Original Article

Visual data mining of multimedia data for social and behavioral studies

Chen Yu[a,*],
Yiwen Zhong[a,b],
Thomas Smith[a],
Ikhyun Park[a] and
Weixia Huang[a]

[a]Indiana University, 1101 East 10th Street, Bloomington, Indiana 47405, USA;
[b]Fujian Agriculture and Forestry University, Fuzhou, Fujian 350002, China.
E-mail: chenyu@indiana.edu

*Corresponding author.

Abstract With advances in computing techniques, a large amount of high-resolution high-quality multimedia data (video and audio, and so on) has been collected in research laboratories in various scientific disciplines, particularly in cognitive and behavioral studies. How to automatically and effectively discover new knowledge from rich multimedia data poses a compelling challenge because most state-of-the-art data mining techniques can only search and extract pre-defined patterns or knowledge from complex heterogeneous data. In light of this challenge, we propose a hybrid approach that allows scientists to use data mining as a first pass, and then forms a closed loop of visual analysis of current results followed by more data mining work inspired by visualization, the results of which can be in turn visualized and lead to the next round of visual exploration and analysis. In this way, new insights and hypotheses gleaned from the raw data and the current level of analysis can contribute to further analysis. As a first step toward this goal, we implement a visualization system with three critical components: (1) a smooth interface between visualization and data mining; (2) a flexible tool to explore and query temporal data derived from raw multimedia data; and (3) a seamless interface between raw multimedia data and derived data. We have developed various ways to visualize both temporal correlations and statistics of multiple derived variables as well as conditional and high-order statistics. Our visualization tool allows users to explore, compare and analyze multi-stream derived variables and simultaneously switch to access raw multimedia data.
Information Visualization (2009) **8,** 56–70. doi:10.1057/ivs.2008.32; published online 12 February 2009

Keywords: visualization; data mining; multimedia; time series; temporal data; pattern discovery

Introduction

With advances in computing and sensing techniques, multimedia data are ubiquitous. In particular, a large amount of high-resolution high-quality multimedia data (video, audio, EEG and fMRI, and so on) has been collected in research laboratories in various scientific disciplines, especially in social, behavioral and cognitive studies. How to automatically and effectively discover new knowledge from rich multimedia data poses a compelling challenge. Multimedia data mining in general consists of two stages. In the first step, researchers extract some derived data from raw multimedia data. This step can be implemented by human coding or by using image/speech processing programs. For example, one may extract a time series of the location of a participant in a video clip to form one temporal continuous variable. However, it is very difficult to extract and formalize all the relevant data from a video. Each video frame may contain several objects and people. One can extract different properties from each entity (an object or a person) in an image frame, such as the location of the entity, the size of the entity, the speed of the entity, or the color change of the entity. Assume

Received: 17 November 2008
Accepted: 17 November 2008

that there are M objects and N people appearing in a video clip, and K properties can be extracted from each, we will have $(M + N)K$ temporal continuous variables. With the increase of M, N and K, extracting all the information from a video clip seems not to be efficient and may not be possible. For another example, speech recordings seem to be relatively simple because we can code them (automatically or manually) as speech transcription. However, in addition to linguistic content, raw speech data also contain other information that may also be useful, such as a speaker's gender and age, the speaker's emotional state encoded in speech prosody, and so on. Moreover, researchers may be interested in complex consequential behavioral patterns arising from either a single modality or several of them. For example, we may be interested in how frequently two entities (objects or people) co-occur together. For another example, we may be interested in the location of an object at the moments that someone is speaking about it – a measure based on the information from two modalities. Combinatorially, there are many possible measures that can be extracted from multimedia data. Thus, how to identify and extract the most informative derived data is a challenge for researchers, which turns out to be a chicken-and-egg problem. To discover new knowledge in scientific studies, researchers may not know in advance what information is most critical and interesting, and thus should be extracted first. But meanwhile, without extracting some data first and computing some results based on these data, researchers may not know where to start.

In the second step of multimedia data analysis, researchers work on derived data (time series, and so on) with the goal to find interesting patterns. Recent advances in machine learning and data mining have provided effective data analysis tools to discover underlying patterns from time series.[1–3] However, these data mining algorithms can effectively search and discover only predetermined patterns and those patterns need to fit a specific definition of statistical reliability. This limitation significantly constrains what can be achieved using standard data mining algorithms because the exploratory nature of discovering new knowledge requires the ability to detect uncommon (but interesting) patterns. This observation points to a similar chicken-and-egg problem like the one in the first step of data processing – without knowing what patterns to look for, data mining algorithms most often cannot work effectively in a purely unsupervised mode. But meanwhile, a huge amount of information must be reduced and summarized in order to find quantitative results. Without the results from some forms of data reduction, researchers may not have concrete ideas on what to look for.

The key to solving this problem is to develop a mechanism that allows researchers to explore the data and gain some insights on how to analyze it.[4–9] In light of this, we propose to use visualization techniques that present the data in various informative ways, making it easier for researchers to employ their own visual perception system and domain knowledge/expertise to detect new patterns that were previously overlooked, to gain new insights, and to generate new hypotheses which will lead to new discoveries. Thus, our solution is to take advantages of both the power of the human perception system and the power of computational algorithms.[10,11] More specifically, researchers can use data mining as a first pass, and then form a closed loop of visual analysis of current results followed by more data mining work inspired by visualization, the results of which can, in turn, be visualized and lead to the next round of visual exploration and analysis. In this way, new insights and hypotheses gleaned from the raw data and the current level of analysis can contribute to further analysis.

As shown in Figure 1, data mining and information visualization are traditionally treated as two independent topics. Data mining algorithms rely on mathematical and statistical techniques to discover patterns while information visualization researchers are interested in finding novel techniques to visually present data in more informative ways. However, it is worth noting that these two topics share the same ultimate research goal – building tools and developing new techniques to allow users (for example, researchers) to obtain a better understanding of massive data. Therefore, the techniques developed in these two fields can be potentially integrated to build a better pattern/knowledge discovery system. In the case of multimedia data, a huge amount of information must be cut and summarized to be useful. But statistics and measures extracted from raw data may exclude embedded patterns or even be misleading. We need a mechanism to represent the overall statistics but still make fine-grained data accessible. Information visualization provides a unique opportunity to accomplish this task. Often, potential users of information visualization are not aware of the benefits of visualization techniques on data mining; or they use those techniques only as a *first* phase in the data analysis process. As shown in Figure 1 (right), we suggest a more interactive mode between data mining and information visualization. In our system, researchers can not only visualize raw data at the beginning but also visualize processed data and results. In this way, data mining and visualization can bootstrap each other – more informative visualization based on new results will lead to the discovery of more complicated patterns which in turn can be visualized again to lead to more findings.

Keim[4] suggested that visual data mining techniques can be classified based on three criteria: the data to be visualized, the visualization technique used, and the interaction and distortion technique used. The organization of the present paper follows this structure. We will first briefly review related work. After that, we will introduce multimedia data sets we used in this study. Next, we will present our visualization tool while focusing on several novel analytical functions we designed and implemented to visualize multimedia multi-stream data. We will explain how these visualization techniques allow users to actively explore the data, highlight certain properties and patterns

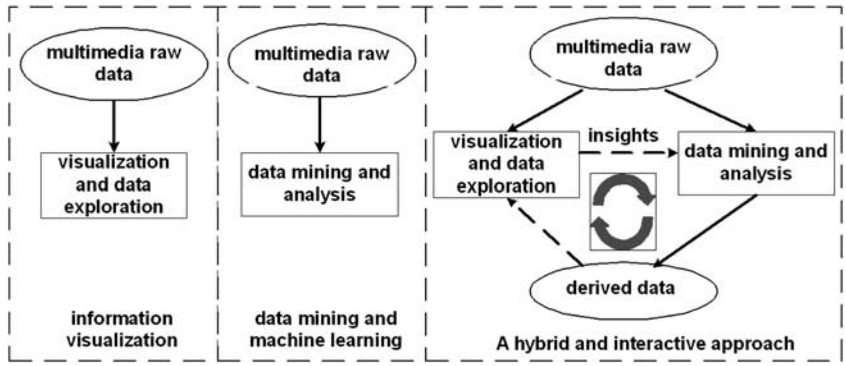

Figure 1: *Left*: Information visualization techniques focus on developing informative ways to visualize data. *Middle*: Data mining algorithms rely on mathematics and statistics to find complex patterns from data. *Right*: Although data mining and information visualization are traditionally treated as two separate topics, they share the same research goal. In light of this, our hybrid and interactive approach builds the links between these two and by doing so forms a closed loop between visualization and data mining.

in multi-stream data and facilitate further data mining. We will also describe both the interface between visualization and data mining and the interface between raw multimedia data and derived data.

Related Work

There are several visualization approaches for multivariate temporal sequences in the literature (for example, Carlis and Konstan[12] and Wang *et al*,[13] see an overview in Aigner *et al*[14]). TimeSearcher [13] is a time series exploratory and visualization tool that allows users to query time series by use of TimeBoxes, which are rectangular query regions drawn directly on a two-dimensional display of temporal data. ThemeRiver[15] is used to visualize thematic changes in large document collections. VizTree[16] is designed to visually mine and monitor massive time series data. It uses symbols to represent time series data first, and then codes those symbols in a modified suffix tree in which the frequency and other properties of patterns are mapped onto colors and other visual properties. Spiral[12,17] is mainly developed to compare and analyze periodic structures in time series data, where the time axis is represented by a spiral, and data values are characterized by attributes such as color and line thickness. Van Wijk *et al*[18] designed a cluster and calendar-based approach for the visualization of calendar-based data. XmdvTool [19] uses multiple brushes to help users gain insights by allowing users to create composite brushes via user-defined logical operations. Those methods deal with linear timing or highly periodic timing, and therefore they are not designed to handle event-based data which is typical in multimedia applications. In general, those methods focus on visualization, navigation or query only (but also see Keogh *et al*[20]). In contrast, our hybrid approach provides an interactive tool to integrate visualization with data mining.

The present paper is also related to data mining. Various algorithms have been proposed to discover sequential patterns.[1,3,21] Those algorithms can only extract temporal patterns that frequently occur in the data. Depending on research topics, researchers may also be interested in searching for unexpected patterns. Most algorithms cannot easily accomplish this task.

Multimedia Datasets

Our visualization program works by extracting various derived variables from raw multimedia data and using colors and other visual properties to visualize multi-stream data in parallel. Before explaining our visualization tool in detail, we will present both representative data sets collected from state-of-the-art behavioral studies and the challenges in analyzing those data that motivated our work.

Raw multimedia data

The example data sets used to illustrate our visualization work were collected from human–human interaction studies in a laboratory environment wherein a pair of participants interacted and talked with each other. We have developed a multisensory setup to record fine-grained behavioral data from two participants. The raw data were collected from various sensing systems, including:

- *Video*: There were three video streams recorded simultaneously with a frequency of 10 frames per second, and the resolution of each frame is 720×480. In each study, we have collected 90,000 image frames.
- *Audio*: The speech of the participants was recorded at a frequency of 44.1 kHz.
- *Body motion*: There were multiple sensors, one on each participant's body parts, for example, the head or hands. Each sensor provided six-dimensional (x, y, z, yaw, pitch and roll) data points at a frequency of 120 Hz.

In each study, we have collected 864 000 position data points.
- *Eye gaze*: An eye tracker recorded the course of a participant's eye movements over time at 60 Hz.

The sensing systems here and multimedia data collected using these equipments represent state-of-the-art techniques used in behavioral studies. Some laboratories may only use 1–2 above recording devices and others may use different models. But in general, many researchers are collecting this kind of multi-stream, multimodal data in their laboratories. Hence, we will use the above data sets throughout the paper as an example to demonstrate not only the functionalities of our system to visualize multimedia data but also how users can visually explore the data to find new discoveries.

Data processing and derived data

Multimedia data sets themselves (for example, video) cannot directly be used to report any quantitative results. Instead, researchers need to formalize and encode the information in the multimedia data into quantitative forms, either manually or with automatic data processing algorithms, and then they can data-mine these coded (derived) data. In this section, we will briefly introduce and characterize derived variables inferred from multimedia data.

Data types
From a multimedia data processing perspective, we find that these temporal data can be categorized into two kinds: (1) a continuous variable $X = \{\langle x_1, t_1\rangle, \langle x_2, t_2\rangle \cdots, \langle x_m, t_m\rangle\}$: : a series of time-value pairs; and (2) an event variable $Y = \{\langle y_1, t_{s_1}, t_{e_2}\rangle, \langle y_2, t_{s_2}, t_{e_2}\rangle \cdots, \langle y_m, t_{s_m}, t_{e_m}\rangle\}$: a set of time intervals with onset and offset timestamps of each instance of a specific event. For example, the location of an object in a video is a continuous temporal variable that may vary over time. The time intervals when a participant is speaking can be captured as an event variable. In each category, we further define two sub-categories based on data types. If a continuous variable X is represented by numerical values x_i, then we call it a continuous time series. For example, the location of an entity (person or object) in the whole video can be represented by two continuous time series, each of which specifies one of the two image dimensions over time. In another case, if x_i is a categorical value, then X is a categorical time series. For instance, if we pre-define n Areas Of Interest (AOIs) in the video, we can then derive a categorical variable with $(n + 1)$ possible values indicating whether an entity is in one of those AOIs (0 – not in any of them). Event variables can also have two data types. One is a binary event. An example of this type is a speaking event indicating whether a person is speaking. The other type is a categorical event. For example, one may be interested in whether a person mentioned a list of key words which can be represented by a categorical event variable wherein each categorical value refers to a speaking event when the speaker utters a particular word in the list. Overall, the above four data types cover general representations of derived variables from multimedia data.

Measurement levels
In addition to different data types, we can also distinguish derived variables with two levels that reflect different levels of abstraction from raw data:

- *Level-1 derived variables* contain those measures that are directly calculated from raw multimedia data. For instance, the magnitude of the speed of a motion sensor is a one-dimensional time series that can be directly calculated based on the three-dimensional location of the sensor over time.
- *Level-2 derived variables* contain the information that is mined from derived variables at level 1. For example, given the time series of the sensor speed as described above, one can further calculate the temporal entropy by applying a sliding window over the whole stream, which will in turn generate a new time series. In this case, each data point in the new level-2 stream corresponds to the entropy over a temporal segment. Overall, the whole stream represents the dynamics of entropy changes over time. For another example, the size of an object in the video over time forms one time series, now one can calculate a binary event variable representing those moments that its size is larger than a certain threshold. This list of temporal events forms a new level-2 event variable, which is based on a further calculation from the level-1 size variable.

In multimedia data mining, researchers can potentially extract various derived variables at level 1 from either a single modality or a combination of multiple modalities. Moreover, various calculations can be applied on those derived variables to generate many more new derived variables at level 2. Taken together, this data processing process reflects the complexity and the challenge in data mining multimedia data. Namely, the potential search space for interesting data patterns is so immerse that in order to analyze this kind of data effectively and efficiently, researchers must find a way to quickly explore both raw data and derived data so that they can then generate concrete hypotheses and insights to guide further data mining efforts in the right direction.

Data processing
With the definitions of both data types and measurement levels, the process of analyzing derived data can be divided into two phases – signal processing (most often generating level-1 variables) and statistical computation (most often generating level-2 variables). In the first phase, given multimedia data, various signal processing and machine learning algorithms can be used to extract various measures, depending on which aspects of the

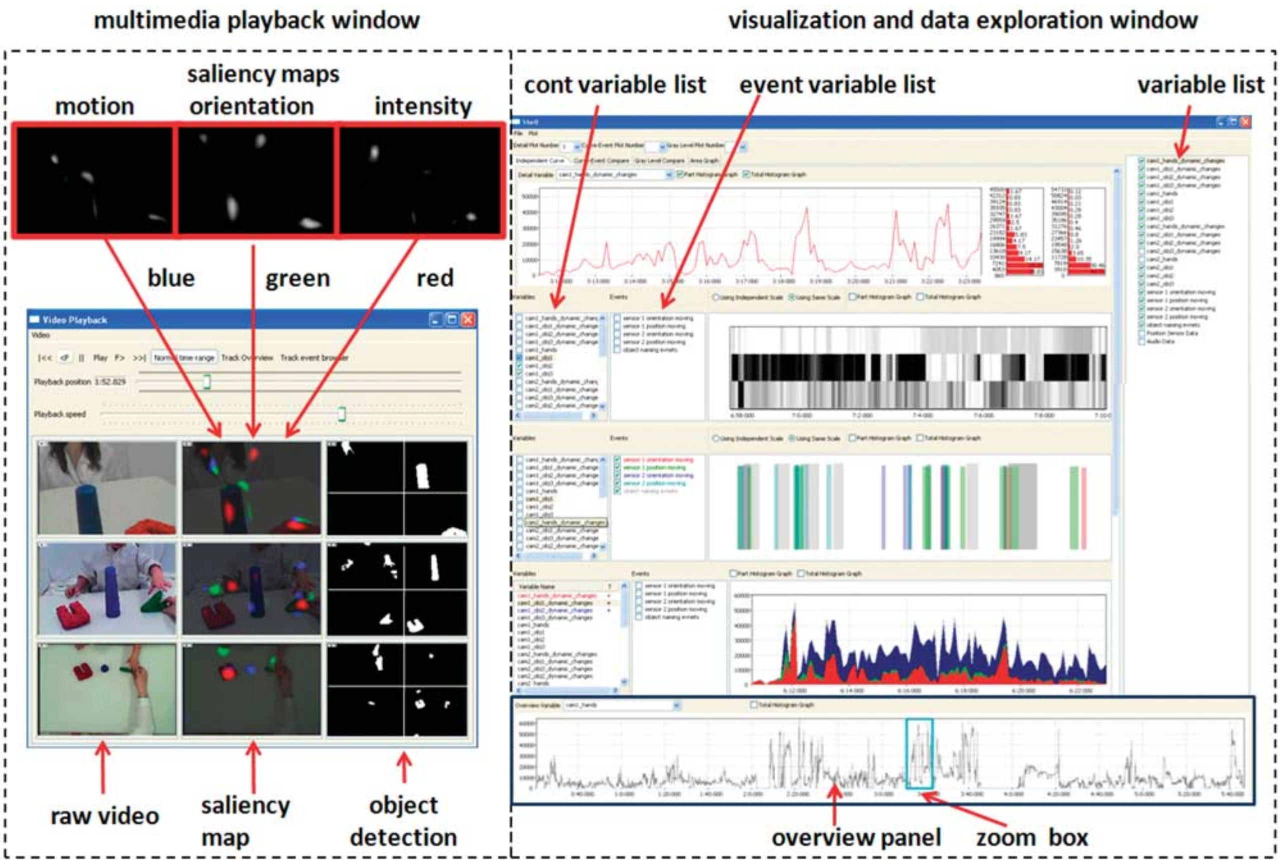

Figure 2: There are two major display components in our visualization system: a multimedia playback window (left) and a visualization window (right). The multimedia playback window is a digital media player that allows users to access video and audio data and play back both raw and preprocessed multimedia data in various ways. The visualization window is the main tool that allows users to visually explore the derived data streams and discover new patterns and findings.

multimedia data that researchers are interested in most. Here we will briefly show several measures used in our data sets as a representative case study:

- *Image/video content*: One set of derived variables describe the content of video. We have developed a computer vision program[22] that can segment individual entities (for example, people or objects) from the background scene. From this segmented image, we derive time series specifying different properties of each entity, such as the location of object A, or the visual size of person B, or the co-occurring moments of object A with person C. These measures can be represented by continuous variables, categorical variables or a binary event variable in the last example.
- *Visual saliency*: We are also interested in extracting visual and perceptual cues at a lower level. To achieve this goal, saliency map algorithms developed by Itti *et al*[23] are applied to our image sequences to generate several saliency maps for each image. In brief, given a video sequence, Itti's saliency map model computes multiple maps which topographically encode for conspicuity at every location in the visual input. Each map is based on one particular visual cue, such as motion, orientation, intensity or color. Thus, one video sequence can generate multiple saliency map sequences. From each saliency map (shown in the multimedia window in Figure 2), we can then treat it as a gray-level image and derive various measures, such as the average saliency value or the location of the most salient spot in each image frame. Moreover, we can also combine the information in object detection and saliency maps to derive measures, such as the most salient object in an image frame.
- *Speech transcription*: We process the transcribed speech to extract some key event variables, such as speaking events (binary) and naming events (categorical).
- *Motion detection*: We also analyze motion data to extract movement patterns (categorical) generated by participants.

In the second phase, various statistical measures are applied to temporal sequences derived from raw multimedia data in the first phase to further quantify multimedia data. For example, we have implemented information theoretic measures, such as entropies of individual variables and transfer entropies between two variables.[24] Since each entropy-based measure is applied on a small local window which is moved over the whole sequence, a new time series is generated as a level-2 continuous variable.

In summary, we record and extract various data and derived results from multimedia data sets. But the ultimate goal of the research is to infer meaningful results from these multimodal multi-stream data. Given the complexity and the pure amount of the data, any temporal and spatial patterns within individual variables and between multiple variables may lead to new discoveries. For such data, researchers are interested in discovering detailed behavioral patterns, such as at what moments a participant is holding and as well as gazing at an object, when two participants jointly attend to the same object, in what ways the actions from one participant cause responsive actions from their social partner and what are shared behavioral patterns across multiple participants in the same situation. The challenge of discovering these patterns from the sea of raw and derived data motivates us to build a visual data mining system that allows us to visually and interactively explore these data to quickly identify new candidate patterns and perform real-time hypothesis testing.

Visual Mining of Multimedia Data

An overview

Ben Shneiderman, one of the best scientists in information visualization and human – computer interface, famously proposed the general principles of building scientific data visualization systems: 'overview, zoom & filter, details-on-command'.[25,26] Our visualization program embodies these principles in the case of multimedia visual data mining, which comprises two major display components as shown in Figure 2: a multimedia playback window and a visualization window. The multimedia playback window is a media player that allows users to access video and audio data and play them back in various ways. The visualization window is the main tool that allows users to visually explore the derived data streams and discover new patterns and findings. More importantly, when users visually explore a data set, these two display windows are coordinated to allow users to switch between synchronized raw data and derived data, which we will discuss more later. We will first introduce the analytical functions in our visualization system.

The visualization and data exploration window is designed based on TimeSearcher.[27] There are three display areas. After users load a multimedia data set, which usually consists of all the data collected plus derived data from that raw data, the variables in the data set are displayed in a window in the upper right corner of the application. Each variable is labeled by its name. Users can select which ones they want to load into individual display panels. These individual display panels and an overview display panel occupy the central area of the display window. The overview display panel at the bottom of the application is the place that users can select any of the loaded variables as a reference to present global trends in the data. Within the overview display, users can drag and resize a 'zoom box' to define the area of interest in the time axis. This zoom box allows users to control the level of detail in the central display area wherein users can select and examine multiple variables simultaneously by zooming into the AOIs defined by the zoom box in the overview panel and comparing multiple data streams side by side. We have developed various functions to visualize derived data streams individually or together to highlight different aspects of multimedia multivariable data.

Visualization of raw video data and preprocessing results

Although most other information visualization systems provide various functionalities to display (and allow users to interactively explore) derived data, we argue that the access to raw data and preprocessed data is critical, in particular for visual pattern discovery of multimedia data, for two major reasons. First, as mentioned earlier, a set of derived variables based on signal processing is just a part of information encoded in multimedia data and researchers need to access raw data to obtain a more complete picture of the data set while considering the possibility of extracting more derived measures that capture different aspects of multimedia data. More importantly, although visualizing derived time series in parallel is a useful way to explore the underlying patterns, especially those temporal patterns, the potential missing information in individual graphable derived data is the overall spatial-temporal correlations between these variables embedded in the same video. Thus, visualizing processed video streams as intermediate results of data mining, such as video with objects segmented from one another and video with particularly salient objects highlighted, provides a unique level to examine the data that allows researchers to explore spatial-temporal patterns in the context of raw video data. The technical problem here is how to use limited display space to display more preprocessing results since each raw image frame maps to multiple processed images. Our solution is to use colors and synthesize a new image with multiple layers. As shown in the middle column in the multimedia playback window in Figure 2, three additional layers are superimposed on the top of the original image, each layer corresponding to one saliency map (a gray-level image) based on one visual cue. Thus, each saliency map adds one more layer on the top of the original image. In our current implementation, the blue channel corresponds to the orientation cue, green to the motion cue and red to the intensity cue. In this way, the combined effects of three layers indicate how salient each area/pixel in an image is and which saliency cues are stronger. For example, a white pixel would suggest this point is very salient for all three visual cues. In contrast, a black pixel means that spot is not salient at all and a yellow (red + green) pixel means the location is salient for both the motion and intensity cues. In fact, this superimposing technique can be generalized to visualize any derived variables back into raw video data to highlight their correlations with other information in the video. For example, we can superimpose the segmented images based on object detection (small binary images shown in the right column of the multimedia playback window in Figure 2) with saliency maps to depict which objects are more salient.

Visualization of derived data

In the following sections, we will first present how our visualization tool deals with continuous data and event data, and then we will introduce how we visualize these two kinds together and how users can perform event-based visual exploration.

Continuous time series data
After loading the data set, a list of continuous variables is displayed next to individual display panels, from which users can select one or more variables to display. Our visualization tool supports three ways to visually explore continuous time series data: (1) as individual data streams, (2) as a set of multiple data streams and (3) as an arithmetic combination of multiple data streams. We will present each mode one by one, followed by depicting the visualization of categorical time series.

Using curves to visualize individual data streams The purpose is to allow users to explore individual data streams and examine both the overall statistics of a data stream and the statistics within a local window – both the global and the local structures. As shown in Figure 3(a), users can examine multiple streams at the same time, one for each display panel. Meanwhile, users can move the zoom box in the overview panel to zoom in and take a closer look within a time window. A novel feature we added here is histogram display. There are two histogram plots – one corresponding to the overall histogram of the time series and the other corresponding to the local histogram of the current window. The local histogram is updated as users move the zoom box while the global histogram is constant. Compared with summary statistics and metrics (means and variances, and so on) extracted from the data, the histogram reveals more fine-grained features such as whether distributions are uniform, normal, skewed, bimodal or distorted by outliers, and as well as the range of the time series and the proportion of each bin. By doing so, our visualization tool transparently allows users to examine the overall distribution of the data and discover potential interesting features which can be then quantified and confirmed by calculating these features in data analysis. Moreover, by putting the two histograms side by side, users can easily compare these two and check whether the local distribution is the same as the global distribution. The top plot in Figure 3(a) shows an example where these two are similar, while in the bottom one they are quite different. In this way, users can easily identify and extract these local durations that can be then used to interpret the underlying implications of these data patterns. Also, the time-query box put forth in Keim[4] can be added here as an advanced feature to explore individual time series.

Using gray-level representation to visualize a set of multiple data streams The main purpose here is to visually display and explore two kinds of information: (1) the possible correlation between multiple data streams and (2) interesting joint patterns/states across multiple data streams. First, users are allowed to select multiple variables and display them as parallel gray-level data streams to explore the potential synchrony and correlation between multiple variables. For example, by viewing two time series together, users can easily detect whether two streams are correlated or not. As shown in Figure 3(b), human observers can immediately detect that the top two variables are closely correlated. Thus, even though the two streams are not perfectly aligned, our visual observation can tolerate and compensate for a certain degree of time shifting between two temporal streams. Moreover, even though the absolute values in the two streams are not perfectly matched, our visual system is still sensitive to the overall similarity between the two. Meanwhile, users can easily notice that the bottom two variables in Figure 3(b) are not correlated because of the unsynchronized changes in the two streams. The advantage here compared with data mining algorithms is that users can dynamically adjust their judgment of the similarity (time shifting or value differences) based on their visual observation. Users can make and test hypotheses in seconds, with no need to take the time to implement a data mining algorithm as an external tool. Moreover, our visual judgment is more flexible than parameterized data analysis algorithms. Users can easily extend this pairwise comparison to more general cases by selecting more than two temporal variables and examining the possible temporal correlations across all of them. To make this function more flexible, our visualization tool also allows users to select how to convert data values into gray levels, which can be either based on an individual data stream itself or normalized across all data streams displayed together.

Second, by simultaneously displaying multiple temporal streams, users can also spot interesting patterns in those data streams. These interesting patterns usually cannot be easily calculated using data mining algorithms because there may be only a few instances of each pattern that are embedded in the sea of irrelevant data. Therefore, most algorithms are less likely to find those patterns because they do not occur often enough. Thus, without any insights about what to look for, researchers cannot take advantage of the power of data mining techniques which

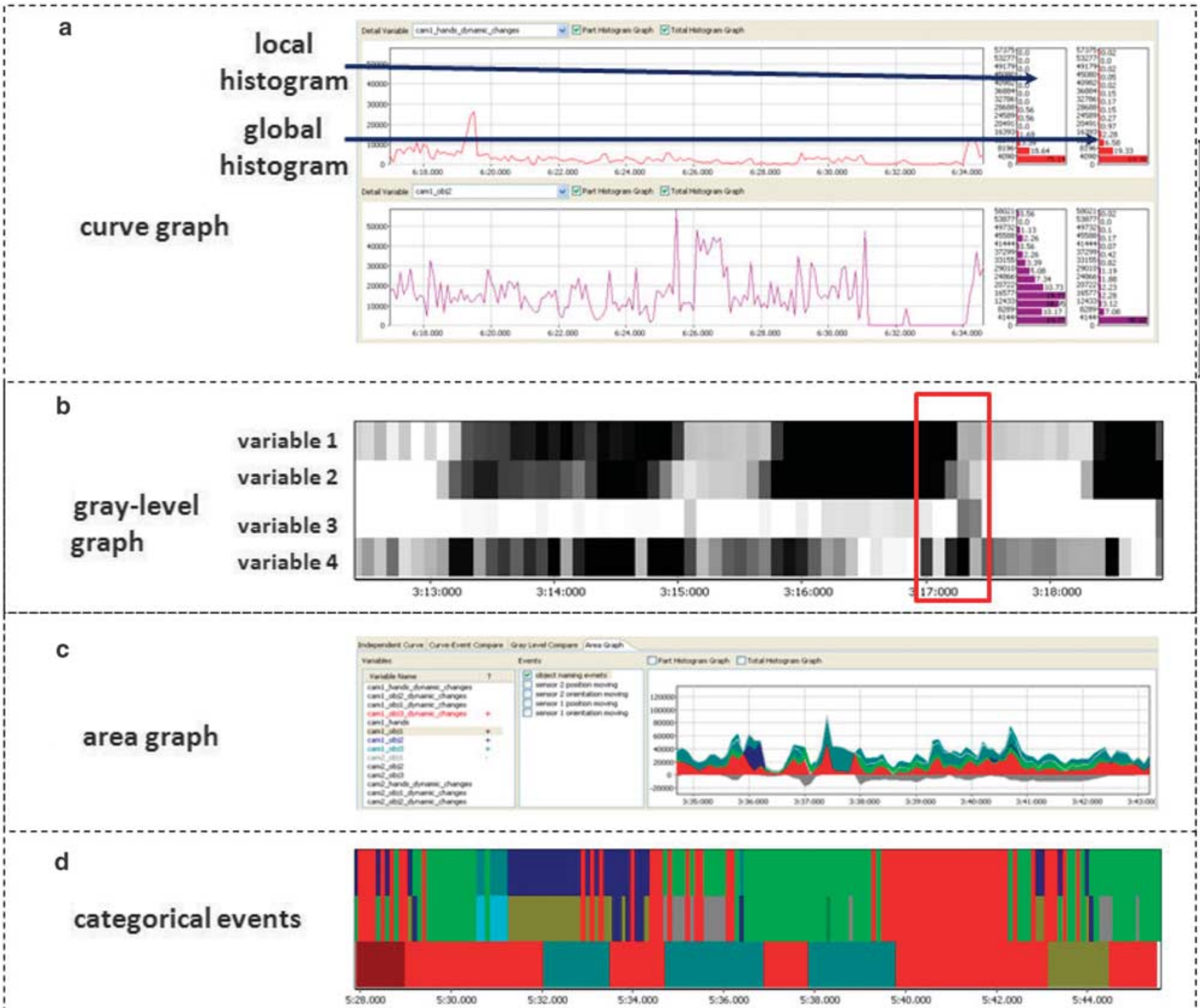

Figure 3: Various ways of visualizing continuous and categorical time series. **(a)** Using *curves* to visualize individual continuous data streams with local and global histograms (structures). We can easily note that the local histogram and the global histogram are similar for the first time series, while they are different for the second variable. **(b)** Using *gray-level* to examine a set of continuous variables. First, we can see that variables 1 and 2 are correlated while variables 3 and 4 are not. Second, the red box highlights a moment that all the variables are changing indicating a potentially interesting pattern in the data streams. **(c)** Using area graphs to visualize both an arithmetic combination of multiple data streams and the proportion of each component. **(d)** Multiple categorical events are displayed in parallel based on the same color coding so that users can easily see the temporal synchrony and correlations between those events.

are most often purely based on statistics and therefore can find only pre-defined patterns or frequent patterns. Nonetheless, the undetected patterns may be particularly interesting for scientific purposes which may lead to new knowledge discovery. Here lies in the advantage of information visualization in which users can first visually spot those patterns (see an example in Figure 3(b) highlighted by a box) and then use data mining techniques to quantify their observations and obtain more quantitative and objective results.

Using area graphs to visualize an arithmetic combination of multiple data streams Our visualization tool also allows users to examine the joint effects of continuous temporal variables by using area graphs. More specifically, users can select multiple continuous variables from the continuous variable list and decide the 'sign' of each variable. We use area graphs to present those variables. A '+' sign (addition) will put a data stream above the time axis and a '−' sign (subtraction) will indicate that the variable should be put below the time axis. In this way, users can combine

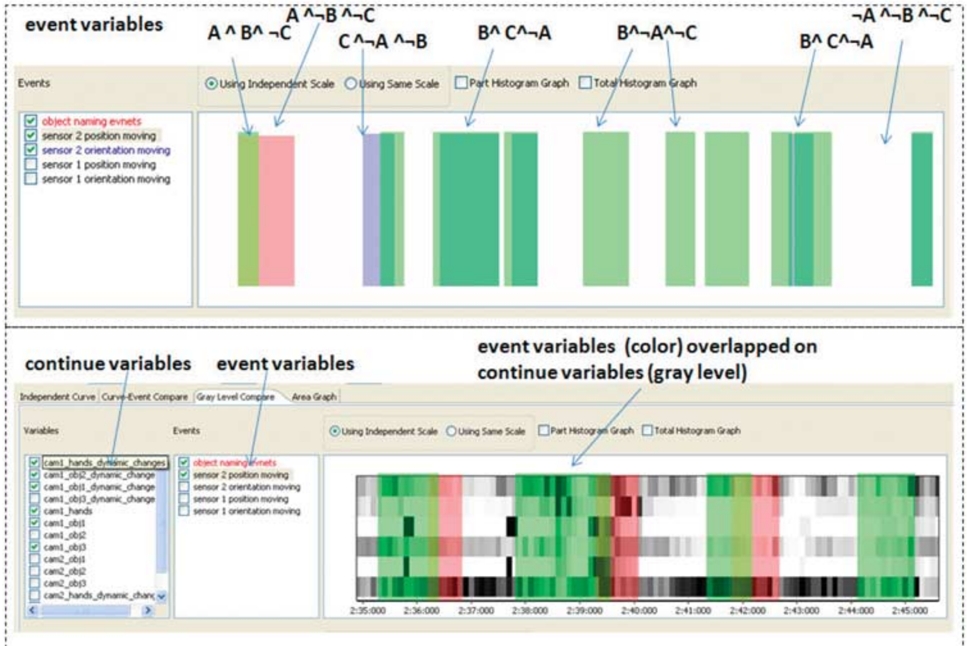

Figure 4: *Top*: Visualizing multiple event variables together over time allows users to see not only individual events but also joint events. *Bottom*: Users can select a subset of continuous variables and a subset of event variables. The display panel will highlight those continuous values at the moments when the selected events happen. In this example, users can visually examine the underlying patterns of those selected continuous variables when either the green event, or the red event, or both, occur.

multiple temporal variables together (by addition and subtraction) in various ways and then visually explore the combined distribution. Figure 3(c) shows two examples in which users can explore not only the overall distribution of the combined variables but also the proportion made up by each component (an individual variable and so on) moment by moment.

Using color representation to visualize categorical time series So far we have only introduced the methods to visualize continuous time series with numerical values. In multimedia data mining, the other derived data type is categorical time series. A different color is used to represent each category so that each categorical stream is represented by a multicolor sequence. Moreover, researchers may be interested in comparing multiple categorical streams. When doing so, they can determine whether those streams should use the same color coding if the derived variables share the same set of categories. For example, one categorical variable may describe which object a person is holding, a second one describing which object the same person is gazing at and the third one describing which object the other person is gazing at. As shown in Figure 3(d), by using the same color coding, researchers can immediately explore at what moments those measures share the same values, and at what moments one variable may be a precursor to other variables.

Event data
Events are presented as bars of color, with their length on screen corresponding to their duration. To select event variables for display, they are shown in a list next to the continuous variable list. Event data can be displayed alone or with continuous data. In both cases, each binary event is presented by a unique color and each categorical event is presented by a set of colors. As shown in Figure 4 (top), users can see not only an event itself but also the conjunction of two events because the overlapping of two colors/areas will generate a new color. For example, if two binary events A and B are labeled as ‘red’ and ‘green’, then users can see four color categories: red – ‘Aˆ¬B’, green – ‘¬AˆB’, yellow – ‘A ˆB’ and white – ‘¬Aˆ¬B’. In a more general case with more events, our color-based event visualization scheme is still able to represent all possible logic conjunctions of those events.

As shown in Figure 4, now that users can visually explore several temporal characteristics of a single event: (1) the occurrence frequency of the event; (2) its duration in each instance and (3) its periodicity across multiple instances. Moreover, by displaying multiple event variables, our visualization tool allows users to explore sequential statistics across events, such as joint probability or transition probability. Joint probabilities are visually represented by the overlapping (new color) of multiple events – how frequently the new combined event (A ˆB) occurs. There are two kinds of transition probabilities: (1) P(A|B) – users can visually examine whether and if

Figure 5: Users can define a new event based on three logic operations on multiple current events.

so how frequently event A occurs after event B (red and green bars are paired while red ones are always followed by green ones); (2) P(A^B|B) – users can explore whether and if so how frequently the joint event A and B always occurs after event B. Moreover, high-order transitional probabilities, such as P(A^B|C^D), can also be visually detected and examined. Although one can argue that those probabilities or frequencies can be calculated easily in data analysis, the practical problem in data exploration and knowledge discovery is that given multiple variables and many possible combinations of them with various logic operations (and, or, not) and in different orders, this combinatorial problem can be formidable. Now researchers can use our visualization tool to significantly reduce their search space by quickly going over the whole stream while rejecting uninteresting and unreliable patterns, and concentrating on interesting ones.

The above visualization is implemented by assigning different colors to different event variables that users select to display. One potential limit is that users can probably keep track of only a small number of distinct colors over time. Moreover, complex patterns involving more variables and logic operations may be hard for users to examine visually based only on color coding. To overcome these limits, we also allow users to define a new event variable by combining existing events with three universal logic operators (AND, OR, NOT). As shown in Figure 5, users can create a new event and assign it a meaningful variable name. After that, the new event variable will be automatically added in the event list to allow users to select and visualize this new variable. For example, users may start with 20 distinct events, but through the first round of visual data mining, they may notice that event A and event B always co-occur together while event C co-occur with the two occasionally; to directly examine the joint activities of these three events, we can definite two new event variables ABC = A^B^C and AB_C = A^B^ (not C). By doing so, users can directly not only compare these two new events to determine the occurrence of event C with the other two but also examine the relations of these two new events with other events. More generally, users can bootstrap this data mining process by applying a hierarchical-clustering strategy on initial event variables, gradually reducing the number of events that always co-occur together, and ultimately identifying the underlying complex patterns defined by high-order combinations of multiple events.

Concurrent visualization of continuous and event variables

The visualization functions described so far concentrate on visualizing either event variables or continuous variables. Here we present an approach to visually exploring the combination of these two. We are interested in exploring the potential complex patterns hidden in continuous variables *conditioned* on event variables – what trends and patterns exist in the continuous variables when certain events happen. Our approach is to use colors to visualize various events while using gray levels to visualize continuous values, and overlap these two. As shown in Figure 4 (bottom), users can select multiple continuous variables displayed in parallel as before to visually spot potential patterns across those data streams. Meanwhile, users can also select multiple events (with different colors) overlapped on the top of gray-level continuous variables to visually examine the underlying patterns of continuous variables at those moments when certain events happen. We allow users to select one or multiple events and also assign a transparency value for each event to highlight certain events over others. The other purpose of re-assigning transparency values is to allow users to trade-off between the visibility of events and the visibility of the underlying continuous data. Since the overall visualization effects are based on the combination of continuous and event variables, both of which change over time, we cannot use pre-defined transparency values that work perfectly across the whole data streams with different variables. Instead, users can dynamically adjust transparency values to find the best visualization effect in the current local window to visually examine both events themselves and as well as the underlying continuous values.

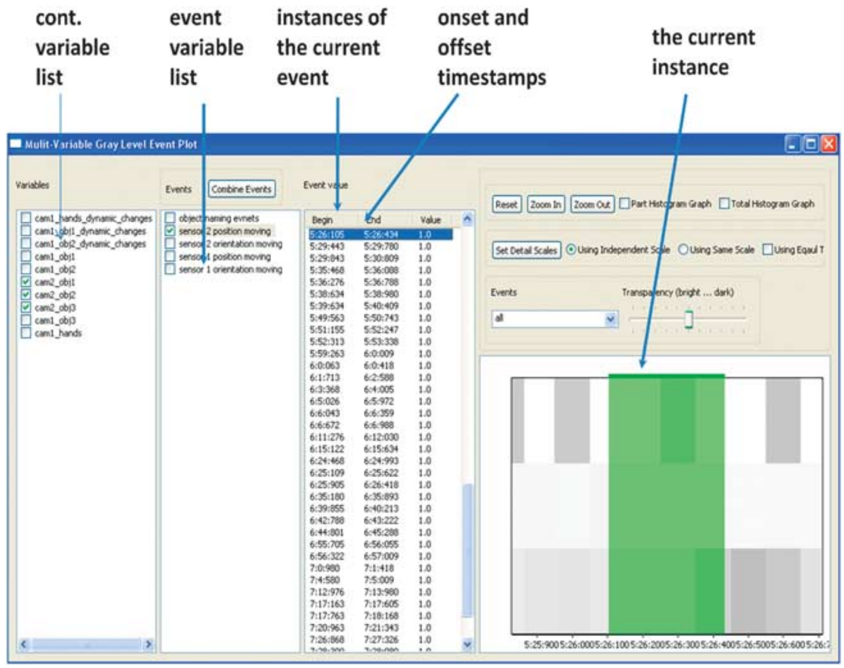

Figure 6: Users can visually explore individual instances of an event one by one and visually analyze the underlying patterns of selected continuous variables case by case.

Event-based interactive visual exploration

We observe that many multimedia data are essentially event-driven. Researchers can quantify their observations in terms of individual events, such as the event of a person entering a room, the event of a certain object appearing in a certain location or the event of two people shaking hands. Therefore, we provide advanced functions to allow researchers to examine individual events thoroughly. As shown in Figure 6, after users select both an event and one or multiple continuous variables, all the instances in this event are listed with start and end timestamps. Users can explore those moments one at a time and the corresponding continuous values are displayed and updated as well. In this way, users can examine in a fine-grained way the patterns of the underlying continuous variables within a certain event. While doing that, users are also allowed to change the time axis to zoom in or out to examine what happen before and after each instance of the selected event while the display panel is centered on the current instance. By visually exploring the data instance by instance, users can directly compare those moments to detect the similarities between these. In this way, this event-based visual exploration allows users to gain insights about complex patterns between continuous and event variables, both frequently occurring patterns and surprising (abnormal) patterns, which cannot be easily achieved by standard statistical data mining techniques. This is because data analysis algorithms may overlook interesting patterns in the data when they extract statistical patterns from the data. Meanwhile, without a better understanding of the data itself, researchers cannot simply keep track of all the details over time when a certain event happens. Visual exploration allows researchers to directly obtain the insights on complex patterns.

Our visualization tool also allows researchers to reorganize the data so that they can examine and comprehend the data better. Users can visually examine each instance of an event, and categorize the instances into groups. Then, the instances can be viewed side by side based on their grouping. As shown in Figure 7 (right) where each column shows the instances within the same group. Users can visually analyze and compare all the instances on the same display and regroup them as needed, a display technique called 'small multiples' which works well for users.[11] This complete picture of all the instances allows users to examine both similarities within each group and dissimilarities across groups, and by doing so facilitate users to improve the overall grouping and thus more accurately capture various patterns from the underlying data. When users finalize their grouping, they can save the results of this visual analysis which can be directly used as a guide for further data analysis. This function is particularly useful in our multimedia data mining with complex patterns. For example, with 20 instances of an event in the video that can be grouped into three categories based on underlying data, using our visualization program, users can easily do so by examining those instances one by one visually and grouping them

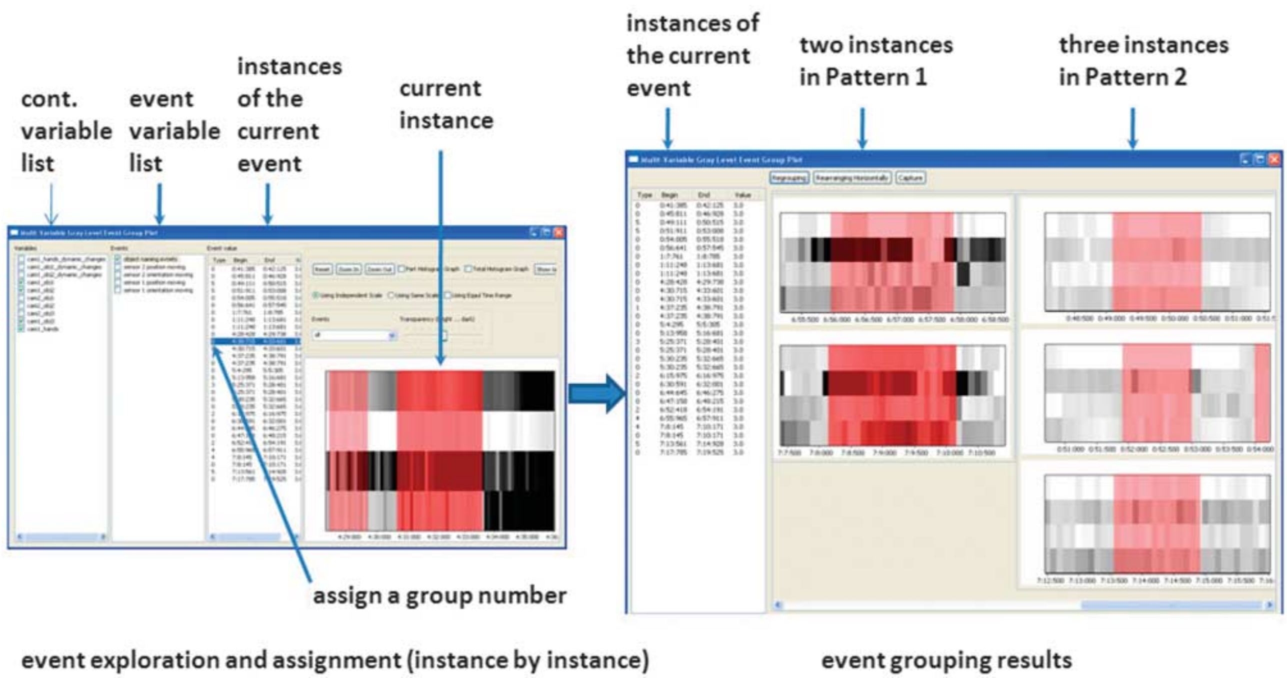

Figure 7: Visual Event grouping. All the instances of the same event can be grouped based on the patterns of underlying continuous variables. The overall grouping results can then be visualized in one single panel where each column includes the instances belonging to the same group.

based on the pattern of each instance. However, if users just 'blindly' compute aggregated statistics from those instances, since we don't know how many patterns in advance which is a critical parameter in any data mining algorithms, users may not be able to discover those three patterns, and in a worse case, researchers could reach wrong conclusions due to wrong parameter selection.

Comparing the data collected from different participants

So far, we focused on how our visualization program can be used to find underlying patterns between different variables derived from the same participant. One important goal in data mining of behavioral data is to find both shared, frequently occurring patterns and surprising patterns *across* participants. Our visualization program provides a unique way to facilitate this task. As shown in Figure 8, we load eye movement data (for example, where a person is looking at) while they were asked to watch the same video clip displayed on a computer screen. Each stream in the figure corresponds to the same derived variable but collected from different participants. In this example, since all of data streams are temporally aligned (watching the same video, and so on), users can easily compare those time series to spot interesting patterns. More specifically, there are at least two immediate outcomes from this visual exploration. First, we can discover those moments that all the participants behave similarly. This pattern can be spotted by examining multiple time series vertically. Meanwhile, by examining those data streams horizontally, we can also easily find those individuals who are different with others (for example, participant 3 in this example). Moreover, this explicit and informative visualization allows us to also discover *in what ways* those individuals are different. In the above example, participant 3 seems to always generate eye movements right before most of people do so – the pattern that can be easily detected through this visualization.

Visualization and data processing

In the previous sections, we have presented various ways to visualize both raw and derived data in our visualization system. This section will revisit the two chicken-and-egg problems in multimedia data mining that were described in the introduction and present our visualization-based solution to these two problems – namely how to determine what derived variables should be extracted and how to determine which derived variables should be closely analyzed to find new patterns. In addition to various analytical functions provided in our visualization tool that facilitate users to effectively examine the data visually, we also provide flexible interfaces between visualization and data mining that allow researchers to smoothly switch between the two. This section introduces two interfaces: (1) between raw data and derived data, and (2) between visualization and data analysis.

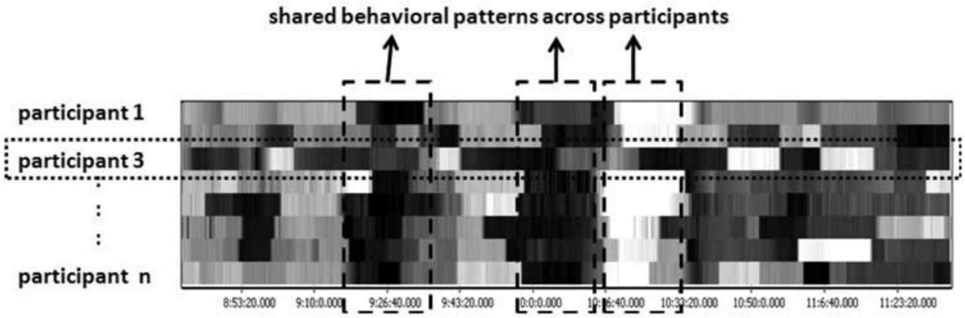

Figure 8: Application of our visualization program to the problem of comparing the same derived variable collected from multiple participants. By visually exploring those multiple streams in parallel, researchers can easily detect the shared behavioral patterns across participants. Namely, most participants generate the same behaviors or perceive the same information at those moments. At those moments, they demonstrate individual differences. Moreover, researchers can also easily identify individuals that are different with the whole group. For instance, participant 3 seems to be an outlier whose data are quite different with other participants.

Synchronization of multimedia data and visual data exploration

We have described various ways for researchers to visually examine and analyze continuous and event variables derived from multimedia data. We note that derived data are extracted from raw multimedia data and can only include a subset of the information available in the raw multimedia data. Therefore, it is important that users can refer to the raw multimedia data while exploring derived data. Our multimedia playback window allows users to play back video and audio data at various speeds, from fast forward/backward to frame-by-frame playback. Users can also control the onset of the playback and stop/restart the video at any moment. On the top of these standard video playback functions, we design and implement one critical component to connect multimedia playback with real-time visual data mining – an automatic control of the interval of video that is played back using the visual data mining tools. A key technical issue in implementing this feature is to synchronize in time video playback with users' ongoing visual exploration. We note two places that users may change the current time in visual exploration and they may want to go back and consult raw multimedia data. Therefore, we have developed corresponding mechanisms to synchronize multimedia playback with visual exploration at these two places. First, users may drag or resize the zoom box in the overview panel to look in detail at particular parts of the data set (Figures 3 and 4). When users do this, our playback program also updates the play head of playback so that if users switch to the media play panel and start to play the multimedia data, they will see raw multimedia data starting exactly with the moment that is being examined in the graphs. Second, when users explore the instances of a certain event page by page (Figures 6 and 7), we also synchronize the media playback with event exploration. In this way, users can watch multiple video and audio segments one by one, each of which corresponds to one instance of the event. The boundaries of a multimedia segment are defined by the onset and the offset of an event.

Access to raw multimedia data is critical in this kind of data mining since multimedia data contain more information and is more transparent to the human perceptual system than derived data. By visually exploring the current derived data and meanwhile examining the corresponding raw data, users can gain insights about what additional variables are missing and should be extracted from raw data. In this way, users can start with a small set of derived data, either continuous or event variables, and gradually augment the data exploration scale inspired by the observations based on the current results.

Visual exploration and data mining

Our visualization tool supports various procedures that allow users to examine both raw and derived data, and gain insights and hypotheses about interesting patterns embedded in the data. All this is accomplished by human observer's visual system. In order to quantify and extend these observations, researchers need to develop and use data mining algorithms to extract and measure the patterns detected in visual exploration. We note that different researchers may have different preferences of programming languages and may prefer to use certain software packages. To increase the flexibility to be compatible with data mining, our system allows users to use any programming language to obtain new results. Thus, data researchers can implement new data mining algorithms using their own analysis tools (from Matlab, to R and to C/C++) and as far as they write the results into text files (for example, CSV form) with pre-defined formats (one for each of four data types explained earlier), our program monitors user's workspace in real time and will automatically load new derived variables into the variable list so that users can immediately visually examine these new results. In this way, our visualization system supports a close and flexible coupling between visual exploration and data mining. The insights gleaned from visualization can be used to guide further data mining. Meanwhile, the results from the next round of data mining can be visualized which allows users to obtain new insights and develop more hypotheses with the data. Overall, our visualization system is 'open-minded' by not adding any constraints, assumptions or simplification on raw and derived data, but instead allows users to guide the direction and systematically explore the data through informative visualization, which is truly the power of visual data mining.

Conclusion

Just like many other scientific fields, a new trend in social, cognitive and behavioral studies is to collect massive data on human behavior partly because of the advances in computing and sensing techniques. Researchers face with the challenge of how to analyze such data efficiently, especially considering the fine-grained patterns researchers look for and the exploratory nature of this kind of research. Meanwhile, many scientists are not aware of the power of visualization and visual data mining, where we can rely on our eyes as the best data mining tool if we can provide the right input to our visual systems. This work is our first effort to build such a tool while we focus on a new framework of visual mining of multimedia data. The key idea is to integrate data visualization and data mining. Based on this idea, we have developed a prototype system with several critical features to facilitate pattern discovery. First, we decided to decompose and represent multimedia data as a set of continuous variables and event variables. Second, we developed various ways to visualize these four kinds of data types separately and jointly. Third, we visualized not only raw multimedia data but also all intermediate and final results of data mining, which allowed researchers to access the 'ground truth' of an experiment along with the results. Fourth, we provided a flexible interface between our visualization tool and data mining tools. Overall, our visualization tool allowed users to not only easily examine and synthesize information into new ideas and hypotheses but also quickly quantify and test the insights gained from visualization. Our next step is to conduct a systematic evaluation of our prototype system. We plan to use the experimental paradigm developed by Saraiya *et al*[28,29] to test what kinds of new findings researchers (both experts and novices) can obtain by using this tool. By doing so, we will have a better idea of what are the advantages and limitations of the current system and what needs to be improved in future work. In addition, we hope that our work will motivate and convince researchers working on multimedia data in general and human behavioral data in particular to use information visualization techniques in their data mining and data analysis.

Acknowledgements

This work was supported in part by National Science Foundation Grant BCS0544995 and by NIH grant R21 EY017843. We thank Alfredo Pereira and Linda Smith for their collaboration in collecting multimedia data used in this paper, Hong-Wei(Henry) Shen for generating information theoretic measures from multimedia data.

References

1 Keogh, E. J. and Pazzani, M. J. (1999) Relevance Feedback Retrieval of Time Series Data. Proceedings of the 22nd Annual International ACM SIGIR Conference on Research and Development in Information Retrieval SIGIR '99; August, New York. Berkeley, CA: ACM, pp. 183–190.
2 Agrawal, A. and Srikant, R. (1995) Mining sequential patterns. P.S. Yu and A.L.P. Chen (eds.), Proceedings of the 11th International Conference on Data Engineering, ICDE; March, New York. Taipei Taiwan: IEEE Press, pp. 3–14.
3 Oates, T. (1999) *Identifying Distinctive Subsequences in Multivariate Time Series by Clustering*. The 5th International Conference on Knowledge Discovery and Data Mining; San Diego. CA: ACM Press, pp. 322–326.
4 Keim, D. A. (2002) Information visualization and visual data mining. *IEEE Transactions on Visualization and Computer Graphics*, 8(1): 1–8.
5 Westphal, C. and Blaxton, T. (1998) *Data Mining Solutions: Methods and Tools for Solving Real World Problems*. New York: John Wiley & Sons.
6 Oliveira, M. C. F. D. and Levkowitz, H. (2003) From visual data exploration to visual data mining: A survey. *IEEE Transactions on Visualization and Computer Graphics*, 9(3): 378–394.
7 Keim, D. A., Mansmann, F., Schneidewind, J. and Ziegler, H. (2006) Challenges in Visual Data Analysis. Proceedings of the International Conference on Information Visualization (IV 2006); Los Alamitos: IEEE Press: Piscataway.
8 Simoff, S. J., Bohlen, M. H. and Mazeika, A. (2008) Visual data mining: An introduction and overview. In: *Lecture Notes in Computer Science*, vol. 4404, Berlin, Heidelberg: Springer, 1–12.
9 Shneiderman, B. (2002) Inventing discovery tools: Combining information visualization with data mining. *Information Visualization*, 1(1): 5–12.
10 Shneiderman, B. (1996) The Eyes Have It: A Task by Data Type Taxonomy for Information Visualizations. The IEEE Symposium on Visual Languages; Silver Spring, MD. Boulder, CO: IEEE Computer Society Press, pp. 336–343.
11 Huettel, S., Mack, P. B. and McCarthy, G. (2002) Perceiving patterns in random series: Dynamic processing of sequence in prefrontal cortex. *Nature Neuroscience*, 5: 485–490.
12 Carlis, J. V. and Konstan, J. A. (1998) Interactive Visualization of Serial Periodic Data. ACM Symposium on User Interface Software and Technology, November, New York. San Francisco, CA: ACM Press, pp. 29–38.
13 Wang, C., Yu, H. and Ma, K. L. (2008) Importance-driven time-varying data visualization. *IEEE Transactions on Visualization and Computer Graphics (Proceedings Visualization/Information Visualization 2008)*, 4(6): 1547–1554.
14 Aigner, W., Miksch, S., Muller, W., Schumann, H. and Tominski, C. (2007) Visualizing time-oriented data – A systematic view. *Computers & Graphics*, 31(3): 401–409.
15 Harve, S., Hetzler, E., Whitney, P. and Nowell, L. (2002) ThemeRiver: Visualizing thematic changes in large document

collections. *IEEE Transactions on Visualization and Computer Graphics*, 8(1): 9–20.

16 Lin, J., Keogh, E., Lonardi, S., Lankford, J. P. and Nystrom, D. M. (2004) Visually Mining and Monitoring Massive Time Series. Proceedings of the Tenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining; 22–25 August, Seattle, WA, USA. Piscataway NJ: IEEE Press, pp. 460–469.

17 Weber, M., Alexa, M. and Muller, W. (2001) Visualizing Time Series on Spirals. Proceedings of 2001 IEEE Symposium on Information Visualization; 21–26 October, San Diego, CA. Piscataway NJ: IEEE Press. pp. 7–14.

18 Van Wijk, J. J. and Van Selow, E. (1999) Cluster and Calendar based Visualization of Time Series Data. Proceedings of IEEE Symposium on Information Visualization; 24–29 October, San Francisco, CA. Piscataway NJ: IEEE Press. pp. 4–9.

19 Ward, M. O. (1994) Xmdvtool: Integration Multiple Methods for Visualizing Multivariate Data. Proceedings of Visualization '94, Piscataway NJ: IEEE Press. pp. 326–333.

20 Keogh, E., Hochheiser, H. and Shneiderman, B. (2002) An Augmented Visual Query Mechanism for Finding Patterns in Time Series Data. Proceedings of the Fifth International Conference on Flexible Query Answering Systems; 27–29 October, Lecture Notes in Artificial Intelligence, Berlin, Copenhagen, Denmark: Springer-Verlag, pp. 240–250.

21 Berndt, D. J. and Clifford, J. (1996) Finding patterns in time series: A dynamic programming approach. In: U.M. Fayyad, G. Piatetsky-Shapiro, P. Smyth, and R. Uthurusamy (eds.) *Advances in Knowledge Discovery and Data Mining*, Cambridge, MA: AAAI Press/MIT Press, pp. 229–248.

22 Yu, C. and Ballard, D. H. (2004) A multimodal learning interface for grounding spoken language in sensorimotor experience. *ACM Transactions on Applied Perception*, 1: 57–80.

23 Itti, L., Koch, C. and Niebur, E. (1998) A model of saliency-based visual attention for rapid scene analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 20(11): 1254–1259.

24 Marko, H. (1970) The bidirectional communication theory – A generalization of information theory. *IEEE Transactions on Communications*, 21(12): 1345–1351.

25 Shneiderman, B. (1994) Dynamic queries for visual information seeking. *IEEE Software*, 11: 70–77.

26 Shneiderman, B. (1995) Tree Visualization with tree-maps: 2-D space-filling approach. *ACM Transactions on Graphics*, 11: 92–99.

27 Hochheiser, H. and Shneiderman, B. (2004) Dynamic query tools for time series data sets: Timebox widgets for interactive exploration. *Information Visualization*, 3: 1–18.

28 Saraiya, P., North, C. and Duca, K. (2005a) An insight-based methodology for evaluating bioinformatics visualizations. *IEEE Transactions Visualization and Computer Graphics*, 11(4): 443–456.

29 Saraiya, P., North, C. and Duca, K. (2005b) Visualizing biological pathways: Requirements analysis, systems evaluation and research agenda. *Information Visualization*, 4: 1–15.