Sangati – A Social Event Web approach to Index Videos

Anjana Anilkumar, Anusha Sreenivasan, Animesh Sahay, Dilip Gurumurthy, Nirupama M P, Subramaniam Kalambur, Dinkar Sitaram
Computer Science and Engineering
PES University,Bangalore
{subramaniamkv,dinkars}@pes.edu

Ramesh Jain
Department of Computer Science
University of California, Irvine
jain@ics.uci.edu

***Abstract*— In its inception, the World Wide Web was a means of sharing text documents. As the Web has grown, the amount of non-text content (primarily image and video) has grown. The reason for the growth of image and video content is the desire of users to share their experiences. Video in particular has a very rich number of associations. For example, a goal scored at the last minute in a football game may have an association of "turning point of match" and should logically be hyperlinked to other similar events, as these associations are part of the experience the user would like to share and may be the reason for the user to take the video at that particular point. The amount of image and video content on the Web far exceeds the text content, and there are applications, such as Facebook, which provide support for image and video sharing. However, hyperlinking of video content does not still capture the full range of hyperlinks latent in the video making it difficult for users to share such experiences. We model a video as a social event web, which is a complex graph with hyperlinks between events. The event web can be linked to the WWW in a natural way. To discover hidden hyperlinks, we perform data mining of the event web using context from the event web and domain knowledge. Experiments on sports videos show the effectiveness of our framework and prove that our approach can provide richer hyperlinks between video segments, thereby extending the reach of the WWW to include the hidden content within videos and enabling users to share such links.**

Keywords— Event web, Situation Recognition, Video Indexing, Context Analysis, Ad-hoc links, Hyperlinks, Sentiment Analysis.

I. INTRODUCTION

Since their invention, computers have been used for many purposes. One major use is the sharing of information, which in the early days of computing was via email. A major advancement was the creation of the World Wide Web, which made possible the sharing of much richer information via hyperlinked images, videos, and text. The past few years have seen the emergence of companies like WhatsApp, Facebook, Flickr, Instagram, Snapchat, and GoPro because these companies provide platforms through which users can share their experience using multimedia [8].

Photos present objects, scenery and people instantaneously; to represent the same in text requires numerous words, which often fail to do justice to the image and the relationships within it. Photo content is significantly more subjective than text; people see what they want to see in pictures. A person's view of a picture could vary over time depending on changes in real-world situations or events[6]. Users capture videos to record significant events in their lives as they are interested in the context offered by it. An event can be thought of as a sequence of sub-events. For example, a graduation ceremony is comprised of sub-events such as the commencement address, conferring of degrees and alumni greetings to name a few. These events form a part of the sequence in the life of every person who participated in them.

Presently, the simple links derived from image and video tags do not capture the full meaning of the event making it difficult to share deeper associations. As an example, consider a video clip showing part of a football match, where a player X scores a goal under extreme pressure to level the match. For the user who took the video clip, the meaning of the clip is likely to include associations such as turning point of the match, one of X's best attempts and a perfect goal. These subtle associations are part of the experience of the match which the user would like to share. Our approach allows such links to be automatically discovered, simplifying sharing of such links. Additionally, data mining of video clips to find all those which fall into the category "one of X's best shots" could give an opposing player or coach the useful information for counter-strategies. Thus, images and videos can be thought of as a web of events.

These hidden associations are not generally part of the links available on common social media platforms. It is therefore necessary to develop techniques to analyze an image or video more deeply and fully capture the context and associations[6] present. This would help users to share richer and more meaningful experiences as well as allow mining of the contexts and associations to extract useful information. In this paper, we show that modelling multimedia documents[15] as part of a social event web can greatly help in situation recognition and in discovering the latent associations between events. We have extended the event web model[18] and have demonstrated the applicability using social event web models for two different sports - cricket and tennis. The event-web consist of events and hyperlinks between them that were then used to instantiate an event web for each video document. Navigation across this web is performed by traversing the hyperlinks across the events. We have determined the type of associations each user expects to find in a video and would like to share. We then show how our event web can be used to efficiently discover these associations. Due to the similarity of the event models, we show that similar techniques can be used to discover these associations, thus resulting in design reuse and obviating the need to completely customize the queries for each sport.

II. Related Work

The objective of video analysis research has been to search and identify events of interest in a video or across a group of videos. An exhaustive survey of the research in the area of video analysis has been carried out by Hu et. al[5]. The identification of key events within the video is based on tracking objects through different frames of the video for use in video surveillance[9, 10]. In the case of traffic videos with fixed cameras, the detection of objects and tracking of their motion to check for violations constitute events. However, the events are only spatio-temporal in nature and no attempt is made to identify relationships across events which is a requirement for a visual web. A similar approach to cluster identical frames[2].

In contrast, sports videos are more dynamic in nature and Li et. al[12] have generically modelled sports videos in terms of “plays” and “non-plays”. Their model is generic enough to be used for a static summarization of the video, but does not lend itself to dynamic summarization that requires semantic knowledge of the events.

Our work is inspired by the research challenges of trying to identify context in events in a visual web[8], to perform dynamic summarization. This requires the need to understand event semantics which can be inferred from videos[16], but is limited to determining object motion. To understand if the player scored a winner, integration with a domain/structural model[3] and multiple sources of information may have to be integrated [1]. We integrated information from various sources such as video commentary, inline text in videos and data from other websites having match descriptions to arrive at our event model. This is similar to the approach taken by Evangelopolous[4] to use speech, object detection and text events to identify salient points in the image. However, they use this for static summarization. Our approach to take into account the context of the game, goes beyond integration of highlights or preplays between segments of plays to describe the context of a particular event.

The approach taken by Wang et. al[17] attempts to identify temporal sections of videos with user supplied tags to identify the semantics of an event. We instead rely on the domain model and other sources of information about the event itself (such as a match commentary) to identify an event.

III. Overview of approach

An image or video clip shared by a user is a representation of an experience. Reconstruction of the full context and associations of an experience is referred to as situation recognition [7]. Crucial to the creation of an event web is the discovery of links between events, a model which defines those that can occur, their attributes, and their structure (e.g., a particular event is a subevent of another) as well as their associations. The event model has a direct relationship to the efficiency of search and navigation.

We define an event(or a sub-event) as a n-tuple <*S, T, Sp, I, C, E*> as defined by Westerman and Jain[18]. The various components of the event are

- Structural(S) – this represents the structure of the domain that is being represented by the video. For each domain (specific sport or classroom), we need to define a domain model. This is illustrated further in the definition of a domain model for each sport.
- Temporal (T) – represents the time period represented by the event or sub-event.
- Spatial (*Sp)* – represents spatial coordinates or geolocations associated with sections of the video.
- Informational *(I)* – includes people who were involved in the event and their attributes, roles.
- Causality (C) – represents how some events are caused by other events.
- Experiential *(E)* – is the experience captured by a section of the video. Interestingly, a same segment of video can have multiple experiences depending on the perspective of the persons watching the segment of video. For a sport, a goal scored can be exciting for fans of one team while reflecting disappointment in the other teams. Our approach focuses on automatic discovery of experiential links.

To represent the experience in its totality, we represent each video as being constructed of events. An event is further decomposed into sub-events that may be atomic or compound. Atomic events are indivisible plays. Compound events are a combination of multiple atomic events or compound events. For example, in our model, a ‘game’ in Tennis is considered as an atomic event, while a ‘set’ is considered as a compound event. Each event is linked to other events in the same video, or events in other videos, creating an event web[6]. Thus navigation and search are possible by querying on the links to get the related events which make up the entire experience. The *temporal, spatial* and *informational* aspects of an event can be gleaned part of the atomic events and are “truths” while *causality* and *experiential* aspects are derived aspects and depend on the perspectives and biases of the person experiencing the event. Hence, events may have either no *experiential* or *causal* information associated with them or may have more than one *experience/causal* information associated. A *social event-web* is one with multiple *experience* aspects which can be associated with one or more events and it often depends on the perspectives of the people observing the *events.* For example, a player fan may select a certain subset of events as being significant from their perspective.

In the following section, we first present details of our event model. To illustrate the generality of our approach in the field of sports, we have applied it to two different sports - cricket and tennis and have constructed similar event models. We then describe using our event model, an event web representation for cricket and tennis videos

IV. EVENT MODEL FOR CRICKET

A. *Cricket Background*

Cricket is a bat-and-ball game, similar to baseball, played between two teams of eleven players on a cricket field. At the centre is a rectangular 22-yard-long pitch with a wicket i.e., a set of three wooden stumps sited at each end[19]. Each match consists of two innings. Each innings consists of a certain number of overs and each over is made up of a minimum of six balls. Each ball is bowled by the bowler of one team (called the fielding team) to the batsman of the batting team. On each ball, the batsman may score a certain number of runs (including 0) or get out. The batting team attempts to score as many runs as possible, whilst the fielding team tries to prevent runs from being scored by retrieving the ball after it has been hit by the batsman. After either ten batsmen have been dismissed or a set number of overs have been completed, the innings ends and the two teams swap roles. The winning team is the one that scores the most runs during the match.

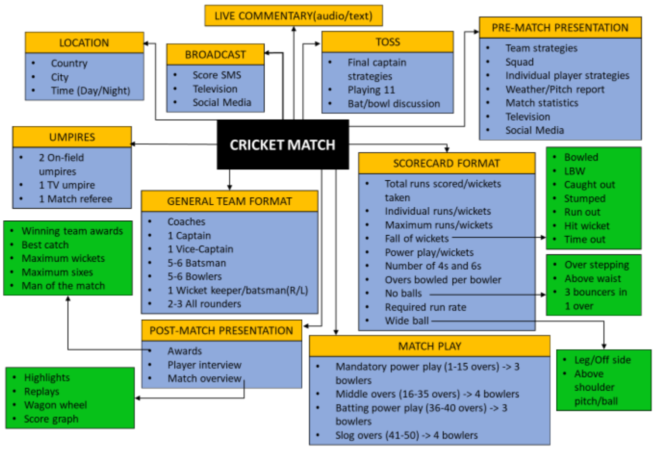

Figure 1: Event Model for Cricket

Figure 1 illustrates our event model for cricket. This model consists of two parts. First, we have the decomposition of a cricketing video into a number of events and sub-events that are ordered linearly in time and represent the temporal component of the event. This includes, for example, the pre-match presentation event (see figure 1), which contains sub-events such as the discussion by the captains of the teams about their strategies, as well as commentary by the commentators on the likely course of the match. It is to be noted that the event model describes possible events, and not the actual events present in any one video stream. For example, in a video stream that does not include a pre-match commentary, there would be no instance of this event. The second part consists of the participants in the match, such as the players on each team. Strictly speaking, these are not part of the event model per se but are required to describe the structural aspect of the model. We describe these in more detail below.

B. *Details of cricket event model*

Match Play The match play event is a compound event that contains the basic events of the match and is used to generate the *Structural* (S) aspect of the model. A match is divided into *overs* and these are logically separated out into 4 sets -- the mandatory power play, the middle overs, batting power play, and the slog overs. Each *over* itself is subdivided into a number of *balls*, each of which is an atomic event in our model. The basic properties of a ball in cricket includes a timestamp and also a location. It has the following attributes: the *players involved* which include the batsman who is batting, the bowler, the number of runs scored, and any additional information (for example, if the batsman gets out, the mode of getting out) and constitutes the *Informational (I)* aspect of the model. The match play event thus encodes part of the semantics of the game. Additionally, given a ball or an over, it is easy to navigate through the model to find out which part of the match it belongs to. The part of the match the ball or over belongs to is contextual information which can be efficiently retrieved for data mining and discovery of links. The *Spatial(Sp)* aspect of the model is defined by the location of the match.

Pre-match and Post match analysis This is a compound event made up of a number of sub-events. Important ones include interviews with the captains of the two teams about team strategy to be followed and the individual player strategy, which is a discussion about the expectations from the key players. Note that by comparing the individual player strategy or the team strategy with the actual course of play, it is possible to discover if the two teams were able to execute their strategy, or whether the individual players played according to the role they were assigned in the strategy. This constitutes the *Experiential (E)* aspect of the model and can be viewed from the perspectives of both the teams.

Toss This is an important compound event, as the question of which team bats first is determined by a coin toss. The aspect of *causality(C)* in the model is gleaned from the toss. While there may exist causality amongst other sub-events, that is a more complex problem and is subject to perception. We restrict ourselves to this simplistic approach to causality.

.

V. EVENT MODEL FOR TENNIS

A. *Tennis Background*

Tennis is a racket sport that can be played individually against a single opponent (singles) or between two teams of two players each (doubles)[22]. The game of tennis is played on a rectangular court with a net running across the centre. The aim of the game is to hit the ball over the net landing the ball within the margins of the court and in a way that results in the opponent being unable to return the ball. The players start on opposite sides of the net. One player is designated as the server, and the opposing player is the receiver. A legal service starts a rally, in which the players alternate hitting the ball across the net. A player wins a point every time that the opponent is unable to return the ball or makes a fault. A fault occurs when the ball is hit into the net or outside the court. A game consists of a sequence of points played with the same player serving. A game is won by the first player to have won at least four points in total and at least two more than the opponent. A set consists of a sequence of games played with service alternating between games, ending when the count of games won meets certain criteria.

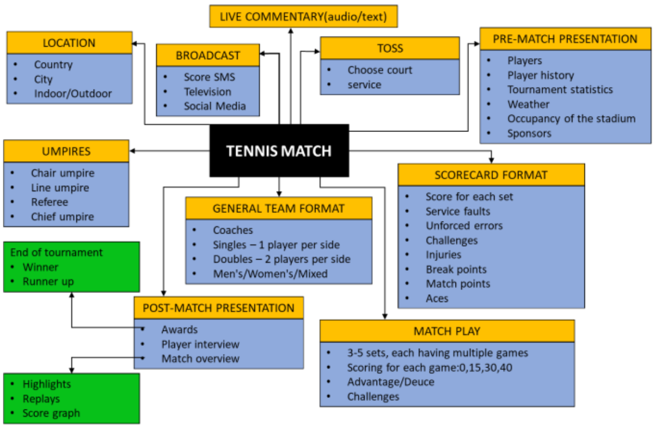

Figure 2: Event Model for Tennis

Our event model for tennis is illustrated in Figure 2. It can be seen that the model has similarities to the cricket model, containing common elements such as the pre-match presentation event. We present details of the model below.

B. Details of Tennis event model

Our tennis event model is similar to our cricket event model. **Match Play:** As in the cricket event model, this is a compound event that contains the actual events of the match and constitutes the *Structural(S)* aspect; this consists of a number of set events. Each set consists of a number of game events. At present, we have found it sufficient in our tennis model to treat games and tie-breaks as atomic events, and not go to the point level. A game is of higher significance than each point, and provides sufficient context; that is, players generally treat each game as a unit and devise strategies for each game. In addition to the basic attributes, *informational* attributes*(I)* of the game involve the side of the court that the player plays on, the server, receiver and referee of the game. Informational attributes of the match include the participating players and the team or country they represent. As in the cricket match play event, the tennis match play event encodes the structure of the match and allows for easy inference of context in the match.

Toss In tennis, the toss determines which player serves first. The winner of the toss can also choose which side of the court to serve from. This is represented by the toss event and highlights the *causality(C)* aspect of the model.

Pre-match presentation, Post-match analysis

These events are similar to the corresponding ones in the cricket model, and contain events such as the match highlights event, player strategy event, and represent the *Experiential (E) aspect* of the model.

VI. SOCIAL EVENT WEB CREATION

We now describe the architecture of *Sangati* and the process of how an event web was created for each video based upon the event models described in the previous section. With the drastic improvement and availability of multimedia technology[14], the volume of visual data getting generated is increasing exponentially. The social event web in *Sangati*, organizes data in the form of events with experiences and allows natural access from the user's perspective. Users participate in a sporting or entertainment event only through observation from a particular perspective. In addition, they could archive these events to experience later, probably from a different perspective each time. This could provide insights leading to valuable knowledge creation.

Figure 3 illustrates the architecture of *Sangati* and consist of the following modules:

1. *Event Recognizer:* This uses the event model of each sport and extracts atomic events from both the Videos and Text commentaries and descriptions of the videos. For example, in tennis, a game is considered as an atomic event. It builds the inherent structural relationships to create compound events from the atomic events. For example, in the case of cricket, multiple balls bowled by a bowler may be combined into overs. The event recognizer generates the *S*, *Sp*, *I*, *T* and *C* aspects of the event model.

2. *Association Builder:* This uses other sources such as text commentary and match analysis to extract qualitative attributes to be associated with each simple or compound event. For example, from the commentary, using sentiment analysis and a rule based system, we classify each ball in cricket as being good for the batsman or good for the bowler. This information is used later for querying. A challenge in merging events from both the video event stream and the commentary is synchronizing events between two different event streams.

These steps are described in more detail in the following section.

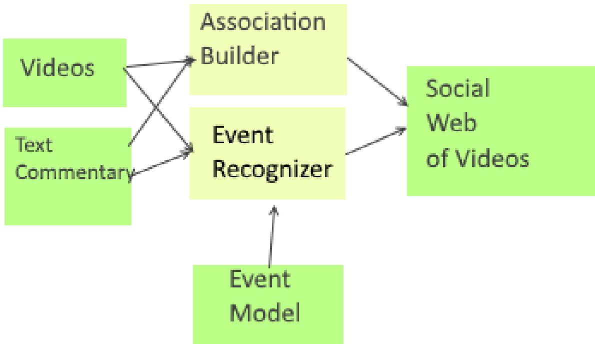

Figure 3: Architecture of Sangati

A. *Detection of Atomic Events and Inherent Structure*

We consider the video to be made up of two parallel streams - an image stream and a text stream (from the audio). These are separately segmented, aligned, and converted into atomic events. The atomic events are then further combined into compound events as described earlier.

Ball by ball commentary of a particular match was obtained from ESPNCricInfo which is a sports news website exclusively for the game of cricket[20]. Similarly, for tennis, the website scoreboard.com was used.

We synchronized the events between the video and the text commentary by identifying text scores from live video commentary[13] and then matching these with events from text commentary stream.

To help viewers keep track of a match, scores are usually displayed on a particular area of the screen. By considering the fact that this position is fixed, the score and other relevant information represented in the video were automatically extracted. The match video was split into frames and the portion of the frame displaying the score was cropped. The frame was converted to grayscale, denoised and suitable filters were applied to improve the clarity of the text displayed. The image was enlarged and supplied to tesseract OCR engine[11] to convert text to an editable format. Based on a set of predefined patterns, the string obtained from the OCR engine was checked for a suitable match and the score table was updated. The data for the detected atomic events was stored in an SQL database. SQL stores data efficiently for a small database with a fixed schema..

B. *Encoding of Inherent Structure*

The *structural* and informational event data generated is then processed by OpenRefine[21]. The result from semi-automatic tagging of match videos reformatted into a row-column format is fed into Google Refine. This tabular format as illustrated in Figure 4 is used to bring out the attributes of each tuple in the commentary. Each column in the table represents attributes like over number, ball number, batsman, bowler whereas each row is a ball.

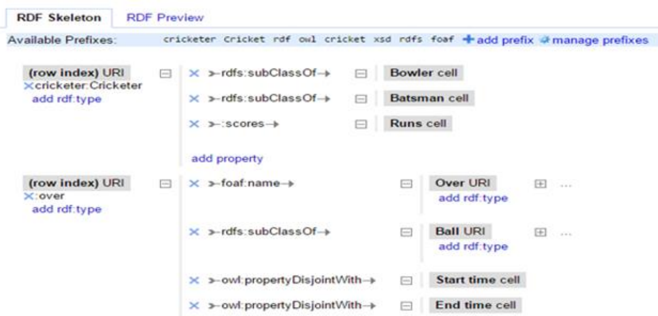

Figure 4: Extraction of event details.

A hierarchical model for the match structure is built using Google Refine. Each level in the match defines the depth of the entity in the cricket match, the smallest entity being the ball in an over and the highest entity being the entire cricket match. The event structure from the event model is represented by an RDF skeleton as shown in Figure 4.

C. *Extraction of qualitative attributes*

The usual commentary that is listed on sites such as CricInfo, CricBuzz, etc., expresses the ball by ball analysis of the entire match. However, there are instances when the commentators engage in casual talks with the players and/or with other commentators. The general topics of discussion are game strategy, team composition, player replacement, pitch and weather conditions, highlights of the game, turning points during the match, runs scored and wickets taken, award ceremony, captain's review, practice sessions. These comments are important for situation recognition and for finding associations between events. A Bayesian classifier with training set as a large dataset with already tagged words is used to tag comments in the commentary with sentiments. For example, "Brett Lee was the best bowler" can be tagged as `<"Brett Lee", "positive">` or "Gilchrist did not play his best today" with the tag `<"Gilchrist","negative">.`

In addition to the comments directly made by the commentator, there are sentiments that can be inferred from the game itself. These sentiments are inferred from a set of predefined rules obtained using general game play principles. A few rules are as follows, `"If runs scored>=3, then it is positive for the batsman"` or `"If a wicket is taken, then it is positive for the bowling side"`. For example, for a tuple with values `<3rd over, 4th ball, 4 runs, Batsman:"Sachin", Bowler:"Brett Lee", Start time: 344.56, End time: 700.67, comment>`, a tag `<"Sachin","batsman","positive">` can be inferred using already defined rules. These represent the experiential aspect of the model. In the following section, we describe how the system recommends video segments to users based on the perspectives.

VII. DATA MINING OF VIDEO LINKS AND ASSOCIATIONS

We now show how the event web constructed from the videos can be used to mine links and associations between the videos in the database.

The topic of experiential association is subjective and in order to get an initial list of possible experiential associations, we conducted a short survey on a group of 15 individuals where they were asked about their associations for cricket and tennis video clips. Our survey revealed that some of the associations requested by the sampled users were "positive instances of a particular player", "best bowler and best batsman of a match", "highlights of a match" for cricket and "summary of a specific set in the match", for tennis. Complex queries such as "similar innings played by a player in an older match", were also requested. Since the objective of the system is to allow richer sharing of experiences by displaying the associations of a particular clip, we describe below how our system can data mine the event web to display these associations.

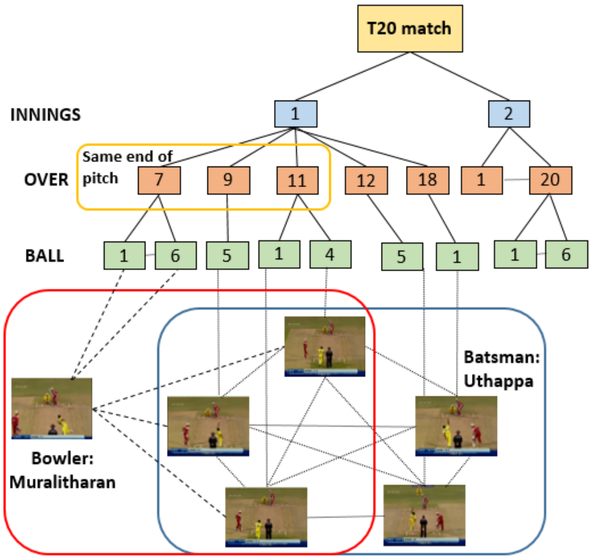

Figure 5: Construction of a social event web for cricket

A. Finding Associations

As an example, we show how the highlights of a match are detected in our system. In the game of cricket, post match presentation events summarize the match as a whole. Our framework detects the mentioned events in the game, queues the video clips and plays them one after the other. For example, consider a T20 cricket match between Royal Challengers Bangalore and Chennai Super Kings. During the post match session, Anil Kumble, the captain of the RCB team, commented “Robin Uthappa’s impressive batting brought the match to our side”. Our framework uses NLP techniques to parse the string and identify the name “Robin Uthappa”. It identifies events in the database that match Robin Uthappa performing the batting role in the *Informational* aspect and a matching sentiment with the query adjective equivalent to “impressive” in the *experiential* aspect; the experience is quantified by testing the adjectives in the query string over a corpus which specifies whether a word has a positive or negative sentiment attached to it as illustrated in Figure 6. Finally, our framework plays the video segments for the same with the help of the time stamp in the annotated file. Note that in this case, there may have been parts of the time Robin Uthappa was batting when his batting was not good (e.g., during the beginning when he is getting used to the pitch) or when his batting was not outstanding. Our system will not display those clips.

An association that is commonly requested by users and can be handled by our system is “highlights of a batsman’s innings” or “positive instances of a player”. As every ball in the match is also tagged with sentiments, all the instances of the game which correspond to the post-match tag can be retrieved using SQL. For example, a tag `<”Virat Kohli”,”batsman”,”Positive”>` from the post-match commentary can be justified by retrieving all the instances of the match where ”Kohli” was the batsman and had a ”positive” sentiment attached to the corresponding tuple. The important thing to note is that we would display only those balls that have a positive sentiment attached to them, and not balls that have a negative or neutral sentiment attached. Similar construction for the event web for tennis is shown in Figure 7.

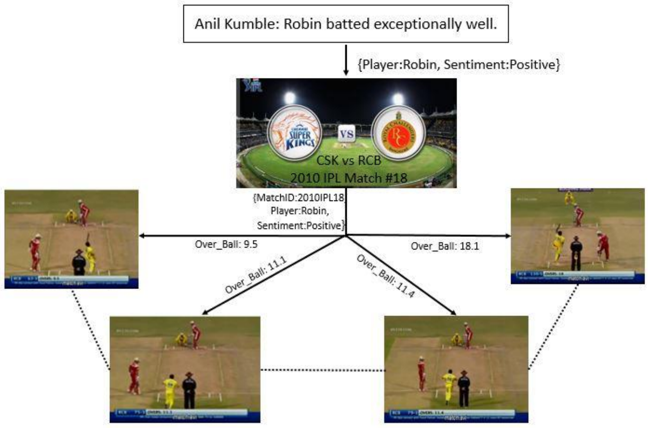

Figure 6: Extracting associations for a Cricket Match

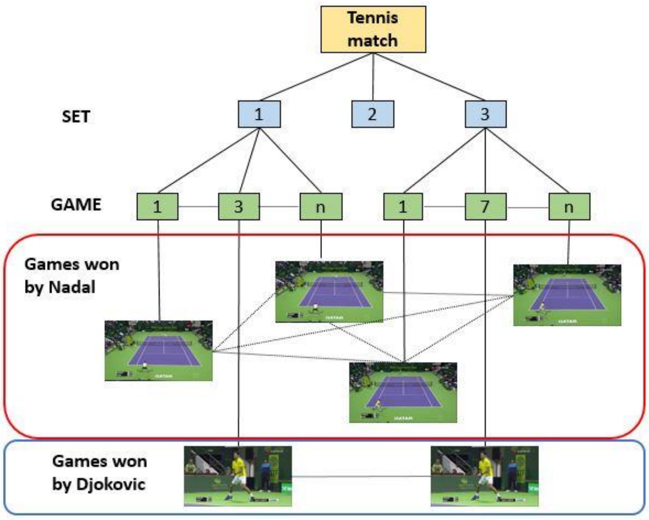

Figure 7: Construction of social event-web model for Tennis

B. Evaluation of our Framework

Our tool permits the users to discover associations and then annotate and share these associations. In order to verify and evaluate the quality of results returned by our system, we took a sample of 7 cricket matches (from IPL 2010 and T20 Blast tournament 2015) and 2 tennis matches. We rely on letting users create, discover and run queries on the system such as “Best batsman of the match”, “positive/negative overs of a player”, etc., and check how accurately our system identifies events for a given set of queries.

To evaluate the results from our system, we use human rating. 10 users of the system (not the same ones who had formulated the queries) were asked to enter the queries and rate the results. The first measure was precision measured on a scale from 0 to 100. We instructed the users to use the rating scale to correspond to a percentage. For example, for the queries "What are the highlights of the match?” a precision rating of 100 indicates that 100% of the highlights were captured. A rating of 0 would indicate that none of the highlights of the match were captured. We also used a recall measure on a scale of 0 to 100. A recall measure of say 70% would indicate that 70% of the events captured corresponded to events which were highlights, and 30% of the events returned were not highlights.

	PRECISION	RECALL
MEAN	93.32	81.7
STANDARD DEVIATION	0.0205	0.0165

Figure 8: Accuracy measures of our evaluation.

Figure 8 shows the average and standard deviation of the precision and recall measures. We observe the low value of standard deviation indicating that users were in agreement.

Another example of an association that can be discovered by our system is "Did a key player perform up to expectations?". The pre-match presentation event contains instances of sentiments associated with various key players. These sentiments need not necessarily be positive; for example, if a player has not played well in previous matches, the sentiment may be negative or neutral. These can be compared with sentiments associated with the player in the match and in the post-match presentation event to find if a player played up to expectations or exceeded expectations. In our future work, we plan to do more queries, and greater numbers of evaluations, as well as use evaluation feedback from users to further refine the features of the system.

VIII. CONCLUSION

In this paper, we have demonstrated that building an event web for a multimedia document, specifically videos, can lead to discovery of deeper associations between the events in the video. This is useful for deeper sharing or experiences as well as for data mining of the video. The social event web is built on an event model, which has been described in detail. When every video in a corpus is represented as an event-web, it will aid users in being able to (a) navigate across multiple videos using ad-hoc queries (b) discover associations/experiences that span across multiple videos thus allowing us to realize our goal of navigating through a visual web of videos. To illustrate the generality of our approach, we have applied this technique to two sports - cricket and tennis, and demonstrated that our approach can satisfactorily handle both sports, and due to the similarity of the event models, there is a lot of design reuse between the two cases.

REFERENCES

[1] Bailer, W., Mayer, H., Neuschmied, H., Haas, W., Lux, M., and Klieber, W. Content-based video retrieval and summarization using mpeg-7. In *Electronic Imaging 2004* (2003), International Society for Optics and Photonics, pp. 1–12.

[2] Borth, D., Ulges, A., Schulze, C., and Breuel, T. M. Keyframe extraction for video tagging & summarization. In *Informatiktage* (2008), vol. 2008, pp. 45–48.

[3] Ekin, A., Tekalp, A. M., and Mehrotra, R. Automatic soccer video analysis and summarization. *IEEE Transactions on Image processing 12*, 7 (2003), 796–807.

[4] Evangelopoulos, G., Zlatintsi, A., Skoumas, G., Rapantzikos, K., Potamianos, A., Maragos, P., and Avrithis, Y. Video event detection and summarization using audio, visual and text saliency. In *2009 IEEE International Conference on Acoustics, Speech and Signal Processing* (2009), IEEE, pp. 3553–3556.

[5] Hu, W., Xie, N., Li, L., Zeng, X., and Maybank, S. A survey on visual content-based video indexing and retrieval. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews) 41*, 6 (2011), 797–819.

[6] Jain, R. Eventweb: Developing a human-centered computing system. *Computer* (2008).

[7] Jain, R. Eventweb: towards social life networks. *Philosophical Transactions of the Royal Society of London A: Mathematical, Physical and Engineering Sciences* (2013).

[8] Jain, R. Let's weave the visual web. *IEEE MultiMedia* (2015).

[9] Jiang, F., Yuan, J., Tsaftaris, S. A., and Katsaggelos, A. K. Anomalous video event detection using spatiotemporal context. *Computer Vision and Image Understanding 115*, 3 (2011), 323–333.

[10] Jung, Y.-K., Lee, K.-W., and Ho, Y.-S. Content-based event retrieval using semantic scene interpretation for automated traffic surveillance. *IEEE Transactions on Intelligent Transportation Systems 2*, 3 (2001), 151–163.

[11] Smith, Ray. "An overview of the Tesseract OCR engine." *Document Analysis and Recognition, 2007. ICDAR 2007. Ninth International Conference on.* Vol. 2. IEEE, 2007.

[12] Li, B., and Sezan, M. I. Event detection and summarization in sports video. In *Content-Based Access of Image and Video Libraries, 2001.(CBAIVL 2001). IEEE Workshop on* (2001), IEEE, pp. 132–138.

[13] Snoek, Cees GM, and Marcel Worring. "Multimedia event-based video indexing using time interval." *IEEE Transactions on Multimedia* 7, no. 4 (2005): 638-647.

[14] Song, J., Dan, A., and Sitaram, D. *Efficient retrieval of composite multimedia objects in the JINSIL distributed system*, vol. 25. ACM, 1997.

[15] Song, J., Dan, A., and Sitaram, D. Jinsil: A middleware for presentation of composite multimedia objects in a distributed environment. *Multimedia systems 8*, 4 (2002), 295–314.

[16] Wang, F., and Ngo, C.-W. Rushes video summarization by object and event understanding.

In *Proceedings of the international workshop on TRECVID video summarization* (2007), ACM, pp. 25–29.

[17] Wang, M., Hong, R., Li, G., Zha, Z.-J., Yan, S., and Chua, T.-S. Event driven web video summarization by tag localization and key-shot identification. *IEEE Transactions on Multimedia 14*, 4 (2012), 975–985.

[18] Westermann, Utz, and Ramesh Jain. "Toward a common event model for multimedia applications." IEEE multimedia 14, no. 1 (2007).

[19] Wikipedia. Cricket — wikipedia, the free encyclopedia, 2016.

[20] Wikipedia. Espncricinfo — wikipedia, the free encyclopedia, 2016.

[21] Wikipedia. Openrefine — wikipedia, the free encyclopedia, 2016

[22] Wikipedia. Tennis — wikipedia, the free encyclopedia, 2016.