



Extending the attribute explorer to support professional team-sport analysis

Pär-Anders Albinsson
Dennis Andersson

*Division of Command and Control Systems,
Swedish Defence Research Agency, Linköping,
Sweden*

Correspondence:
Pär-Anders Albinsson, Division of Command
and Control Systems, Swedish Defence
Research Agency, P.O. Box 1165,
SE-581 11 Linköping, Sweden.
Tel: +46 13378346, +46 734447746;
Fax: +46 13 378550;
E-mail: paalb@foi.se

Abstract

Advances in interactive systems and the ability to manage increasing amounts of high-dimensional data provide new opportunities in numerous domains. Information visualization techniques are especially useful in situations where analysts seek patterns and information of interest in massive data sets. In this article, we propose an extension of the original Attribute Explorer (AE) technique by Spence and colleagues to take on the challenges presented in the domain of professional team-sport analysis. We describe the implementation of an extended AE and use football game-event data to highlight the new possibilities.

Information Visualization (2008) 7, 163–169. doi:10.1057/palgrave.ivs.9500178

Keywords: Attribute Explorer; team-sport analysis

Introduction

Analysing team-sport data, an increasingly focal activity in professional sports^{1,2} is accompanied with typical issues tackled in the information visualization (IV) discipline. In this article, we extend an existing IV technique in the design of a tool for supporting analysis of football game events.

Identifying and analysing game events in team sports is a common approach in the efforts to improve the quality of professional teams. Recent advances in automatic video analysis promise faster and less labour-intensive access to large amounts of sport-event data. However, with increasing amounts of high-dimensional data comes a challenge in data exploration. Dealing with complex data sets, analysts may need support not only to *answer* their questions but also to *discover* them.

The Attribute Explorer (AE), an interactive data-presentation tool developed by Spence and colleagues, has shown promising exploration capabilities for large multivariate data sets.³ Noting the similarities between sport analysis and earlier successful applications of the AE, we initiated a design process to develop a tool based on the previous work. During this process we identified and implemented several extensions to the AE pertaining to the domain of game-event analysis. In this article, we present the resulting tool and discuss the extensions by exemplifying them on data collected from a video recording of an international football friendly. While football was chosen for this demonstration it is the authors' intent to keep the discussion general enough to suit other complex team sports, such as American football, basketball, rugby and hockey.

Professional team-sport analysis

Professional team sports, such as football, have turned into a highly competitive and lucrative global business. Revenue is critical, as for all businesses,

Received: 9 July 2007
Revised: 13 March 2008
Accepted: 13 March 2008
Online publication date: 17 April 2008

and primarily generated by getting and keeping supporters and sponsors. Professional teams focus on achieving good results to maintain the supporter crowd and to reach better sponsor deals. Also, the increasingly tougher competition forces teams to find new and improved ways of optimizing performance to sustain good results. Several major clubs have full-fledged research labs for this purpose.

One of the research initiatives concerns observing matches and collecting game events for later analysis.¹ Events such as passes, shots, interceptions and ball runs are time-stamped and given certain attributes, such as actors involved, location, quality and distance. Statistical methods can then be used to investigate how events and series of events relate to quality, performance and other concepts. Sport physiologist Paul Balsom believes that knowledge about player movement patterns is crucial for developing effective training methods.⁴ Human observation and memory limitations reduce the usefulness of collecting these data during the actual game.⁵ Instead, video recordings provide a more manageable environment for identification and structuring of events after the games. Still, going through video feeds manually is time-consuming even when supported by computer systems.¹ However, considerable research efforts are ongoing on automatic identification of game events, both based on live video feeds and on various positioning devices.^{6–9} Combining manual and automatic approaches – skilled observers and well-tuned algorithms – will lead to faster access to larger amounts high-quality data.

When databases grow and their data elements get increasingly detailed, new challenges arise. Rather than on difficulties of collecting and storing data, the issues will pertain to getting an overview of data, navigating among data and finding patterns of significance in the data. This problem is particularly true in the relatively young domain of computer-supported football analysis where much of the complex relations among game-event data are still to be understood: “The problem for the sports analyst is to locate the dynamic patterns within these data that give rise to the invariant features that exist in sports contests” (McGarry and Franks,² p. 271). If one can define the questions on beforehand, statistical methods usually do the trick; otherwise, one needs help to formulate the questions as well. The current manager of the Swedish national football team, Lars Lagerbäck, anticipates that the growing data sets and the technological advances will provide better context possibilities that enable identification of cause-and-effect relationships.¹⁰

The Attribute Explorer

Within the field of IV a central theme pertains to designing representations of the reality that are as informative as possible with regard to the users. When dealing with complex data, their representation is decisive: “... solving a problem simply means representing it so as to make the solution transparent” (Simon,¹¹ p. 132). The IV

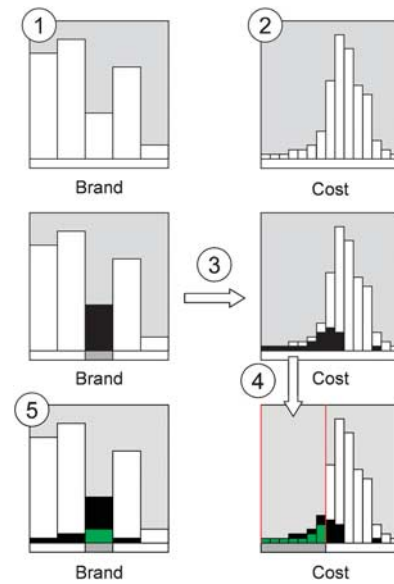


Figure 1 A simple example of the use of the AE. The example shows a two-dimensional data set, the brand and cost of cars, to explain the basic principles of the AE. See the article text for explanation.

literature shows a multitude of innovative techniques to represent and interact with large multidimensional data sets, and the possibilities increase with the rapid development in computer graphics and interactive systems.³

One technique that shows promising potential for “acquisition of insight into multivariate data” is the AE (Spence,³ p. 62). Originally, the AE aimed to support the selection of one object from many on the basis of its attributes,^{12,13} but has since then also been employed for more exploratory uses. One such example includes the Swedish Defence Research Agency’s use of the AE as an investigative tool into communication data (Albinsson and Morin,¹⁴ Morin and Albinsson,¹⁵ Spence,³ pp. 216–220).

The AE (Figure 1) is an interactive data presentation tool that employs the concept of *linked histograms*. There is one histogram for each dimension (or attribute) of a data set, where the height of each bar corresponds to the number of data elements that fall under that interval. When applying constraints in one dimension, the corresponding changes are displayed in the other dimensions. Data elements presented in green represent *full hits* – that is, they satisfy all constraints applied in all dimensions. Shades of grey, from black to white, represent the number of failing dimensions; a black element fails in only one dimension and a white in all.

Consider the example in Figure 1. Imagine that we want to find a car to buy from a set of many by using the AE. In this case the car data set has two dimensions: *Brand* and *Cost* of car; consequently, the AE provides two histograms. The first dimension (1) shows five brands of cars, whereas the other (2) shows the price range. There

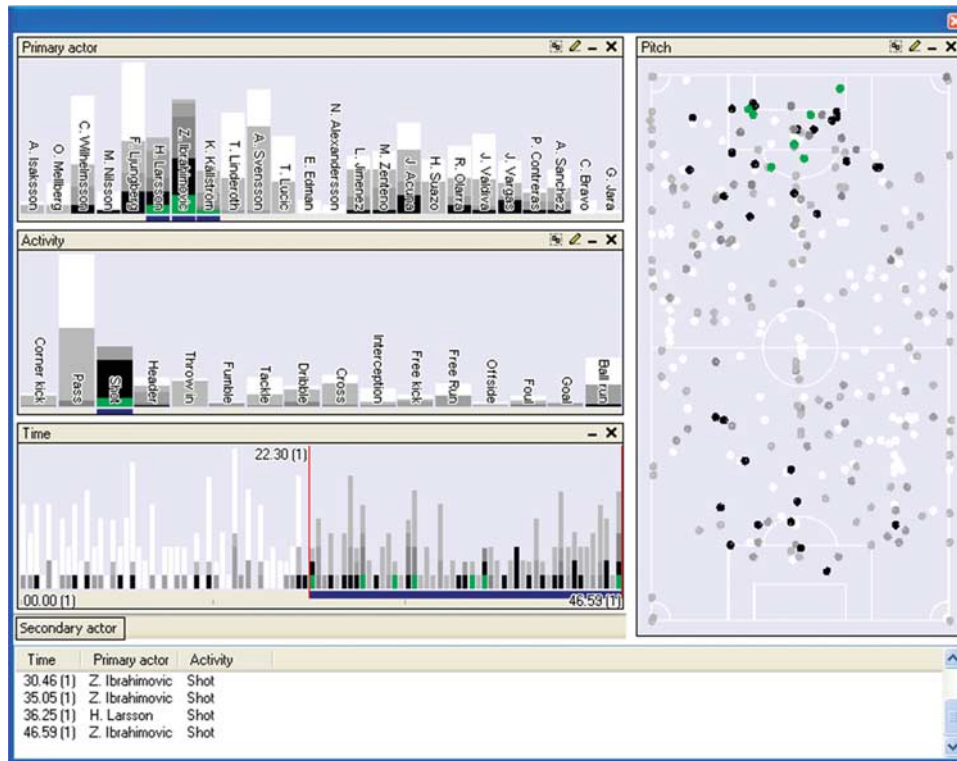


Figure 2 A screenshot of the football AE. Three dimensions are shown to the left (primary actor, activity and time) and the pitch dimension to the right. A minimized dimension (secondary actor) is visible in the below left. Bottommost, a list presents the current hits (coloured green): shots by three players during the second half of the time period.

are some cheap cars, some expensive ones and many in between. Initially, all histograms are white, since there are no selections and, thus, all elements fail in all dimensions. Changing the constraints for one dimension, by selecting a certain brand, causes a part of the histogram bar to turn black, which indicates that the data elements (i.e., the cars) of that selection only fail in one dimension. This change also appears in the cost dimension (3), where the corresponding elements turn black. Looking at these black elements, one notices that most cars of the selected brand are relatively inexpensive. Say that we are looking for a cheap car and therefore apply constraints to the price dimension accordingly (4). The black elements within the range are going to be full hits (and turn green), because they only failed in this dimension previously. The green elements represent cars of the selected brand within our acceptable price range. Black elements under these constraints represent *other* cheap brands. Finally, we switch to the brand dimension again (5) and notice the corresponding green hits. Here, the black elements (failing in the cost dimension) represent the remainder of the cheap cars of other brands.

A comprehensive description of the original AE cannot be covered in this article; instead unfamiliar readers are encouraged to look into the works of Spence and colleagues.^{3,12,13}

Extending the AE

Ongoing research on transferring military-training computer-support tools – where the AE is used for communication analysis as one component of several – to the domain of professional football¹⁶ inspired us to take a closer look at the specific possibilities of the AE.

Encouraged by the successful use of the AE in the military domain¹⁴ and from discussions with professional football practitioners and experts we started an iterative design process based on the challenges expressed in current sport research as summarized earlier. During the design process we identified several new needs concerning the AE and we implemented solutions for further investigation using realistic football event data gathered by stepwise going through a video recording of an international football game (Figure 2). The following sections comprise our extension of the original AE technique to support team-sport analysis.

Distributed attribute values

The original AE concept¹² mainly concerns data elements that have one value per attribute. That is, one single data element is represented by one single element in one single bar for each dimension. In our case, the nature of game events forced us to allow a data element to

be distributed in several bars of the same attribute. For example, a single game event can have multiple values in the activity dimension, such as a *pass* that is also a *header* or a *shot* that is also a *free kick*. Similarly, other attributes such as actor sometimes need multiple values.

The design decisions that arise from allowing distributed attribute values include how to deal with the colour coding. When a bar that contains elements with distributed values is selected in a dimension, should the other bars that contain the same element in the same dimension be colour coded similarly? A negative effect from such a solution is that bars outside current selections still can be presented as full hits. For example, if an analyst has selected all passes as full hits, every event in the *header* bar that also is a *pass* will turn green even without the user selecting the *header* bar, thereby breaking the conventional colour coding. Analogously, black elements (normally one-limit failed) no longer ensure the user that selecting it will produce a full hit.

An unappealing alternative involves requiring all distributed attribute values within a dimension to be selected to result in a full hit. This solution causes obscure behaviour since it nullifies the established colour coding. Selecting distributed attribute values would not update the colour-coding level until all values were selected. Selecting *pass* in the activity dimension would not include game events that also are *header* or *volley*, making the use cumbersome and unfamiliar.

Instead, after considering different compromises, we chose to design the colour coding to account for distributed attribute values. An element in an unselected bar which is already selected in another bar within the same dimension reduces its colour-coding value by 1. Thereby, a full-hit element in a selected bar will be presented in green whereas unselected bars containing the same element will use black colouring. In this way, black still represents one failed limit (see e.g. the black elements in the *corner kick* bar, in Figure 3, which also are *passes*).

Excluding selections and attribute duplication

Traditionally, the AE uses inclusive selections. That is, one selects ranges that represent what the elements should fulfil. In our case, as a consequence of the previous issue of distributed attribute values, selections need to support *exclusions* as well. Consider a *header* game event that also has the value *pass*. If one selects *pass* to investigate all passes during a match, the previous *header* event would fall under this selection since it also has the value *pass*. Since there are situations where it is of interest to rule out, for example, all *passes* that are *headers*, we added excluding-selection functionality to the AE. By using a modifier key, a bar can be marked as an excluding selection, indicated by a red colour below the bar rather than the blue for normal selection (Figure 3). In the previous example, if the *header* bar is marked as an excluding selection, only *passes* without the additional value *header* would become full hits.

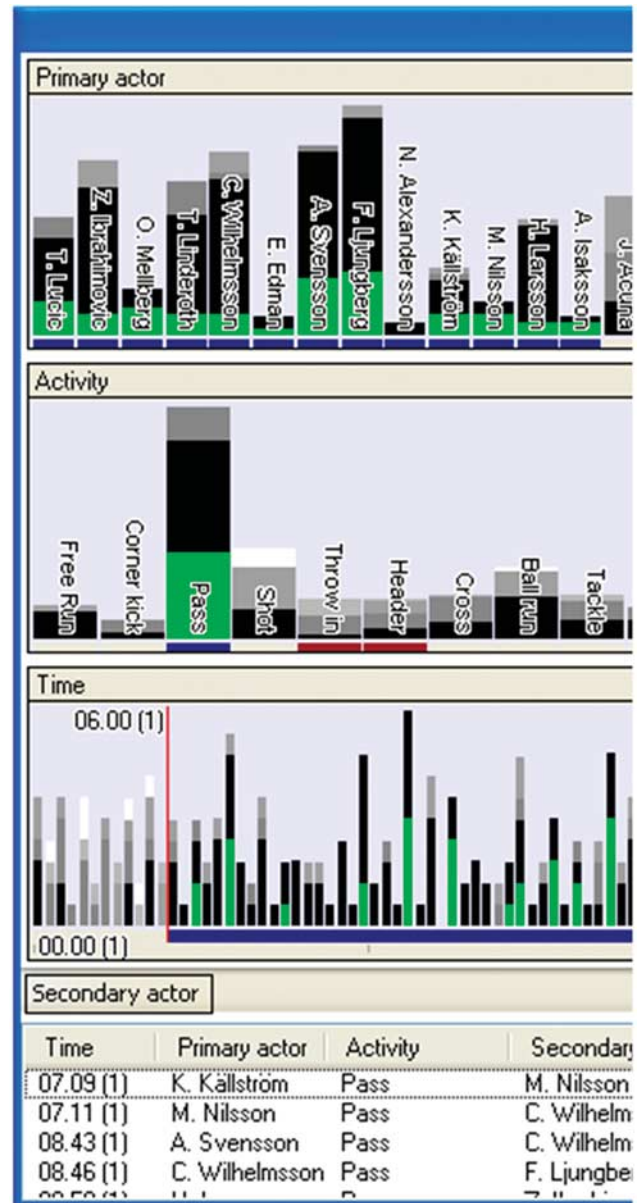


Figure 3 Excluding selections. A user has selected *pass* in the activity dimension and marked *header* and *throw in* as excluding selections. The result presents passes that are not headers or throw ins.

Excluding selections allow users to apply logical *not* operations in addition to the conventional *or* operations within a dimension. Owing to the distributed attribute values the need arises for logical *and* operations within a dimension. We may, for example, seek information regarding *passes* that also are *headers*. But selecting both event types in the activity dimension results in a logical *or*. To enable *and* operations we add the function of *attribute duplication*. A new identical view for the same dimension is (temporarily) added to the AE

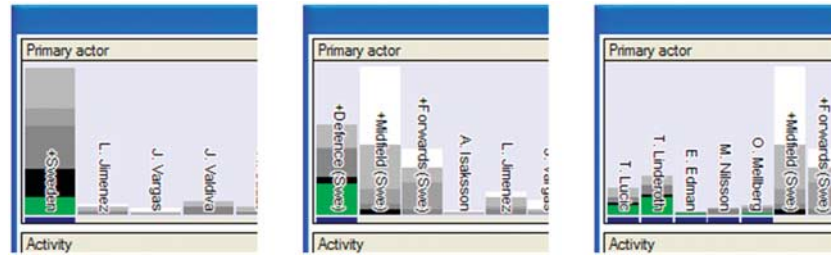


Figure 4 Dynamic aggregation. To the left we see an aggregated bar containing all players of the Swedish team. The result from expanding this bar is shown in the middle where different roles now are visible. By selecting one of the roles and expanding further, individual players appear as shown to the right.

where user selections are modified independently of the original dimension. The duplicate dimension behaves exactly like a normal AE dimension. By selecting *passes* in the original dimension and *headers* in the duplicate, the result comprise passes carried out as headers.

Heterogeneous data sets

As described earlier, game events have a large number of potential attributes. Managing numerous attributes is not a problem for the AE concept. In the game-event case however, there is a need to work concurrently with data elements represented in varying subsets of all dimensions. For example, a *shot* event may have only two dimensions, primary actor and activity, while a *pass* event is represented in the secondary-actor dimension as well (the receiver of the pass). Analogously, *interception* events are not applicable for a *length* dimension.

Allowing such heterogeneous data sets forces us to take some design decisions to not end up with ambiguous colour coding in the AE. For example, when selecting the activity *pass*, and two particular players as primary and secondary actors the result would present all the passes between them as green hits. However, using the traditional AE concept, any *shot* event by the selected players would be presented as black elements since they only fail in one dimension. This behaviour is deceptive since one may be misled to incorrectly believe that the *shot* activities are represented in the secondary-actor dimension.

Instead we introduce the concept of *forced inclusion*. Forced inclusion ensures that, once users select elements in one dimension, only elements that are represented in this dimension are considered within the global selection boundaries. This behaviour affects the previous example by, as soon as a secondary actor is selected, preventing *shot* events to get closer to a hit.

In the case where a user aims to focus only on a subset of the available dimensions of a data set, we added the possibility to exclude dimensions and minimize their views (as shown in Figure 3). In the example with *pass* and *shot* events, excluding the secondary-actor dimension will treat the pass events as if they were not represented in

this dimension; in other words, the AE will disregard the excluded dimension for all calculations.

Dynamic aggregation of attribute values

In our earlier work on communication data exploration, we recognized the need for hierarchical dimensions in the AE.¹⁴ The tree structure used in that work was static, however, and could not be rearranged or modified interactively during exploration. For game-event data, the need for dynamics is greater. The player dimensions, for example, are highly organized. On the leaf level there are individual players, which belong to collections like centre back, fullback, centre midfielder, wing midfielder, which in turn belongs to defence and midfield. On the highest levels we find entire teams or national squads. Obviously, it would be unpractical to not allow analysts to arrange, group and expand players interactively during exploration. This dynamism is especially important when there are large amounts of bars in a dimension (imagine, e.g., the player dimensions in a data set covering game events from a whole season).

Therefore, we designed a general grouping functionality for all discrete dimensions (Figure 4). Users can create unlimited levels of groupings, move bars around and expand, and contract defined groups. When the analyst selects a number of bars, and then uses the group command, the selected bars are visually hidden and a new virtual bar is created (named by the user), which contains the elements from all previous bars. Any selection operation performed on the new bar affects the all aggregated elements from the previous bars.

Two-dimensional attributes

The original AE presented the possibility to use a geographical map dimension in the case of a home-finding application.¹² Our idea is to treat the map dimension analogously to the traditional dimensions, and allow both colour-coded presentation of elements and brushed selection possibilities (Figure 5).

The map dimension – the pitch in this case – is central in team-sport analysis. Putting a lot of focus on the pitch,

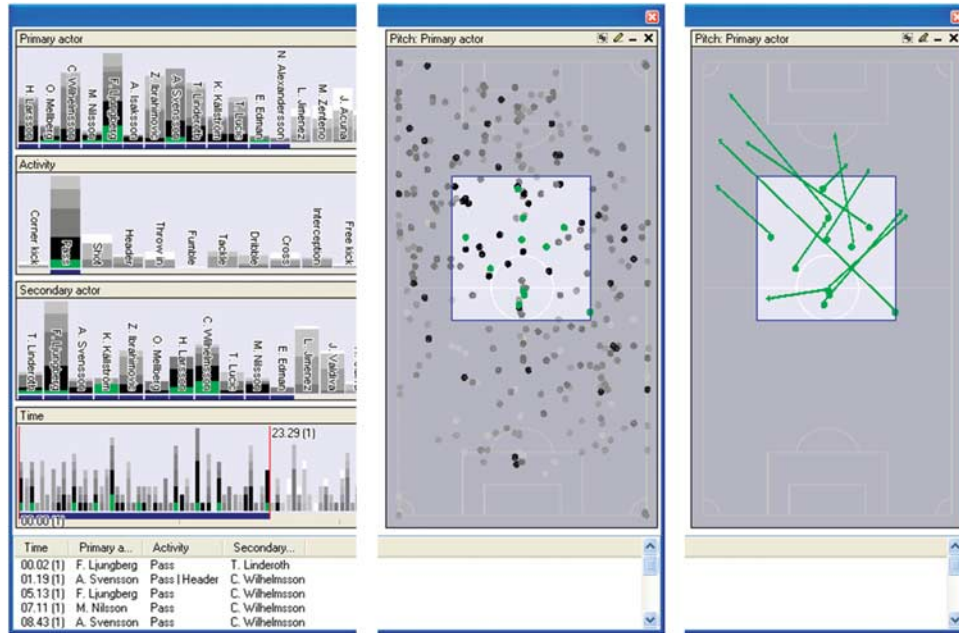


Figure 5 The pitch dimension in action. To the left is the current selection of events in four dimensions. Additionally, the user has selected a part of the upper midfield as shown in the pitch dimension in the middle. The pitch dimension is linked to *primary actor* (as indicated in the pitch title bar). Consequently, the hits comprise passes within the Swedish team originating from the upper midfield. The pitch dimension to the right exemplifies another configuration of the same view where pass vectors are shown for each event and where only the hits are presented.

we found that for game-event data, this issue is somewhat more complex than just providing one more dimension for the game events. Instead, *several* of the game-event attributes can be positioned on the map simultaneously. For example, for the game event “player 1 passes the ball to player 2”, player 1 has a position on the map, the pass a vector, and player 2 another position. In this case, a single event comprises three objects in the pitch dimension. To handle the presentation of these multiple map objects, we added the possibility to interactively choose which game-event attributes to show in the pitch dimension (e.g. showing only positions of primary actors, or the vectors and polygons of activities). Furthermore, to avoid clutter when showing many attributes simultaneously, we added the possibility to only show the events closest to a hit (Figure 5).

Because of the multiple game-event attributes in the pitch dimension, we also need to define how selections in the map affect them. For example, picture an analyst that configures the pitch to show positions of primary and secondary actors and the vectors of their activities and decides to look more closely on midfield pass receives. If the analyst selects the appropriate midfield section in the pitch view, she would like that selection to apply only for the receivers’ (secondary actors’) positions to get the desired selection of events. However, if the analyst instead gets interested in passes *from* the midfield she would like the current map selection to apply for the *primary actors’*

positions. We designed the map view to support these actions by allowing the user to dynamically select which attribute to couple to the map.

As a consequence of this solution, multiple pitch dimensions may be used simultaneously to tackle more advanced queries. For example, the analyst may use one pitch dimension to mark *primary actors’* positions in the upper midfield area, and a second pitch dimension to mark *secondary actors’* positions in the penalty area. In this way, the analyst can investigate passes *from* a certain region *to* a certain region. Similar to the other AE dimensions, the pitch view supports excluding selections, allowing the analyst to look into events that did *not* take place in a certain area.

Coordination to other views

Since there are large numbers of existing tools and techniques for team-sport analysis we view the AE as a component to be coupled to other components, similarly to the earlier work on communication analysis.¹⁵ Video analysis, for example, is an important part of team-sport analysis. By adding a synchronized video view to the AE, an analyst can quickly playback a specified sequence of the game by selecting an event from the list of hits. The video view will then use the timestamp of that event as an offset into the video feed and start playback from that position. If multiple video feeds are available

for the topical event, several views will be opened for playback.

There are of course a multitude of other potential analysis components to be coupled to the AE. A range of possibilities concerning football analysis are discussed by Albinsson and Andersson.¹⁶

Discussion

All data used in this initial investigation of AE extensions have been manually entered by the authors themselves from analysing a television broadcasted game using general protocols described in current literature.^{1,2} A natural next step is to interview football analysts to find out in more detail what information and patterns are relevant and let them use the technique on a more extensive set of data. Testing the technique on other sports would also be important to identify the added value of the proposed extensions to the AE and to investigate possibilities for further development.

Even though the team-sport analysis domain motivated our work on extending the AE, we believe that the extensions are general enough to suit other domains where large amounts of temporal and spatial multidimensional data are generated. Examples include such diverse fields as military intelligence analysis, forest-harvesting planning, GPS-positioned digital photo libraries, and of course other sports.

We have presented possibilities and challenges of introducing the AE for team-sport analysis. To summarize, the challenges were confronted by designing the following extensions to the original concept:

- *Allowing distributed attribute values:* A single data element may be represented by multiple values in the same dimension.
- *Excluding selections and attribute duplication:* Users may use logical *and*, *not* and *or* operations.
- *Allowing heterogeneous data sets:* Data elements are allowed to have varying number of dimensions.
- *Dynamic aggregation of attribute values:* Data elements may be dynamically structured hierarchically.
- *Two-dimensional attributes:* Data elements are represented and interacted with in a geographical dimension.
- *Coordination to other views:* Resulting hits in the AE view may control other views.

A general concern pertains to whether these additions to the AE result in a too complex concept for the proposed use. One may argue that once the necessary data are collected, well-known statistical methods can be applied to give an answer to the questions of the analyst. This argument is especially valid in cases where the questions

and problems are well defined. From our review of the literature, however, questions are often ill defined in the complex domain of team-sport analysis, much like in the domain of command-and-control analysis in our earlier work with the AE.¹⁵ We believe that, once provided to analysis experts, visualization techniques like the one presented will prove to be useful in understanding game-event data enough to derive new metrics that adequately capture their complexity; that is, finding questions to ask.

Acknowledgements

We thank Robert Spence for valuable comments on an early draft of this article.

References

- 1 Hughes M. Notational analysis. In: Reilly T, Williams AM (Eds). *Science and Soccer*. 2nd edn., Routledge: London, 2003; 245–264.
- 2 McGarry T, Franks I. The science of match analysis. In: Reilly T, Williams AM (Eds). *Science and Soccer*. 2nd edn., Routledge: London, 2003; 265–275.
- 3 Spence R. *Information Visualization: Design for Interaction*. 2nd edn. Pearson: Harlow, 2007.
- 4 Balsom P. *Fotbollens träningslära*. Svenska FotbollFörlaget AB: Uppsala, Sweden, 2003.
- 5 Franks I, Miller G. Eyewitness testimony in sport. *Journal of Sport Behavior* 1986; **9**: 39–45.
- 6 Assfalg J, Bertini M, Colombo C, Del Bimbo A, Nunziati W. Semantic annotation of soccer videos: automatic highlights identification. *Computer Vision and Image Understanding* 2003; **92**: 285–305.
- 7 Ekin A, Murat Tekalp A, Mehrotra R. Automatic soccer video analysis and summarization. *IEEE Transactions on Image Process* 2003; **12**(7): 796–807.
- 8 Nitta N. Semantic content analysis of broadcasted sports videos with intermodal collaboration. PhD Thesis: Osaka University, Japan, 2003.
- 9 Yu X, Yan X, Hay TS, Leong HW. 3D reconstruction and enrichment of broadcast soccer video. *12th ACM Conference on Multimedia* (New York, NY), ACM Press: New York, 2004; 260–263.
- 10 Hedlund C, Landin S. GPS som tekniskt hjälpmedel inom fotbollen. Master Thesis, Högskolan Dalarna, Sweden, 2005.
- 11 Simon H. *The Sciences of the Artificial*. 3rd edn. MIT Press: Cambridge, MA, 1996.
- 12 Spence R, Tweedie L. The attribute explorer: information synthesis via exploration. *Interacting with Computers* 1998; **11**: 137–146.
- 13 Tweedie L, Spence R, Williams D, Bhogal R. The attribute explorer. *ACM Conference on Computer-Human Interaction* (Boston, MA), ACM Press: New York, 1994; 435–436.
- 14 Albinsson P-A, Morin M. Visual exploration of communication in command and control. *Sixth International Conference on Information Visualisation* (London, UK), IEEE: Los Alamitos, 2002; 141–146.
- 15 Morin M, Albinsson P-A. Exploration and context in communication analysis. In: Bowers C, Salas E, Jentsch F (Eds). *Creating High-Tech Teams: Practical Guidance on Work Performance and Technology*. APA Press: Washington DC, 2005; 89–112.
- 16 Albinsson P-A, Andersson D. Computer-aided football training: exploiting advances in distributed tactical operations research. *Sixth International Conference of the International Sports Engineering Association* (Munich, Germany), Springer: New York, 2006; 185–190.