Paper:
Legg, P., Chung, D., Parry, M., Bown, R., Jones, M., Griffiths, I. & Chen, M. (2013). Transformation of an Uncertain Video Search Pipeline to a Sketch-Based Visual Analytics Loop. *IEEE Transactions on Visualization and Computer Graphics, 19*(12), 2109-2118.
http://dx.doi.org/10.1109/TVCG.2013.207

Transformation of an Uncertain Video Search Pipeline to a Sketch-based Visual Analytics Loop

Philip A. Legg, David H. S. Chung, Matthew L. Parry, Rhodri Bown, Mark W. Jones, Iwan W. Griffiths, Min Chen

Abstract— Traditional sketch-based image or video search systems rely on machine learning concepts as their core technology. However, in many applications, machine learning alone is impractical since videos may not be semantically annotated sufficiently, there may be a lack of suitable training data, and the search requirements of the user may frequently change for different tasks. In this work, we develop a visual analytics systems that overcomes the shortcomings of the traditional approach. We make use of a sketch-based interface to enable users to specify search requirement in a flexible manner without depending on semantic annotation. We employ active machine learning to train different analytical models for different types of search requirements. We use visualization to facilitate knowledge discovery at the different stages of visual analytics. This includes visualizing the parameter space of the trained model, visualizing the search space to support interactive browsing, visualizing candidature search results to support rapid interaction for active learning while minimizing watching videos, and visualizing aggregated information of the search results. We demonstrate the system for searching spatiotemporal attributes from sports video to identify key instances of the team and player performance.

1 Introduction

With the advances in video capture and storage, many application areas have come to rely on collecting vast amounts of video data for the purpose of archive and review [16]. For instance, CCTV security systems can be capturing up to 24 hours of video each day from multiple camera sources. Likewise in sport, analysts will collect video data from multiple cameras for every match of a season. They may also be collecting and reviewing match videos from opposition teams to analyse traits in their performance that can then be exploited in their own strategy. This typically results in hundreds of hours of video footage (in our case 85TB and growing). The task of then reviewing this video content in order to identify interesting or important passages of play now becomes a significant challenge for an analyst to perform. Often the analysts will need to collect together a selection of example video clips that illustrate team strategy, to then analyse whether it resulted in positive or negative outcome for the team. However, identifying such clips from the video soon becomes a tedious and laborious task for the analysts to perform manually.

Many professional sports currently adopt an approach known as *notational analysis* [22]. Notational analysis is a process that allows for the occurrence of time-based events to be manually recorded in correspondence to a synchronised video capture that can then be reviewed at a later time. Additional notes can be recorded with a timestamp, such as technique deployed or the number of players involved. Whilst this approach is widely used, it is a highly intense process for the analysts to conduct during the game when their expertise could be utilised much more effectively for crucial decision-making. Most importantly, notational analysis does not offer any means of directly searching the video content; it merely allows the user to search the tagged annotations that are associated with a video timestamp. Neither does it allow the user to obtain any more information than what was initially recorded in the first instance.

Video search enables finding segments from a collection of videos based on particular search criteria. One approach to video search is to use a sketch-based search, whereby through sketching the user can indicate a particular spatio-temporal pattern that can be associated with the video content. For instance, the user may sketch out a particular path of motion to find when people travel in that direction, or draw a region to query when ‘action’ occurs in that area. Developing a system that can support such open search queries poses some challenges. Firstly, the parameter space of possible sketches that the user could perform is significantly large. Secondly, the result expected by the user may differ from that returned due to other factors that can be difficult either for a user to encode them into the sketch, or for a system to interpret. Hence, the video search pipeline introduces a substantial amount of uncertainty and may not suitably fit to the analysts needs.

In this work, we propose a novel visual analytics approach to sketch-based video search. We incorporate intuitive sketch-based interactions that allow a user to search queries based on spatio-temporal attributes (e.g., motion, position, distance, trajectory, and spatial occupancy), with application to sports video data. The system computes the similarity between the user sketch and the video content to present visual feedback against the video timeline, along with normalized manevoure visualizations that depict segments from the video. The user can then browse the visual feedback and the normalized manevoure visualizations to explore the video data. To facilitate learning, the user can also choose to accept or reject particular results to train the system. This enhances the accuracy of the similarity visualization accordingly and incorporates human knowledge into the model. Such a step results in a tailored solution that begins to reduce the ambiguity that user sketches could have on an untrained system, creating a more powerful analytical system for the end-user. Traditional notational analysis could also be incorporated into the proposed system to combine the benefits of both methodologies.

The paper is organised as follow: Section 2 provides a review of related works on sketch-based video search, video retrieval techniques using active learning, and video visualization and summarization techniques. Section 3 presents the system requirements and discusses the motivations behind the conducted work. Section 4 discusses the visual analytics system, and details the variety of components (video processing, user interaction, video search, visual analytics loop and visualization) that the system incorporates. Section 5 presents a case study conducted with the analysts of the Welsh Rugby Union, that identifies three scenarios where the sketch-based visual analytics can reveal novel insight to match analysis. Section 6 provides a quantitative evaluation of the system based on time, recall and precision measures. Section 7 gives discussion on the proposed system and how this can be extended for future use. Finally, Section 8 provides conclusion of our work.

- *Philip Legg was with the Department of Computer Science, Swansea University, and is now with the University of Oxford.*
- *David Chung, Matthew Parry and Mark Jones are with the Department of Computer Science, Swansea University.*
- *Rhodri Bown is with the analysis team at Welsh Rugby Union.*
- *Iwan Griffiths is with the Department of Engineering, Swansea University.*
- *Min Chen is with the e-Research Centre at the University of Oxford.*

Manuscript received 31 March 2013; accepted 1 August 2013; posted online 13 October 2013; mailed on 27 September 2013.
For information on obtaining reprints of this article, please send e-mail to: tvcg@computer.org.

2 Related Works

We consider the related works of three particular areas of research: visual and sketch-based search, video retrieval techniques and video visualization (summarization) methods.

The review by Schoeffmann *et al* [30] presents the current state of the art for video browsing interfaces and applications. Sivic *et al.* studied the use of visual search techniques for images and video content [32, 31, 33]. Deng and Manjunath [15] use colour, texture and motion for content-based video search. Chang *et al.* [8, 9] present VideoQ as a tool for content-based video search using visual cues. Chen and Chang present Vismap as a visualization tool for image and video retrieval systems [12]. Zhong and Chang [41] use spatio-temporal video search for object detection. More recently, Lee *et al.* [23] use sketch-based search techniques for retrieval of 3D objects. Hsu *et al.* [21] propose reranking methods for visual search. Vrochidis *et al.* study user feedback as part of visual search for interactive video retrieval [38]. Hao *et al.* [18] present IVQuery that combines visual interaction with automated analytics for common data visualization tools. Bowman *et al.* [7] present a visual analytics system for identifying relational patterns in MRI data.

Chang *et al.* present some of the recent advances and challenges of semantic image and video search [10] such as semantic concept classification and local image features for object representation. Tesic *et al.* [37] use textual queries in conjunction with underlying semantics taken from visual queries on image/video data to enhance video search results. Snoek *et al.* propose MediaMill that supports semantic querying of video based on lexicon of 100 automatically detected semantic concepts [35]. They also extended the work using machine learnt concepts to build a multimedia thesaurus [34]. Much interest has come of using machine learning techniques for active learning in search query tasks. Hauptmann *et al.* [19] present a video search system that incorporates active learning from a human user to improve search query re-ranking. Similarly, Rodrigues *et al.* [29] study content-based data retrieval in conjunction with visual analytics, using multivariate visualizations, parallel co-ordinates, scatter plots and table lens, to examine the search query metrics in greater detail and identify suitable matching results. Wang *et al.* [39] present a boosted multi-task learning framework for face verification based on limited training data that aims to reduce overfitting. Zhang *et al.* [40] adopt an interactive learning process for content-based image retrieval based on a SVM classifier, that illustrates how recall and precision rates are improved after several learning iterations. Similarly, Lui *et al.* [25] extends the SVM approach beyond considering only relevant and irrelevant results, to incorporate search ranking results for video search queries. Zhong and Chang [42] propose a framework for scene detection based on domain-specific knowledge and machine learning techniques that they apply to tennis and baseball sports videos. However, this only demonstrated the ability to recognise serves (tennis) and pitches (baseball) rather than facilitating the searching of video for key events of interest. Lastly, Baluja *et al.* [3] study the task of large video search and recommendation on YouTube content, using a graph-based algorithm that incorporates user ratings and interactions.

Video visualization is an emerging area that aims to summarize large video content through the use of visualization. The recent survey by Borgo *et al.* [5] provides a good overview to the topic area. The topic was first introduced by Daniel and Chen [13] who demonstrate it for surveillance and television broadcast summarization. Chen *et al.* [11] extend this further by considering visual signatures of motion flow in the video visualization. Botchen *et al.* [6] investigate the inclusion of action-based detail in video visualization. Sports video visualization has also become popular, where Parry *et al.* [27] used video visualization for Snooker match summarization, and Legg *et al.* [24] use it for in-match performance analysis in Rugby. Most recently, Hoferlin *et al.* [20] looked at sketch-based video visualization for the analysis of video content, highlighting the fact that sketch-based search queries would be well-suited to video visualization applications.

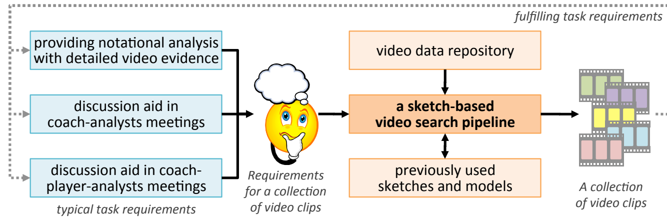

Fig. 1. The tasks of searching video clips are performed routinely by sport analysts in order to meet various objectives for collecting videos such as event analysis, match planning and meeting presentations. The specification for the video clips to be retrieved are typically not well defined beforehand and may change dynamically during the search. Such tasks exhibit some common characteristics of many visual analytics applications.

3 System Requirements

Traditional sketch-based image or video search relies on a trained analytical model to find a set of appropriate segments from the video. Once trained, the model is used for working on all subsequent search tasks. In such a case, the model has to work for all sketch inputs. However, the model is not expected to be reliable because:

- The parameter space of possible sketches suggest that the learning must be based on a very large training data set in order to achieve the necessary statistical significance. However, there are a limited number of rugby videos, and more importantly, analysts do not have sufficient human resources to annotate video training data.
- The results that a user would expect to achieve are usually influenced by other facts that are difficult to pre-determine in a search specification (for instance, the minimum and maximum numbers of players involved), or difficult to be determined (at least currently) from the video by the computer automatically (e.g., a particular type of tactical move). Hence, visual inspection of the search results in essential.

For over 2 years we have worked in close collaboration with the Welsh Rugby Union to develop a suite of applications that integrate visualization into their training processes through notational explorations [24], high level dashboard representations and more recently, this proposed visual search approach. Fig. 1 illustrates the iterative analysis process of video search that would facilitate coaching, training and player feedback. Their problem is that they have a huge ever-expanding library of video match data (over 85TB) that is currently under-utilised due to the lack of effective interfaces into the data. Such a process should significantly reduce the burden of manually searching this video data repository, whilst being flexible enough to support the various needs of collecting video clips such as for event analysis, match planning and meeting presentations. Through regular meetings with the WRU, we developed a set of requirements for the visual search system. These included desirable characteristics for the software, technical requirements and visual requirements.

Fig. 1 illustrates typical tasks of searching for video clips by sports analysts. The requirements usually are defined in an ad hoc manner with objectives to support various needs for event analysis, match planning and meeting presentations. The essential requirement is a sketch-based search system to order to provide a high degree of flexibility in defining a search query. The user should be able to sketch a play and use that as input to the system to make it as intuitive and interactive as possible. Search should be on general motion of the team, position of the team on the pitch, and distances between players. This should allow for exploration of groupings within the team to provide greater flexibility (e.g., forwards and backs). The results of this search should be a set of matching videos based on similarity

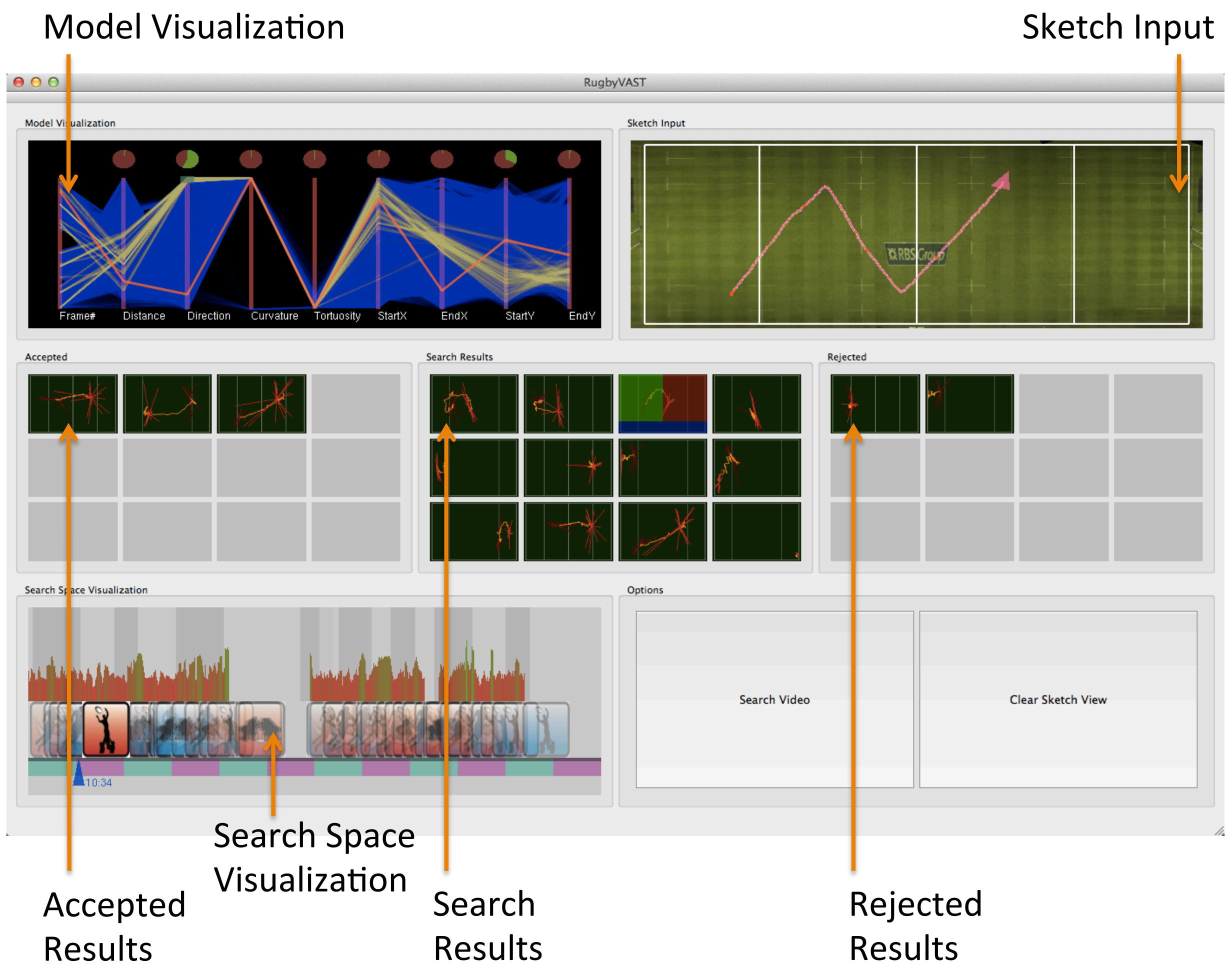

Fig. 2. The user interface consists of a number of interactive panels to encourage data exploration. The sketch input panel allows a user to draw a particular search query using intuitive sketch tools. The search results panel displays the top 12 video segments that correspond with the user sketch based on the current model parameters. The model visualization allows the user to analyse how each video segment corresponds to the different similarity metrics within the model. The search space visualization shows the overall similarity against the video timeline, in conjunction with match event data to provide context to the game. When the user accepts or rejects results, they are moved to the corresponding panels and the model is updated to reflect their decision, depicted by the weighting parameters in the model visualization.

metrics for video comparison. These videos should closely match the requirements and intention that the user specifies in the sketch.

Further discussions involved the fact that without a large enough training set the system may not achieve the desired level of accuracy. We therefore pursue the idea of visual analytics to present the search results to the user and allow further interaction to tune the system through the acceptance or rejection of results. The user should be presented with an overview visualization that depicts the play, along with additional interactive analysis tools that would help inform their decisions. This should include a search space visualization that shows the search similarity in conjunction with match event data recorded using notational analysis, and a model visualization that depicts how each video frame compares against the similarity metrics that the system employs. The user should be able to interact with the overview visualization in conjunction with video keyframes. When results are accepted or rejected, the training model is refined by weighting the similarity metrics in favour of accepted results. The user should also be able to examine the data in finer detail, with linkage back to the video content.

4 Visual Analytics System

We developed a visual analytics system that brings various components from the traditional video search pipeline into a closely integrated system to enhance the exploration and discovery of video data. By utilizing iterative visual analytics, not only can the user explore and retrieve the desired video data, but through repeated use they also facilitate supervised learning (active learning) and improve the underlying analytical model that is used for future search queries.

4.1 Overview

The visual analytics interface (Fig. 2) consists of the followings:

- Sketch Input — this panel allows the user to draw a search query using intuitive sketch-based tools to convey motion, position and distance.
- Model Visualization — this panel uses parallel co-ordinates to convey how the video corresponds to the individual similarity metrics that the model comprises of.
- Search Space Visualization — this panel uses a timeline to convey the overall similarity as defined by the model in conjunction with match event data to provide context to the game.
- Search Results — this panel shows the top 12 video segments based on overall similarity as defined by the model, illustrated using a Normalized Manoeuvre Visualization (NMV).

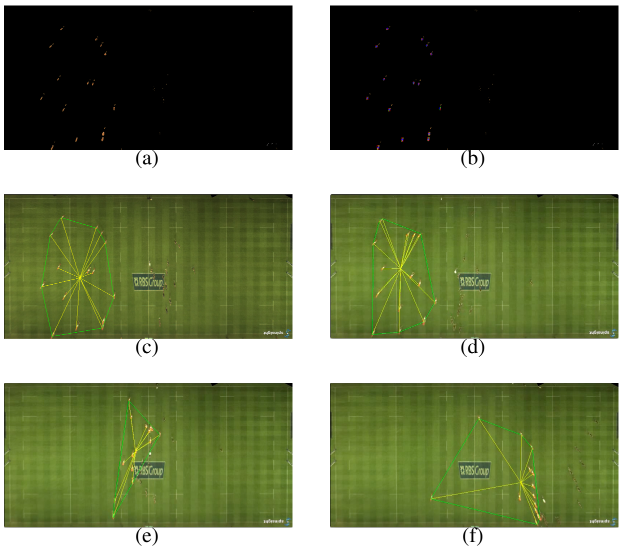
(a) (b) (c) (d) (e) (f)

Fig. 3. Detection and tracking of players from top-down view. (a) Image converted to HSV colour space to extract shirt colour of the home team (e.g., red). (b) Connected component analysis used to determine position of each shirt. (c) Convex hull applied to point set and centroid is calculated. (d)-(f) Subsequent video frames using same process.

- Accepted Results and Rejected Results — these panels show the accepted and rejected results as chosen by the user, when applicable.

Each panel supports interaction from the user, to encourage greater exploration of the data. Panels are also linked so that interaction in one will update the other panels accordingly. In the sketch input panel, the user defines the search query by drawing. By sketching, all panels are updated in accordance to the new search query. In the model visualization panel, the user can 'brush' any axis to highlight the selected parallel co-ordinate polylines. The corresponding video segments are also highlighted in the search space visualization. The user can also manually define the weight of an axis, or lock the weight of an axis. This interaction will adjust the search space visualization accordingly. In the search space visualization, the user can navigate along the timeline to analyse the match event data at a particular time, in conjunction with the overall similarity plot. Double-clicking will display a popup window of the video for the selected time. In the search results, the user can select a particular result (depicted by the NMV) and use the coloured selection tool to choose either accept, reject, or to open a popup window of the video at that particular time should more detail be required. Accepting or rejecting a result will update the learning model, by computing new weighting functions based on the correspond similarity metric values. This in turn will update the weights presented in the model visualization, and the overall similarity as shown in the search space visualization.

4.2 Video Processing

As a pre-processing stage to the visual analytics, the system needs to extract player and team spatial information for each frame of the video data. To facilitate this, we use a synthetic top-down view that is generated using perspective transformation of a stitched pitch view captured by three cameras (for this work we use the Spinsight camera system). By having a fixed-camera top-down view, we have a point of reference to the pitch for all player and team motions.

Fig. 3 shows the process of extracting team and player spatial information from the Spinsight video. In (a) the video frame is converted to HSV colour space, and then we use a threshold to preserve only the shirt colour that we wish to detect (e.g., red for the Welsh Rugby Union). In (b) we use connected component analysis to identify each player as a 'blob' within the scene. This allows us to determine the centroid of each player. Once players are identified, we can calculate various statistics including the centroid of the team. In (c), we visualize a convex hull placed around the players to convey the team shape and region coverage, with yellow lines connecting each player to the team centroid forming a star visualization. Images (d)-(f) show subsequent video frames and the change in team shape that can be analysed by this approach.

For each video frame we record each individual player position and calculate the team centroid. An 80 minute rugby match captured at 20 frames per second would consist of 96000 video frames, where each frame will have up to 15 player positions, and a team centre position, therefore simply rendering positions leads to severe occlusion. Therefore we must explore better visualisation approaches for our visual analytics system.

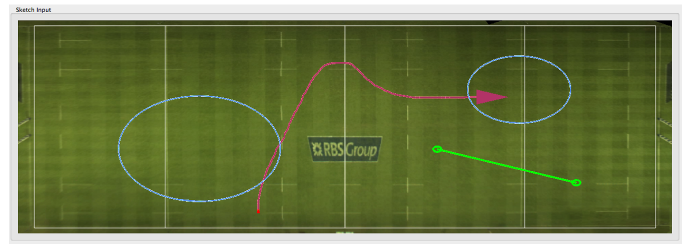

Fig. 4. User sketch interface. The user can define three forms of action: motion, defined by the red arrow, position, defined by the blue circle, and distance, defined by the green cord. The user can also choose for each tool whether to apply this to the team centroid, the forwards (players in front of the team centroid) or the backs (players behind the team centroid).

4.3 Sketch-based Interaction

At the heart of our visual analytics system is visual search using sketch-based interaction. It is therefore crucial that this functionality is intuitive and carefully designed so that analysts can fully exploit the potential capabilities. In coaching scenarios it is quite common for coaches to draw and depict the motions and positions that the team should form in order to achieve a successful result. Coaches tend to draw arrows that show where particular players should run to, or circles that show the area that particular player(s) should be at. Such sketches are often referred to as the coach's playbook. In this system, we take this concept of sketching a rough sequence of play as the input for querying the video content to find periods of play that correspond with this.

Fig. 4 shows an example of the sketch-based interface. The rugby pitch is displayed in the background in order to provide context to the coaches and analysts. By using the mouse, or touchscreen interaction, the user can draw directly onto the pitch area. The system supports three different drawing tools and that can be selected from a tool palette using the right mouse button. These tools are:

- **motion path**, depicted using a red arrow. This search would determine when the team centroid traverses the drawn path within a specified time duration (default duration is within one minute).
- **positional region**, depicted using a blue circle. This search would determine when all players in the team are positioned within the drawn region.
- **distance measure**, depicted using a green chord. This search would determine when there is the drawn length between the most-forward and most-back players. Other distances could be measured including distance across width of pitch, and distance between forwards and backs groupings.

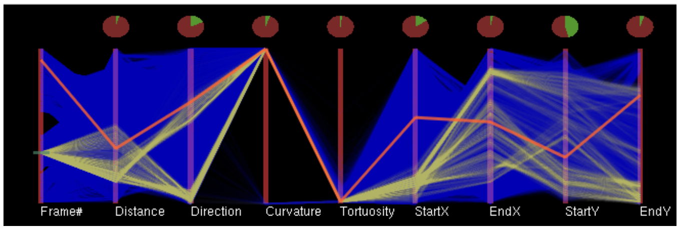

Fig. 5. Model Visualization. Parallel co-ordinates show each video frame (first axis) against each similarity metric (subsequent axes) in our model. Each polyline represents a video segment (which consisting of a video frame and duration). Weighting of each similarity metric is shown using a dial view above the axis. This indicates its contribution to the overall simiarlity measure. User can brush polylines to explore data (shown in yellow). The selected search thumbnail is shown in red.

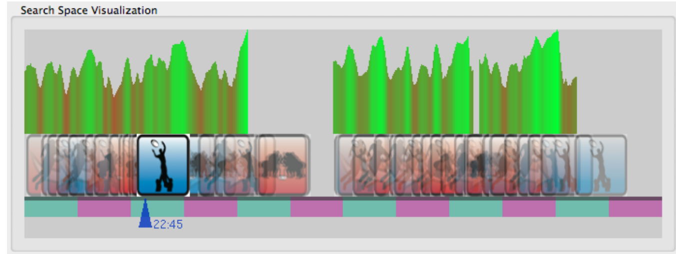

Fig. 6. Search Space Visualization. This plot indicates the search similarity between the user sketch and the current model (based on video data and training data). Colour and height are used to emphasise similarity (red to green indicating minimum to maximum). The coloured segments along the bottom indicate ten-minute time periods, and the blue arrow shows the current selection of the user.

Tools can be combined to create more detailed search queries, for instance, a user may wish to draw multiple motion paths, or specify two regions and a motion path connecting them. The system uses a logical OR to be able to handle multiple tools. Motion paths are re-sampled to give 100 equally spaced points along the path. This is important for computing the similarity metrics discussed in Section 4.4.

4.4 Search Similarity and Model Training

Given a sketch, the system needs to compare this against the extracted video data. One approach would be to compute whether the user sketch is equal to each segment in the video. However, due to the rough nature of sketching and also noise in the video, it is quite clear that this would not yield particularly meaningful results. Instead, we use a *similarity measure* to calculate to what extent the user sketch resembles the content of the video.

Due to the arbitrary nature of sketching, there are many possible search queries that could be defined by the user sketch. Therefore, it may be that there is not one similarity measure that can sufficiently deal with all possible user sketches. Instead, we devise a training model for our system that incorporates multiple similarity metrics with associated weighted contribution. We define the user sketch to the set of points P. When searching the video, we define the points from the current video segment to be V. We normalize the point sets so that each consists of $n = 100$ points. For the video segment V, this is calculated over the user-specified time duration (default is one minute). The similarity metrics incorporated into the model are:

- **mean distance** — the mean absolute distance between P and V.
- **direction** — the directional difference between P and V.
- **curvature** — the curvature difference between P and V.
- **tortuosity** — the tortuosity difference between P and V.
- **start X** — the absolute difference between $P_{1,x}$ and $V_{1,x}$.
- **end X** — the absolute difference between $P_{n,x}$ and $V_{n,x}$.
- **start Y** — the absolute difference between $P_{1,y}$ and $V_{1,y}$.
- **end Y** — the absolute difference between $P_{n,y}$ and $V_{n,y}$.

The *absolute distance* is computed between the corresponding points in P and V, from which we compute the mean value. Similarly, the *direction* is computed based on the vector from the start point to the end point of each point set, from which we calculate the difference value. For computing *curvature* and *tortuosity*, we adapt the approach taken by McLoughlin *et al.* [26] for computing dissimilarity in interactive streamline seeding. Curvature is computed by $\mathbf{v} \times \mathbf{a}/|\mathbf{v}|^3$ where $\mathbf{v}$ is the local velocity and $\mathbf{a}$ is the local acceleration computed by multiplying the local velocity gradient with the local velocity. Tortuosity of P is computed as $\frac{1}{||p_n - p_1|| \sum_{i=1}^{n-1} ||p_{i+1} - p_i}$, where p_i is the i^{th} point in P, and n is the number of points on the line. Tortuosity of V can be computed similarly. The final four metrics measure absolute distance between the start and end positions of P and V, i.e., $(p_{1,x}, p_{1,y}), (p_{n,x}, p_{n,y}), (v_{1,x}, v_{1,y}), (v_{n,x}, v_{n,y})$. This allows us to examine the independent components of the team spatial positioning. This provides the ability to disassociate spatial positioning from similarity. For example, the analyst may wish to know when the team traverse a particular motion path, but is not necessary concerned with where on the pitch it took place. Where a search consists of multiple sketch elements, the system can either choose the mean result or the minimum result for each measure. By default we select the minimum result, to emphasise that the system has matched closely with an element of the sketch. Each metric is scaled to be between 0 and 1.

We incorporate each of the eight similarity metrics into our training model by a weighted contribution. For each similarity metric $s_1 \dots s_n$, there is an associated weighting term $w_1 \dots w_n$. Initially, all weighting terms are equal (e.g., $1/n$, where n is the number of similarity metrics). If the user accepts a search result, the weight of each metric becomes $w_n = w_n(1+\omega)$ where ω is the result of the similarity metric for the selected video segment (scaled between 0 and 1). Likewise, if the user rejects a search result, the weight of each metric becomes $w_n = w_n(1-\omega)$. Weights are then normalized so that $\sum w = 1$. Through the process of accepting and rejecting search results, the model adjusts to favour the metrics that closely resemble the user's acceptance criteria.

4.5 Visualization

After performing the search query, the system will provide visual feedback. The system features a number of visual components that encourage exploration of the data, provide video summarization, and facilitate supervised learning. There are three visual components that will be discussed in this section: model visualization, search space visualization, and normalized manoeuvre visualization.

4.5.1 Model Visualization

Fig. 5 shows the model visualization. This provides a visual representation of the similarity metrics used for computing the visual search. We use parallel co-ordinates since they provide a well-established method for intuitive representation of multi-dimensional data. The first axis indicates the frame number, and then the subsequent axises each denote one of the similarity metrics used within the model. This allows for visual inspection of the current state of the model, and provides a facility for the user to identify which similarity metrics contribute well to the outcome result for a particular search task. By allowing the user to explore this, it provides a verification that the model is performing as expected, rather than simply treating this as a black box approach. Since we use parallel co-ordinates, we can also incorporate standard interaction techniques including brushing, so that the user can highlight polylines of interest for further exploration.

For each similarity metric in the parallel co-ordinates plot, there is also a dial view positioned above the axis. This depicts the current weighting that the similarity metric contributes to the trained model. As discussed in Section 4.4, all weightings are initially equal. The

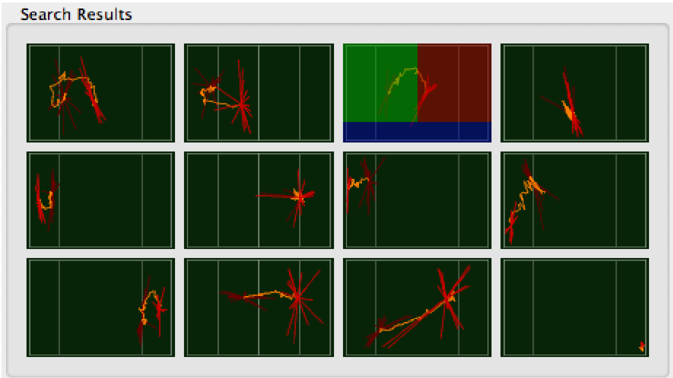

Fig. 7. Search results illustrated using the Normalized Manoeuvre Visualization. The NMV shows the team starting position (dark red star), the team motion (orange path), and the current position (light red star). Individual player positions are also depicted (red dots).

Fig. 8. Thumbnail viewer of NMV. The user can interact with the NMV by adjusting the duration slider to animate the motion path. The corresponding video keyframes are also displayed. The user can overlay additional information onto the video, including the NMV and distances between the team, the forwards and backs, in both the x and y directions.

model then adjusts in accordance to accepted and rejected search results. The user can also choose to adjust the weights manually through interaction with the visualization, or to lock a weight so that the system can not adjust the current value for that particular metric.

4.5.2 Search Space Visualization

Fig. 6 shows the search space visualization. The visualization incorporates the outcome of the analytical model and shows the overall similarity based on the weighted contributions of each metric. The similarity is shown in the plot using both height and colour (where red is low similarity and green is high similarity), to emphasise key video segments from the data that correspond well. The plot is dynamic in that as the system (or the user) modifies the weighting contributions then the similarity plot here will update accordingly. Below the similarity plot we incorporate glyph-based visualization to illustrate match event data that corresponds with the video [24]. Due to the confined space, we modify our approach from [24] to develop a condensed representation that utilises the intended user interaction. As the user navigates the timeline at the bottom of the visualization, the current match event is displayed at the front of the view for clear understanding. Context is provided by showing the remaining event glyphs using transparency. This allows the user to understand what match events occur at each point in the video, combining notational analysis data with the spatial search performed by the user. For instance, a user may search for when a particular motion occurred, but may be interested in only when this occurs as a result of a lineout event. The combination of notational analysis data and spatial information now allows an analyst to investigate such scenarios, which was not previously possible. This also helps inform the user for accepting or rejecting search results.

User interaction of this view is performed by dragging the timeline position marker to navigate left and right along the timeline. By double clicking at a particular time, the system displays the video at that current time. The user can also filter the search space by selecting region of interest from the similarity plot.

4.5.3 Normalized Manoeuvre Visualization

Fig. 7 shows the Normalized Manoeuvre Visualization (NMV), that is used to depict the search results to the user. The system presents the top 12 search results as defined by the overall similarity measure (the user can modify the number of results to show if they should wish). The NMV serves as a thumbnail presentation of a period of play from the video data. This allows for much more information to be encoded than simply presenting a video keyframe. This is particular so in sport where many video frames will have a very similar appearance, despite that there may be very different action occurring in the match. The visualization comprises of three key components: start position, team motion, and current position. The starting position is depicted using a dark red star that shows the team's initial positions at the begin of the selected time period. The orange track shows the team centroid over the specified duration by the user (default value is one minute). The brighter red star shows the current position of the team. In addition, the user can also choose to show individual player positions over the period with red positional markers. By default these are not shown in the thumbnail view due to the condensed representation of the data, of which individual positional markers may not appear clear). The user can choose to view the NMV for the entire team, the forwards, the backs, or a combination of all three. The same visualization convention is used for forwards and backs using blue with a green path, and pink with a purple path, respectively (as shown in Fig. 11). The visualization provides a quick summary of the action during the video segment in a concise manner. The visualization shows some important cues that can help the analysts, including ruck positions (shown by the clustered purple motion path), and also helps to quickly identify whether the team have pushed forward or dropped back on the pitch.

User interaction of this view allows the user to select a particular thumbnail of interest. On selection, an overlay on the thumbnail is displayed that presents the user with three colour-coded options: accept (green), reject (red), and more detail (blue). The first two options move the thumbnail from the search results into the appropriate category for display. Should the user wish to examine the thumbnail further, by selecting more detail a popup window is displayed (Fig. 8). The popup window displays an interactive NMV where the user can adjust the duration parameter, and also view the corresponding video keyframes. This allows the user an additional level of exploration should they require it, in order to make a well-informed decision on whether to accept or reject the presented result.

5 Case Studies

To evaluate the use of our system we have worked in close collaboration with the Welsh Rugby Union who operate with a strong emphasis on performance analysis. Traditionally, the analysts have worked primarily with notational analysis data and so this system presents a novel process for conducting analysis in the future. We tasked the analysts with studying team performance from video data collected during the 6 Nations tournament. From spending time with the system, it became clear that the analysts were now able to explore and interrogate the data in a more interesting and detailed approach, that helped to reveal new insights into their performance. In the follow sections we present three cases where the analysts were able to derive new insight and understanding by using the system.

Fig. 9. Example to show the analyst's workflow. The analyst wanted to see when the team advance down the pitch, and sketched this motion. The analyst found a suitable result (highlighted by the green circle) by examining the search space visualization and the NMV thumbnails. From browsing the NMV, the analyst could see the correspondence in the motion, and by adjusting the duration this revealed that two tries had actually been scored, in quick succession of each other.

5.1 "When do the team move in a particular form?"

Fig. 9 shows the process taken by the analysts. The analysts began by sketching different search queries using the drawing tools available to become familiar with how the system functioned. To begin with the analysts concentrated on studying motion paths within the video data. Their first inclination was to find when the team advance down the pitch in order to score a try. From sketching the desired motion, the system presented thumbnails of most similar searches. In conjunction with the search space visualization, the analysts found a particular example that occurred following a lineout. The analysts observed that the motion corresponded with the sketch, and from browsing the NMV discovered that two tries had actually occurred over a very short period of time. The team had quickly advanced up the pitch following the restart to score again. The analysts accepted the search result and in doing so trained the model for future iterations.

From performing some initial sketches, the analysts came to realise that their motion tended to be more zig-zag than straight advancement. In Rugby, play is often broken down into phase ball events, whereby each ruck or maul is followed by a new phase. What the analysts found particularly interesting was the ability to visualize the motion of these phases leading up to ruck and maul events, the paths travelled by the team, and the distances between ruck and maul events.

5.2 "How did the opposition manage to score?"

Here, the analysts wanted to begin to explore weaknesses in their own performance. To achieve this, the analysts used the regional position tool to indicate when players were in their own defensive half. The system returned proposed results that the analysts could then examine. Since selection of the results is linked to updating the search space visualization, analysts could identify the current event from the notational analysis data. Fig. 10 shows four keyframes taken from a particular example that they uncovered. Following an opposition lineout, the opposition managed to push the home defence back due to poor positioning of the team. The NMV reveals the position of the mauls, and the distance gained by the opposition between each maul. Eventually, the play resulted in the opposition being awarded a penalty from which they scored. The NMV can help analysts to overcome such shortcomings in future. More importantly, the NMV allows for the analysts to convey to the players what has happened so that they understand what they need to improve upon.

5.3 "What is the shape of the forwards and backs?"

Fig. 11 shows an example of studying the performance of the forwards and backs independently. The analysts were searching for pitch ad-

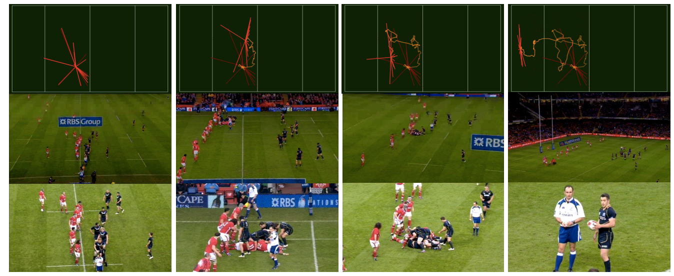

Fig. 10. Four keyframes to illustrate the opposition scoring using the NMV (Top-to-bottom: NMV showing full team, wide-angle video keyframe, close-up video keyframe). From the opposition lineout, the NMV shows the home team being pushed back towards their defensive half. The analysts can now assess whether the team were positioned correctly in order to defend such an attack.

vancement, however were also interested to know where the backs were positioning when the forwards were in a lineout, ruck or maul. The example shown begins with a lineout being played. In the NMV, the blue and purple stars represent the forwards and backs respectively. The full team NMV is also shown for completion. As the passage of play continues, the NMV shows that the backs maintain wide coverage across the width of the pitch. This provides strong support for when the ball is passed out of the maul by the forwards to the backs who can then drive forward. If the backs are well spread then they are more likely to be able to break through the opposition and score. The passage of play resulted in the home team scoring a try.

In all three scenarios, the analysts found that having multiple tools integrated into the system gave them much more choice and freedom in how to explore the data and draw suitable conclusions on player performance. Whilst the analysts would accept or reject search results as they were displayed, they found that the thumbnail browser provided a more familiar mechanism for exploring the data. This is not surprising since the browser is a much more interactive experience than the search result thumbnails. Also, the purpose of accepting and rejecting results is not necessary for the user's exploration, but for the improvement of the system over time, and so it is a process that is almost transparent to the analyst. From their experience with the system, the professional rugby analysts were positive about the use of visual analytics for sports performance analysis. *"As an analyst, I normally have to watch many matches to splice together a series of clips that show particular trends in our play. By using this system, we could quickly identify moments in the game that correspond with what we are looking for, and so could dramatically reduce our workload"*. The analysts remarked that they liked the different interactions that the system provided, offering a number of possible methods for exploring the video data. *"The visualization of video data helps to reveal information that may be missed from watching video alone. In particular, being able to study and interact with the player motion using the [NMV] visualization helps convey what happened back to the players."*

6 Estimation of Precision and Recall

To supplement the qualitative evaluation in Section 5, it would be desirable to have some quantitative indicators to estimate the benefits of using this visual analytics system. Because the actual users of the system, i.e., sports analysts in Rugby, have much higher level domain knowledge than ordinary participants of a would-be user study, it is generally difficult to conduct a controlled user study for such an application. It is also non-trivial to alleviate various confounding effects (e.g., basic knowledge about rugby games, knowledge about parallel coordinates). We thus conducted a small scale empirical study to observe 3 users (all with reasonable knowledge of both rugby and visualization) in using the system. We adapted the common measures of *time* (T), *precision* (P) and *recall* (R) as quantitative indicators. In content-based information retrieval [28, 2], precision and recall are

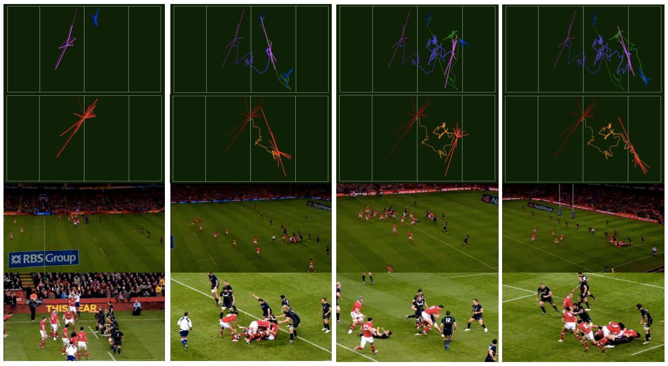

Fig. 11. Four keyframes to illustrate analysing forwards and backs using the NMV (Top-to-bottom: NMV showing forwards and backs, NMV showing full team, wide-angle video keyframe, close-up video keyframe). When the forwards are in the lineout, or in a maul, the analysts can analyse the positioning of the backs in order to cover the pitch whilst support the forwards.

defined as:

$$P = \frac{|C|}{|B|} = \frac{|A \cap B|}{|B|} \quad R = \frac{|C|}{|A|} = \frac{|A \cap B|}{|A|}$$

where given an input query, A is the set all relevant entities (e.g., documents, images, videos, etc.) in a database (including those not retrieved), B is the set of retrieved entities (including those not relevant), and C is the set of correct answers among the retrieved entities. Our empirical study consists of two parts.

The first part focused on comparing a conventional video search approach (finding the relevant video clips by watching videos) and the use of the visual analytics system in an ideal condition. The main stimulus was an 80 minute match video, and a sketch. We manually identified five video clips as the ground truth (each of 1 minute long). We finely tuned our system such that all five video clips are among the top 12 highest scored results, and so appear in the first batch of the the thumbnail display. Our objective for this part was to estimate how long it would take a user to find these five clips in this ideal condition. The test show that on average, the three participants require 2.4 minutes to choose 5 clips from the the 12 thumbnails. The system took less than one minute to return the search results. In this case, the analytical part of the system resulted in $T = 1, P = 5/12, R = 5/5$. The visual analytics process, with combined human and machine reasoning, resulted in $T = 3.4, P = 5/5, R = 5/5$. We then asked participants to estimate how long it would take them to find the 5 clips by watching the video directly with typical video seeking facilities such as fast-forward, pause and rewind. The estimation was between 40-80 minutes. This is equivalent to $40 \leq T \leq 80, P = 5/5, R = 5/5$. This indicates a significant time saving in video search in an ideal condition. We noted from this experiment that participants often made their decision by relying on the thumbnail and the first 10-20 seconds of the video clips. Hence even if the five correct clips were among several multiples of 12 thumbnail options, it would still be quicker than watching the video sequentially.

The second part of our study was designed for identifying the number of interactions that are required to transform the default parameter setting to an optimal parameter setting by using both active learning and model visualization. We counted each decision, accept or reject, as one interaction. We also counted each manual change to a parameter in the model visualization as an interaction. We tested 4 different sketches and found on average it took 14 interactions. This suggested that the dynamic model refinement can benefit search tasks involving several videos, while for a new search task without a pre-trained model, the benefit of dynamic model refinement may not be able to reach its optimal condition with the first 1-2 videos.

Whilst the two studies presented are based on a relatively small set of samples, they offer an informative set of quantitative indicators that demonstrate the benefits of using visual analytics in this application.

7 Discussion

For this application, video search requires the tracking of spatiotemporal components from video data, and a cost-effective method for comparison of the user sketch and the underlying tracking data. Tracking multiple players on a rugby field is a non-trivial task, even with the Spinsight video, due to player occlusion, player contact, and camera distortion. Since the tracking results are likely to have some inaccuracies, the role of visual analytics becomes increasingly important as it allows the user to analyse and verify the generated results, using closely-integrated visual interactive tools.

The system also requires an appropriate method for comparison between the user's input sketch and the video data. We present a model-based approach that consists of multiple similarity metrics that could all potentially yield suitable results. However, to actually define a single similarity measure could result in the system becoming very restrictive, since one measure may not necessarily be suitable for the vast combination of sketches that the user may want to search for. Since it may also be difficult to encapsulate the user's true intentions in the input sketch this becomes even more challenging. By adopting the visual analytics approach the user is equipped with the capability for observing the performance of the model and to correct retrieval errors dynamically. At the same time, the system learns from the user's interaction for accepting and rejecting a retrieved result, improving the accuracy of the model gradually. In addition, the user also has flexibility of modifying the impact that different similarity metrics have on the outcome. For example, the user could choose to increase the impact that direction, curvature and tortuosity have on the similarity measure, and decrease the impact of spatial positioning and mean distance. The results would share similar motion but not necessarily with the same spatial correspondence.

Referring to the task requirement defined in Section 3, we noticed that the most significant benefit is that the user can now perform the tasks in Fig. 1 much quicker than before. Since these tasks are performed routinely, the merit of saving time has indirectly enabled analysts to perform their tasks more accurately as they can examine more videos in each search if they wish to. As search queries vary significantly in their explicit specifications (e.g., sketch, videos), as well as implicit specification (e.g., the type of a match, players involved), we are yet to observe any significant benefit derived from model reuse. We hope that the gradual improvement of the trained models, and the analysts' knowledge about these models will improve the reuse of the trained models.

We should acknowledge that although the feedback from the analysts was very positive, the system is not without its limitations. Player tracking is a non-trivial task and may have some inaccuracies. In particular, when collisions occur between players (e.g., scrums and tackles) it is extremely difficult to perform unique tracking of individuals. For this reason the current system focuses primarily on overall team shape and movement, or team groupings such as forwards and backs, rather than specific individual players. Whilst the analysts express great interest in understanding team shape and movement from the current system, they would certainly welcome the addition of individual player tracking to support more in-depth sketch-based search queries. This could also introduce additional sketch tools such as sequential player actions.

Another potential issue with the proposed system is scalability. Currently, an analyst can search and examine the content of an entire match video. However, they may also want to search and examine content from multiple videos to identify similar movements and techniques that occur between matches. The current system allows for search multiple videos sequentially and individually, thought it could be extended to account for multiple match analysis in a more integrated manner. The search space visualization could be adapted to show either a combined view of all matches, or split on a row-by-row basis to depict the events from each match. Likewise, the parallel co-ordinates view of the similarity model could be adapted so that each polyline depicted a more appropriate time period (e.g., five seconds) rather than showing a polyline for every video frame. We do expect that the number of metrics will increase slightly. However, we do not anticipate any fundamental difficulty in terms of scalability as the analysts' knowledge about different metrics and ability to use parallel coordinates plots will also improve. Nevertheless, it would be helpful to introduce more sophisticated analytical facilities to support analytical tasks based on parallel coordinates plots (e.g., [1, 17, 4, 36, 14]).

8 Conclusion

In this work we have demonstrated the concept of visually searching and analysing video through sketch-based search queries and a visual feedback loop. We present a number of interactive channels that the user can engage with to encourage further exploration of the data, including a model visualization, a search space visualization, search results using NMV thumbnails, and a thumbnail browser that provides linkage back to the original video content. The system adopts active learning through the user interaction, by allowing the user to accept or reject results as determined by the current state of the model. The training model is based on similarity metrics that perform comparison between the user sketch and the video data. The contribution that each similarity metric plays on the overall similarity is determined by a weighting function, that the training model adapts through the learning process. This concept of combining active learning and visual analytics is well-suited to many other applications where vast amount of data is to be processed whilst also facilitating human understanding.

In the past, sports analysts have relied heavily on notational analysis that is laborious to collect and is restrictive to only the occurrence of particular events. Our system allows the analysts to directly search the video content based on the spatiotemporal data of player and team movements during the match. The system provides a closely-integrated suite of tools that the analysts can study in order to gain much deeper exploration of the video data and identify crucial performance indicators. Future work will explore the possibilities of further data integration from different information streams and the use of visual analytics for deeper statistical exploration of player performance.

Acknowledgment

The authors would like to thank the Welsh Rugby Union for their collaboration on this work. Thanks also goes to the Welsh Assembly Government (RIVIC) for funding this research.

References

[1] M. Ankerst, S. Berchtold, and D. Keim. Similarity clustering of dimensions for an enhanced visualization of multidimensional data. In *Information Visualization, 1998. Proceedings. IEEE Symposium on*, pages 52–60, 153, 1998.

[2] R. A. Baeza-Yates and B. Ribeiro-Neto. *Modern Information Retrieval*. Addison-Wesley Longman Publishing Co., Inc., Boston, MA, USA, 1999.

[3] S. Baluja, R. Seth, D. Sivakumar, Y. Jing, J. Yagnik, S. Kumar, D. Ravichandran, and M. Aly. Video suggestion and discovery for youtube: taking random walks through the view graph. In *Proceedings of the 17th international conference on World Wide Web*, WWW '08, pages 895–904, New York, NY, USA, 2008. ACM.

[4] J. Blaas, C. Botha, and F. Post. Extensions of parallel coordinates for interactive exploration of large multi-timepoint data sets. *Visualization and Computer Graphics, IEEE Transactions on*, 14(6):1436–1451, 2008.

[5] R. Borgo, M. Chen, B. Daubney, E. Grundy, G. Heidemann, B. Hoferlin, M. Hoferlin, H. Leitte, D. Weiskopf, and X. Xie. State of the art report on video-based graphics and video visualization. *Computer Graphics Forum*, 31(8):2450–2477, 2012.

[6] R. Botchen, F. Schick, and T. Ertl. Action-based multifield video visualization. *Visualization and Computer Graphics, IEEE Transactions on*, 14(4):885 –899, july-aug. 2008.

[7] I. Bowman, S. Joshi, and J. Van Horn. Query-based coordinated multiple views with feature similarity space for visual analysis of mri repositories. In *Visual Analytics Science and Technology (VAST), 2011 IEEE Conference on*, pages 267–268, 2011.

[8] S.-F. Chang, W. Chen, H. J. Meng, H. Sundaram, and D. Zhong. Videoq: an automated content based video search system using visual cues. In *Proceedings of the fifth ACM international conference on Multimedia*, MULTIMEDIA '97, pages 313–324, New York, NY, USA, 1997. ACM.

[9] S.-F. Chang, W. Chen, H. J. Meng, H. Sundaram, and D. Zhong. A fully automated content-based video search engine supporting spatiotemporal queries. *IEEE Transactions on Circuits and Systems for Video Technology*, 8:602–615, 1998.

[10] S.-F. Chang, W.-Y. Ma, and A. Smeulders. Recent advances and challenges of semantic image/video search. In *Acoustics, Speech and Signal Processing, 2007. ICASSP 2007. IEEE International Conference on*, April 2007.

[11] M. Chen, R. Hashim, R. Botchen, D. Weiskopf, T. Ertl, and I. Thornton. Visual signatures in video visualization. *Visualization and Computer Graphics, IEEE Transactions on*, 12(5):1093 –1100, sept.-oct. 2006.

[12] W. Chen and S.-F. Chang. Vismap: an interactive image/video retrieval system using visualization and concept maps. In *ICIP (3)*, pages 588–591, 2001.

[13] G. Daniel and M. Chen. Video visualization. In *Visualization, 2003. VIS 2003. IEEE*, pages 409 –416, oct. 2003.

[14] A. Dasgupta and R. Kosara. Pargnostics: Screen-space metrics for parallel coordinates. *Visualization and Computer Graphics, IEEE Transactions on*, 16(6):1017–1026, 2010.

[15] Y. Deng and B. S. Manjunath. Content-based search of video using color, texture, and motion. In *Proceedings. International Conference on Image Processing (cat. No.97cb36144)*. IEEE Signal Process. Soc, IEEE Signal Process. Soc, Jan 1997.

[16] N. Dimitrova, H.-J. Zhang, B. Shahraray, I. Sezan, T. Huang, and A. Zakhor. Applications of video-content analysis and retrieval. *IEEE MultiMedia*, 9(3):42–55, July 2002.

[17] G. Ellis and A. Dix. Enabling automatic clutter reduction in parallel coordinate plots. *IEEE Transactions on Visualization and Computer Graphics*, 12(5):717–724, Sept. 2006.

[18] M. Hao, U. Dayal, D. Keim, D. Morent, and J. Schneidewind. Intelligent visual analytics queries. In *Visual Analytics Science and Technology, 2007. VAST 2007. IEEE Symposium on*, pages 91–98, 2007.

[19] A. G. Hauptmann, W.-H. Lin, R. Yan, J. Yang, and M.-Y. Chen. Extreme video retrieval: joint maximization of human and computer performance. In *Proceedings of the 14th annual ACM international conference on Multimedia*, MULTIMEDIA '06, pages 385–394, New York, NY, USA, 2006. ACM.

[20] M. Hoferlin, B. Hoferlin, G. Heidemann, and D. Weiskopf. Interactive schematic summaries for faceted exploration of surveillance video. *Multimedia, IEEE Transactions on*, PP(99):1, 2013.

[21] W. H. Hsu, L. S. Kennedy, and S.-F. Chang. Reranking methods for visual search. *IEEE MultiMedia*, 14(3):14–22, July 2007.

[22] M. D. Hughes and I. M. Franks. *Notational analysis of sport*. London: E. & F.N. Spon., 1997.

[23] J. Lee and T. Funkhouser. Sketch-based search and composition of 3D models. In *EUROGRAPHICS Workshop on Sketch-Based Interfaces and Modeling*, June 2008.

[24] P. A. Legg, D. H. S. Chung, M. L. Parry, M. W. Jones, R. Long, I. W. Griffiths, and M. Chen. Matchpad: Interactive glyph-based visualization for real-time sports performance analysis. *Computer Graphics Forum*, 31(3pt4):1255–1264, 2012.

[25] Y. Liu, T. Mei, X.-S. Hua, J. Tang, X. Wu, and S. Li. Learning to video search rerank via pseudo preference feedback. In *Multimedia and Expo, 2008 IEEE International Conference on*, pages 297–300, 2008.

[26] T. McLoughlin, M. Jones, R. Laramee, R. Malki, I. Masters, and C. Hansen. Similarity measures for enhancing interactive streamline seeding. *Visualization and Computer Graphics, IEEE Transactions on*, PP(99):1–1, 2012.

[27] M. Parry, P. Legg, D. Chung, I. Griffiths, and M. Chen. Hierarchical event selection for video storyboards with a case study on snooker video visualization. *Visualization and Computer Graphics, IEEE Transactions on*, 17(12):1747 –1756, dec. 2011.

[28] C. J. V. Rijsbergen. *Information Retrieval*. Butterworth-Heinemann, Newton, MA, USA, 2nd edition, 1979.

[29] J. Rodrigues, L. Romani, A. J. M. Traina, and C. Traina. Combining visual analytics and content based data retrieval technology for efficient data analysis. In *Information Visualisation (IV), 2010 14th International Conference*, pages 61–67, 2010.

[30] K. Schoeffmann, F. Hopfgartner, O. Marques, L. Boeszoermenyi, and

J. M. Jose. Video browsing interfaces and applications: a review. *Journal of Photonics for Energy*, pages 018004–018004–35, 2010.

[31] J. Sivic. *Efficient Visual Search of Images and Videos*. PhD thesis, University of Oxford, 2006.

[32] J. Sivic and A. Zisserman. Video Google: Efficient visual search of videos. In J. Ponce, M. Hebert, C. Schmid, and A. Zisserman, editors, *Toward Category-Level Object Recognition*, volume 4170 of *LNCS*, pages 127–144. Springer, 2006.

[33] J. Sivic and A. Zisserman. Efficient visual search for objects in videos. *Proceedings of the IEEE*, 96(4):548–566, 2008.

[34] C. G. M. Snoek, B. Huurnink, L. Hollink, M. de Rijke, G. Schreiber, and M. Worring. Adding semantics to detectors for video retrieval. *Multimedia, IEEE Transactions on*, Aug. 2007.

[35] C. G. M. Snoek, M. Worring, J. van Gemert, J.-M. Geusebroek, D. Koelma, G. P. Nguyen, O. de Rooij, and F. Seinstra. Mediamill: exploring news video archives based on learned semantics. In *Proceedings of the 13th annual ACM international conference on Multimedia*, MULTIMEDIA '05, pages 225–226, New York, NY, USA, 2005. ACM.

[36] A. Tatu, G. Albuquerque, M. Eisemann, J. Schneidewind, H. Theisel, M. Magnor, and D. Keim. Combining automated analysis and visualization techniques for effective exploration of high-dimensional data. In *Visual Analytics Science and Technology, 2009. VAST 2009. IEEE Symposium on*, pages 59–66, 2009.

[37] J. Tešić, A. P. Natsev, and J. R. Smith. Cluster-based data modeling for semantic video search. In *Proceedings of the 6th ACM international conference on Image and video retrieval*, CIVR '07, pages 595–602, New York, NY, USA, 2007. ACM.

[38] S. Vrochidis, I. Kompatsiaris, and I. Patras. Optimizing visual search with implicit user feedback in interactive video retrieval. In *Proceedings of the ACM International Conference on Image and Video Retrieval*, CIVR '10, pages 274–281, New York, NY, USA, 2010. ACM.

[39] X. Wang, C. Zhang, and Z. Zhang. Boosted multi-task learning for face verification with applications to web image and video search. In *Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on*, pages 142–149, June 2009.

[40] L. Zhang, F. Lin, and B. Zhang. Support vector machine learning for image retrieval. In *Image Processing, 2001. Proceedings. 2001 International Conference on*, volume 2, pages 721–724 vol.2, Oct.

[41] D. Zhong and S.-F. Chang. Spatio-temporal video search using the object based video representation. In *Proceedings of the 1997 International Conference on Image Processing (ICIP '97) 3-Volume Set-Volume 1 - Volume 1*, ICIP '97, pages 21–, Washington, DC, USA, 1997. IEEE Computer Society.

[42] D. Zhong and S. fu Chang. Structure analysis of sports video using domain models. *IEEE ICME*, pages 22–25, 2001.