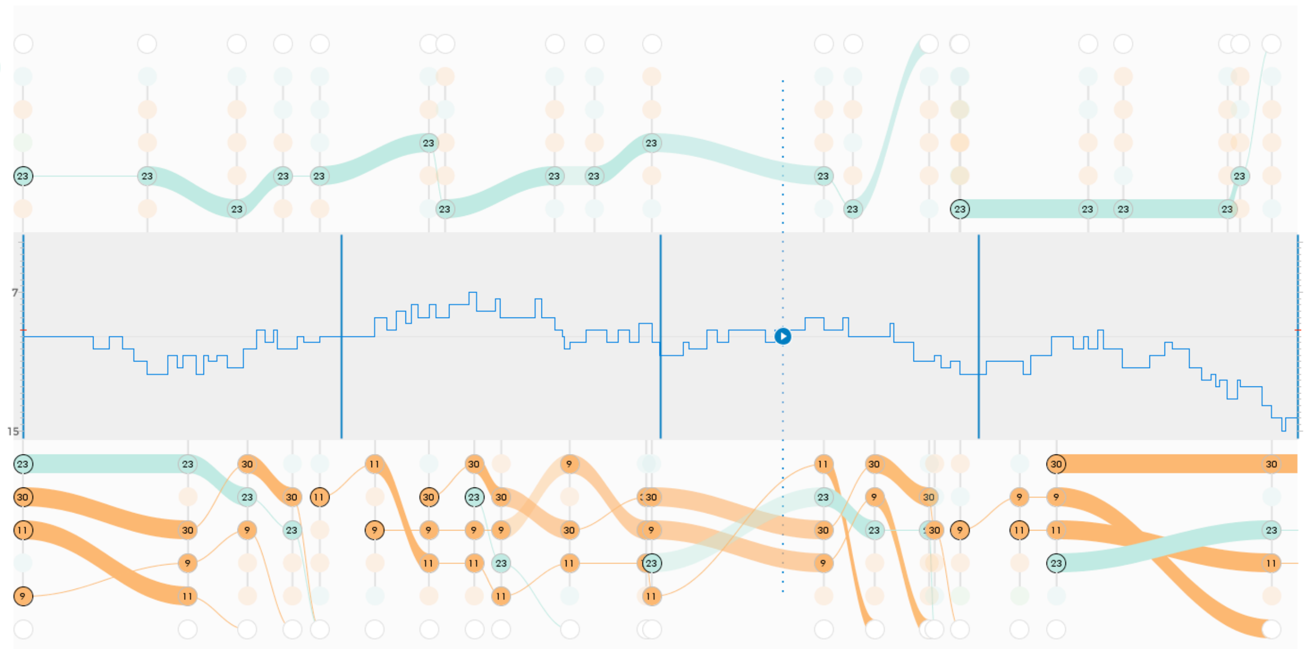


Fig. 6. Player #23 is the only leading scorer in the Cavaliers while in the Warriors there are several leading scorers such as #30, #11 and #9.

poor teamwork (indicated by low assists and high turnovers) but had one or more players of strong personal abilities. In fact, the Thunders had great players such as Kevin Durant and Russel Westbrook. In terms of the Mavericks, the team had low turnovers and low rebounds. The former reflects their strength, while the latter indicates that it contributed to the high points made by its opponents.

2) Inter-team view: We can also obtain some observations in the inter-team view (see Figure 3). First, the entries in the lower triangle are mostly in red, while the entries in the upper triangle are mostly in green. The red section in the lower triangle pattern indicates that strong teams usually have advantages over teams weaker than themselves. The Spurs is an obvious exception from this pattern: among the five teams that obtained higher ranks, this team won over three teams and made draws with the remaining two teams. It is because the Spurs lost a few games to those who ranked behind them. To be specific, the Spur lost 3 games out of 4 to the Pelicans, a team ranked medium in the NBA league. An assumption can be made that the Pelicans might have something that restricts the Spurs' performance.

B. 2015 NBA Finals Game 5

In 2015 NBA Finals Game 5, the Golden States Warriors won over the Cleveland Cavaliers at home at 104-91. The starting lineup of the Warriors was three guard players and two forward players, and the starting lineup of the Cavaliers is two guard players, two forward players and one center player. Based on the configuration, both teams used a small lineup. According to the score difference curve, the game was intensive in the first three periods, and the scores increased alternatively. The Warriors conquered the game in the fourth period and remained ahead of the Cavaliers until the end of the game.

By ranking players by points, we can locate the shooters in the two teams (see Figure 6). Among the Cavaliers, player #23 played almost every minute of the game and had an outstanding performance (see Figure 4 (c)). He obtained high ranks in points, defensive rebounds, and assists. The Cavaliers was able to catch up with the Warriors in the first three periods because of his performance. However, he missed several baskets in the fourth period probably because of the great physical consumption. Short of his contribution, the Cavaliers had no other leading scorer and could not reverse the situation. By contrast, the three guard players (#30, #11 and #9) of the Warriors were all leading scorers and contributed great performances in the fourth period. Clearly, defending three scorers is more difficult than defending only one.

The only starting center player (player #20 Timofey Mozgov) played only a couple of minutes in the beginning. He was substituted onto the court at the end of the third period. According to the actions and video review, the Cavaliers could not protect their defensive rebounds at that time. He was probably substituted to improve the team's rebound and gain control of the situation. However, he only contributed fouls and turnovers (see Figure 4 (c)). Thus, he was substituted again.

The finals MVP is player #9 Andre Iguodala of the Warriors. He made 14 points, 8 rebounds, 7 assists and a game-high 3 steals in Game 5. Three minutes after the game started, the Warriors led the game for the first time at 6-0, and player #9 made a significant contribution. Although he did not make any shots, he contributed two assists out of the three shots, two rebounds, and one steal. Especially at 3:49 of the first period, he launched the offense of the team. As shown in Figure 7 session A, he made a defensive rebound and passed

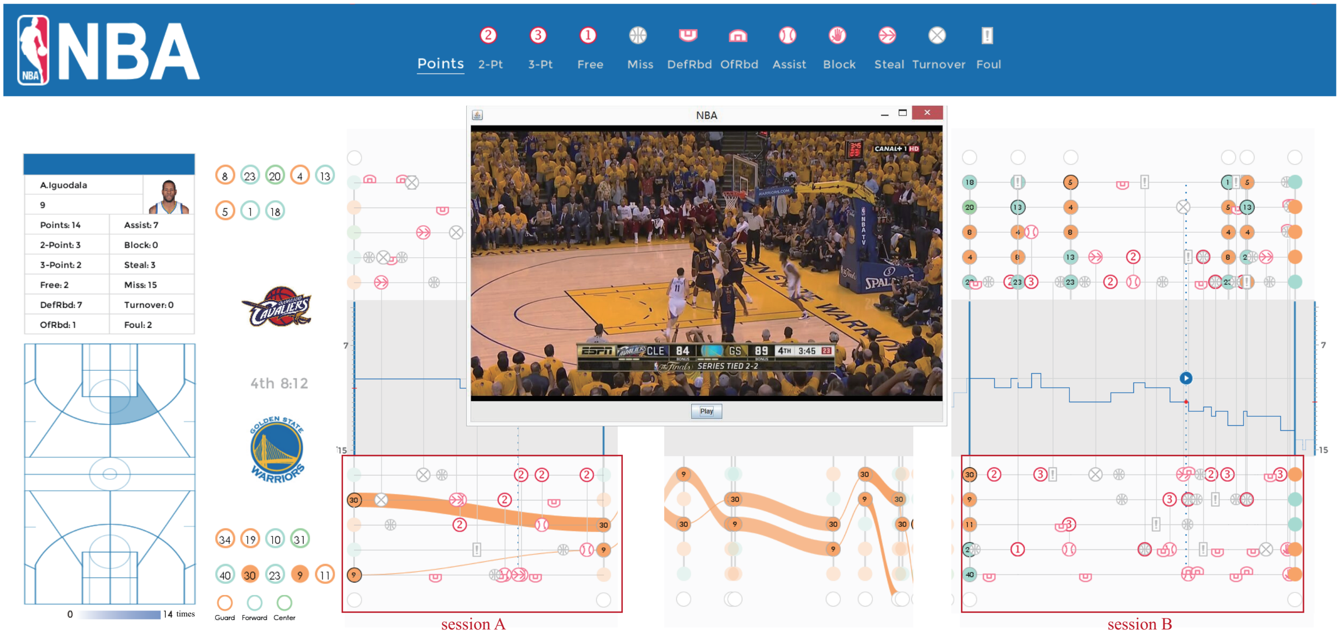


Fig. 7. Player #9 contributed to this game by being the key man who switched from defense to offense. We can infer his decisive role in the team at 3:49 of the first period (session A) and at 7:45 of the fourth period (session B).

the ball to #30, who made the shot. In the next round, he stole the ball from the Cavaliers and made another assist to the Warriors. In the round after that, he made another defensive rebound and helped the team make its third shot. We can infer his decisive role in the team at 7:45 of the fourth period, when the Warriors remained ahead and extended their lead. He first made a three-point shot and then a three-point play (see Figure 7 session B). Although he missed the free throw, the five points he made helped the team to lead the game again. Later, his teammates extended the lead and sealed their victory. Thus, we can conclude that player #9 contributed to this game by being the key man who switched from defense to offense.

VIII. User Feedback

Two basketball devotees were invited to use our system and give feedbacks over the design. They were satisfied with the season view, especially with the comprehensive comparisons in multiple aspects of all teams. They thought the box score component reveals the offense-defense characteristics of the teams: some teams are better at shooting but their weak defense restrict their paces towards the finals. Regarding the game view, they liked the video function and thought the session by substitution design is really crucial to analyze the tactic change of a team. However, they also pointed out that the game view is lack of supporting descriptions over the situation. They suggest textual narrations from social network or social media to complement the visualization. They also would like to know the exact shooting locations for each score.

IX. Conclusion

In this paper, we integrate multi-media heterogeneous NBA game datasets, such as PBP text data, action area data, basic statistics, and game video data of basketball games, to enable a comprehensive visual narration of a complete basketball season. On the basis of the datasets, we propose a full representation of the whole basketball season, which reveals games in multiple levels of details: the season level, the game level and the session level. In future works, we plan to integrate player location track data to enable an in-depth analysis of coaches' tactics.

Acknowledgment

This work is supported by National 973 Program of China (2015CB352503), Major Program of National Natural Science Foundation of China (61232012), National Natural Science Foundation of China (61422211), National Natural Science Foundation of China (61303134), Zhejiang Provincial Natural Science Foundation of China (LR13F020001), the Fundamental Research Funds for the Central Universities.

References

[1] D. Oliver, *Basketball on paper: rules and tools for performance analysis*. Potomac Books, Inc., 2004.

[2] P. Kvam and J. S. Sokol, "A logistic regression/markov chain model for ncaa basketball," *Naval Research Logistics (NRL)*, vol. 53, no. 8, pp. 788 – 803, 2006.

[3] J. Igor, T. Slavko, and P. Ante, "Analysis of basketball game states and transition probabilities using the markov chains," *Physical Culture - Journal of Sport Sciences & Physical Education*, vol. 66, no. 1, pp. 15 – 24, 2012.

[4] K. Goldsberry, "Courtvision: New visual and spatial analytics for the nba," in *MIT Sloan Sports Analytics Conference*, 2012.

[5] R. Sisneros and M. Van Moer, "Expanding plus-minus for visual and statistical analysis of nba box-score data," in *SportVIS - Workshop on Sports Data Visualization*. IEEE, 2013.

[6] M. Bashuk, "Using cumulative win probabilities to predict ncaa basketball performance," in *MIT Sloan Sports Analytics Conference*, 2012.

[7] G. R. Hamilton and C. Reinschmidt., "Optimal trajectory for the basketball free throw," *Journal of sports sciences*, vol. 15, no. 5, pp. 491–504, 1997.

[8] Matej Perše and Matej Kristan and Stanislav Kovačič and Goran Vučkovič and Janez Perš, "A trajectory-based analysis of coordinated team activity in a basketball game," *Computer Vision and Image Understanding*, vol. 113, no. 5, pp. 612 – 621, 2009.

[9] "Fathon's visualization over basketball data," http://fathom.info/latest/6985, accessed: 2014-03-22.

[10] "New york times's NBA shot analysis," http://www.nytimes.com/interactive/2012/06/11/sports/basketball/nba-shot-analysis.html, accessed: 2012-06-12.

[11] "Hotshotchart," http://www.hotshotcharts.com/, accessed: 2014-03-22.

[12] R. Cava and C. D. S. Freitas, "Glyphs in matrix representation of graphs for displaying soccer games results," in *SportVIS - Workshop on Sports Data Visualization*. IEEE, 2013.

[13] K. Wongsuphasawat, "A narrative display for sports tournament recap," in *SportVIS - Workshop on Sports Data Visualization*. IEEE, 2013.

[14] C. Perin, R. Vuillemot, and J.-D. Fekete, "Soccerstories: A kick-off for visual soccer analysis," *IEEE Transactions on Visualization and Computer Graphics*, vol. 19, no. 12, pp. 2506–2515, 2013.

[15] H. Janetzko, D. Sacha, M. Stein, T. Schreck, D. Keim, O. Deussen *et al.*, "Feature-driven visual analytics of soccer data," in *IEEE Symposium on Visual Analytics Science and Technology*. IEEE, 2014, pp. 13–22.

[16] M. L. Parry, P. A. Legg, D. H. S. Chung, I. W. Griffiths, and M. Chen, "Hierarchical event selection for video storyboards with a case study on snooker video visualization," *IEEE Transactions on Visualization and Computer Graphics*, vol. 17, no. 12, pp. 1747 – 1756, 2011.

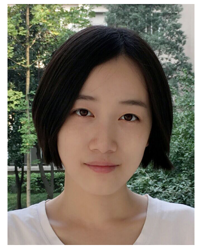
[17] H. Pileggi, C. D. Stolper, J. M. Boyle, and J. T. Stasko, "Snapshot: Visualization to propel ice hockey analytics," *IEEE Transactions on Visualization and Computer Graphics*, vol. 18, no. 12, pp. 2819–2828, 2012.

[18] C. Dietrich, D. Koop, H. T. Vo, and C. T. Silva, "Baseball4d: A tool for baseball game reconstruction & visualization," in *IEEE Symposium on Visual Analytics Science and Technology*. IEEE, 2014, pp. 23–32.

[19] T. Polk, J. Yang, Y. Hu, and Y. Zhao, "Tennivis: Visualization for tennis match analysis," *IEEE Transactions on Visualization and Computer Graphics*, vol. 20, no. 12, p. 2339, 2014.

[20] D. Holten, "Hierarchical edge bundles: Visualization of adjacency relations in hierarchical data," *IEEE Transactions on Visualization and Computer Graphics*, vol. 12, no. 5, pp. 741–748, 2006.

Wei Chen is a professor in State Key Lab of CAD & CG at Zhejiang University, P. R. China. He has published more than 60 papers in international journal and conferences. He served as Steering committee of IEEE Pacific Visualization, Conference chair of IEEE Pacific Visualization 2015, Paper Co-chair of IEEE Pacific Visualization 2013. For more information, please refer to http://www.cad.zju.edu.cn/home/chenwei/. Contact him at chenwei@cad.zju.edu.cn.



Tianyi Lao is a master student at the State Key Laboratory of CAD & CG, Zhejiang University. Her current research focuses are information visualization and visual analytics, especially visual analytics of sports data. Contact her at laotianyi@zju.edu.cn.

Jing Xia a Ph. D. student at the State Key Lab of CAD & CG, Zhejiang University. Her research interests include Information Visualization and Visual Analytics, especially visual analytics of temporal data and spatial-temporal data. Contact her at xiajing_cad@zju.edu.cn.

Xinxin Huang received her master degree at the State Key Laboratory of CAD & CG, Zhejiang University. Her research focuses are information visualization and visual analytics, especially visual analytics of sports data. Contact her at huangxinxin07@gmail.com.

Biao Zhu is a Ph. D. student at the State Key Laboratory of CAD & CG, Zhejiang University. His current research focuses are information visualization and visual analytics. Contact him at zhubiao@zju.edu.cn.

Wanqi Hu a master student at the State Key Laboratory of CAD & CG, Zhejiang University. His current research focuses is information visualization. Contact him at huwanqi@zju.edu.cn.

Huihua Guan is a graduate student at the State Key Laboratory of CAD & CG, Zhejiang University. Her current research focuses are visual analytics, information visualization and human-computer interaction. Contact her at ghh@zju.edu.cn.