Temporal Thumbnails: Rapid Visualization of Time-Based Viewing Data

Michael Tsang, Nigel Morris, Ravin Balakrishnan
Department of Computer Science
University of Toronto
www.dgp.toronto.edu
mtsang | nmorris | ravin @dgp.toronto.edu

ABSTRACT

We introduce the concept of the *Temporal Thumbnail*, used to quickly convey information about the amount of time spent viewing specific areas of a virtual 3D model. Temporal Thumbnails allow for large amounts of time-based information collected from model viewing sessions to be rapidly visualized by collapsing the time dimension onto the space of the model, creating a characteristic impression of the overall interaction. We describe three techniques that implement the Temporal Thumbnail concept and present a study comparing these techniques to more traditional video and storyboard representations. The results suggest that Temporal Thumbnails have potential as an effective technique for quickly analyzing large amounts of viewing data. Practical and theoretical issues for visualization and representation are also discussed.

Categories and Subject Descriptors: I.3.6 [**Computer Graphics**]: Methodology and Techniques – *Interaction Techniques*; H.5.2 [**Information Interfaces and Presentation**]: User Interfaces – *Interaction styles*. **General Terms**: Design, Algorithms

Keywords

Visualization, viewing analysis, temporal thumbnail, representation refinement.

1. INTRODUCTION

Obtaining and understanding information about how people view something, such as a new consumer product, can be very beneficial. For example, even simple information like knowing how long someone spent looking at a certain part of a product could be very valuable. Traditionally, such information has been collected via techniques such as focus groups, surveys, and in-person or videotaped observations of people examining new product offerings. While these tried and tested methods are undoubtedly useful, the increasing use of the world wide web for disseminating product information opens up a very promising avenue for collecting immense amounts of data on how potential customers view these products.

AVI '04, May 25-28, 2004, Gallipoli (LE), Italy

For example, a manufacturer could showcase a virtual model of a new product on the web, and collect data on all virtual camera movements – in both space and time – made by those who viewed the model. Automobiles are frequently displayed in this manner.

While collecting such data is relatively easy, subsequent analyses and visualization of the data is non-trivial. Existing techniques that work well for small datasets are not likely to scale to larger quantities. For example, determining how much time people spend examining different parts of objects becomes very difficult as the length of time and number of people increase. The common technique of analyzing video footage of users examining objects can take linear time in relation to the length of the recording. Sifting through the hundreds of hours of video that an online collection system will produce would be rather onerous, while describing and uniquely identifying similar content can be very difficult. Yet, an online collection process may be valuable as a potential treasure trove of information. Thus, it is critical that we develop appropriate techniques for quickly and accurately analyzing such viewing data.

In this paper we develop, and evaluate techniques that distil viewing information into an easily understood format that may help to address the analysis of how people view models. In particular, we focus on 3D models and scenes. Our techniques have the goal of extracting the characteristic impression of a viewing session by mapping the time dimension into a unique *Temporal Thumbnail*. Temporal Thumbnails are interactive sketches that are time-and-space sensitive and are representative of interactive virtual model examination sessions, similar to how traditional thumbnail images can quickly convey a low-fidelity sketch of static two-dimensional images.

2. RELATED WORK

Summary and Compression. There are several research projects that have investigated the problem of making large amounts of information easier to handle and understand by extracting and presenting the information in a meaningful but condensed way. Some systems summarize video using text annotations or storyboard pictures [8]. Ueda et al. [15] use semantic techniques to automate structure extraction from video, and include a moving icon representation. Other techniques have attempted to reduce the time taken to process recorded information by compression of audio or video [1, 7, 13].

Time and interaction visualization. Stoev and Straber [14] present a case study of visualizing historical data. They allow users to interactively examine the spatial and temporal components of recorded data with control over the time increments and camera flythroughs. This technique, however, is time intensive since large datasets will require considerable interactive exploration. VisVIP [5] represents web site traversal as directed splines connecting nodes. Other work explores visualization using glyphs to represent data and utilizes color and density to display information to users [9-12]. Chi [3] describes a technique for visualizing web site usage using a branching tree structure, where each edge corresponds to user navigation. The edges of the tree change color and thicken to reflect increased traffic. Healey and Enns [12] describe a system for mapping environmental parameters to glyphs for users to visualize spatial data. They present an example of tracking typhoons, where they show how glyph density, regularity and size can be utilized to represent the wind speed, pressure and precipitation due to the typhoon.

Eye Tracking. Knowledge of where a user is looking has been used to measure attention given to different parts of an interface [4]. Vertegaal et al. [16] present a video conferencing system that utilizes eye tracking information to direct user viewpoints toward the targets of their conversations within a virtual conference room. There also have been systems designed to adapt the level of display detail to a user's gaze, for example [2]. DeCarlo and Santella [6] describe a system for non–photorealistic painting of images based on users' perception. Their system tracked eye fixations and displayed a finer level of detail within those regions of interest. Our system similarly assumes that attention is related to viewing information, and uses viewpoint information to visualize attention.

In contrast with much of this previous work, our Temporal Thumbnails are effective only in a static, well-defined domain. However, they have the advantage of requiring no specialized extraction algorithms, semantic understanding, or extensive customization. Our system takes the approach of summarization by collapsing the temporal dimension into a spatial representation overlaid onto the subject of the recording. This approach takes advantage of a well-defined domain, availability of spatial information, and locality of interaction using glyphs and color to represent aggregate temporal and spatial data.

3. TEMPORAL THUMBNAIL DESIGNS

One can imagine using a recorded session of users examining a scene to reconstruct a characteristic picture of the viewing interaction. We take the recording and reconstruct the time component as a visual representation on the corresponding parts of the 3D scene. Just as a thumbnail image is a rough approximation of a full resolution picture, this reconstructed representation is meant to show a characteristic approximation of the time-based interaction. We term this reconstruction a "Temporal Thumbnail" and have developed three different visualization techniques that demonstrate this concept: the *Camera Glyph*, the *View Intersection Glyph*, and the *Temperature Map.*

3.1 Camera Glyph

Our first technique, the camera glyph*,* presupposes that knowing the position of the viewer (i.e., location of the virtual camera in 3D space) is helpful for analysis. Time is represented as dots placed in the scene at the camera's position at each time-step. As shown in Figure 1-top, we can see the position of a viewer at different times during the examination. If the viewer dwells at the same position, the glyph becomes more opaque at each time step. Assuming an object-centered view, we can determine roughly what part of the model is being examined by projecting the view from the glyph into the center of the model. The density of the glyphs in a particularly area provides a rough indication of the importance of that area.

When we have a lot of data represented by any kind of glyphs, occlusion may occur. To reduce the effect of obscuring clouds of camera glyphs, we allow the user to adjust their transparency by using a slider. This allows the user to manipulate the representation to best fit the task and the data set being viewed.

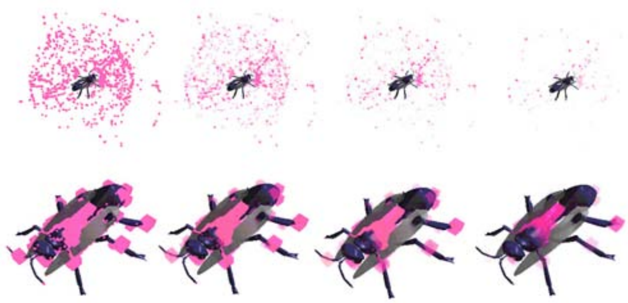

Figure 1. (top) Camera glyphs show the position of the viewer in 3D space as a function of time spent at that location. (bottom) View intersection glyphs represent via highlights on the model itself the relative time spent looking at that location. The transparency level for both techniques is increased in the glyphs from left to right to reveal progressively more of the underlying model.

3.2 View Intersection Glyph

Our second technique, the view intersection glyph, is generated by projecting a ray from the center of the viewing camera into the scene and depositing a glyph at the intersection of the ray and the 3D model being viewed. This operation is performed at each time-step, creating a layer of glyphs on the *surface* of the model (Figure 1-bottom). If the viewer dwells for a long time at the same position larger glyphs are created, while fleeting movements across the model result in smaller glyphs.

As with the previous technique, a slider adjusts the transparency of the glyphs and enables the user to see densest areas of the visualization, while adjusting for the problems that occur when too many glyphs obscure the view of the underlying model.

The application of view intersection glyphs assumes that users center the view close to the subject of interest. This may not always be true, and also may be problematic if the model has multiple long and thin protrusions.

3.3 Temperature Map

The process of constructing the temperature map is conceptually similar to holding a strong spotlight to 'heat up' the model geometry during examination, and then viewing the resulting temperature gradient. To implement the temperature map, we recreate the view and identify the visible triangles of the model at each time-step. We then increase the attention score of these triangles based on their distance from the centre of the screen. We used the Gaussian distance function, which gives a strong peak around the centre of the screen and tails off smoothly to the edges, which roughly approximates the amount of interest for each part of the screen as the object is being examined.

In order to represent this data we mapped the attention scores onto the surface of the model by coloring the geometry using a simple, three-level temperature metaphor. This technique takes further advantage of the pre-attentive nature of hue [11], allowing differences to be immediately apparent. Red 'hotspots' represent areas of strong interest, green areas represent a moderate interest, and blue shows the 'coldest' or least examined areas (Figure 2).

Figure 2. (left) An aggregation of virtual spotlights highlights the area of interest on the model (right) The temperature map shows those same areas using a colored spectrum.

3.4 Summary of the Three Designs

Camera Glyphs and View Intersection Glyphs represent time as objects in the scene. Strengths of these two techniques include speed and precision. Weaknesses of these glyph techniques include the loss of time ordering information and the potential to obscure the view. Specific to the floating Camera Glyphs are problems with associating each glyph to the corresponding part of the model. View Intersection Glyphs, on the other hand, may suffer from the assumption that the center of the view is the precise area of interest.

The Temperature Map represents time as a property of the model itself. It has the advantage of speed, but the disadvantage of losing ordering information. Another disadvantage is that the Temperature Map representation replaces the model's true colors.

4. EVALUATION

We performed a usability evaluation of these three Temporal Thumbnail techniques. In addition to comparing the three designs amongst themselves, we also included two existing techniques in the evaluation: *video analysis*, and *storyboards* (Figure 3). Video analysis is high-fidelity, and preserves both audio and visual modes of information. The storyboard representation lays out visual stills of the video, which allows for non-linear navigation, but with a reduced visual fidelity.

Through the evaluation, we differentiated the five different techniques in terms of speed, accuracy, and user confidence. (See Figures 4 and 5). Details of this usability study can be found at www.dgp.toronto.edu/research/temporalthumbnails.

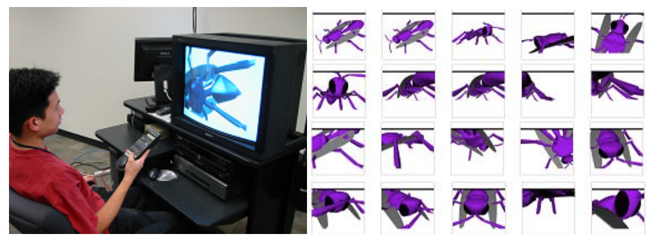

Figure 3. Traditional techniques: (left) Simple video analysis. (right) Storyboards.

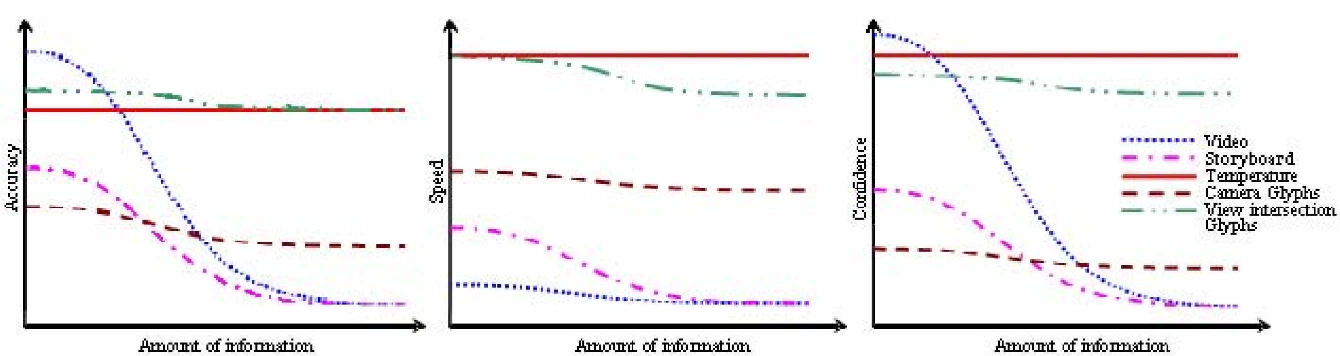

Figure 4. Predicted (from left to right) accuracy, speed, and confidence for the five visualization techniques, measured against the amount of viewing data being summarized.

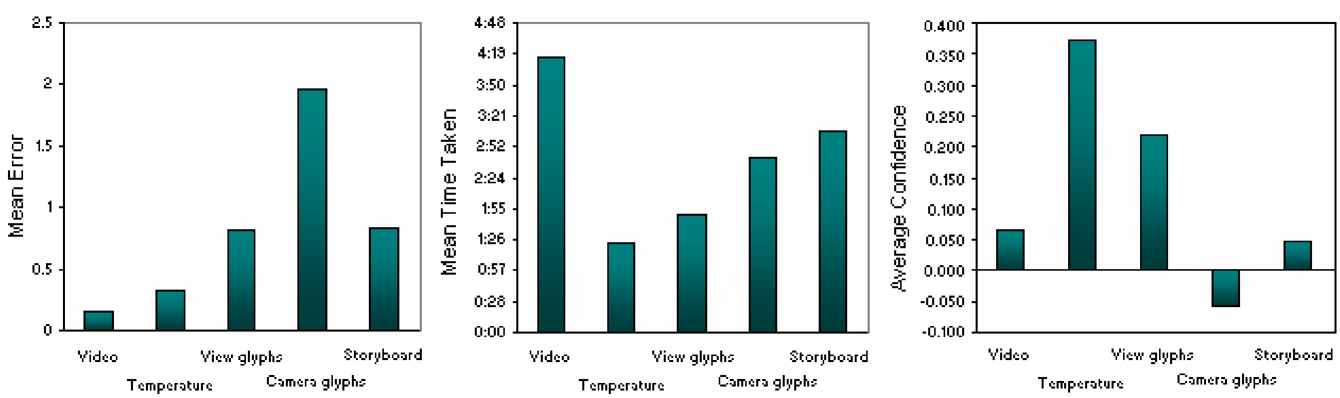

Figure 5. Evaluation results.

5 DISCUSSION

Video is the most flexible and sensory rich medium that we evaluated, with characteristics not available in most of the other visualizations. Perceptual advantages include tone and content of audio, mouse-cursor movements, deictic gestures, and time-based context and ordering information. However, although time in video can be rescaled by compression [13], it nevertheless requires review time proportional to the amount of information. This leads to memory problems and content overload, with the additional difficulty of sifting through large recordings.

In contrast, the storyboard technique presents time through the layout of information in space. Study participants consistently reported that the large amount of information simultaneously presented by the storyboard feels "visually overwhelming" and made it difficult to remember and sort information.

In both Camera and View Intersection Glyphs, time is represented as a scene object. This has the advantage of precise location, but also has the possible disadvantage of obscuring the view. By projecting the viewpoint onto the model, View Intersection Glyphs were more conducive to our particular task than the Camera Glyphs. This is evidenced by the strong difference in both confidence and speed. Our temperature map technique represents time as the color property of the model geometry. In addition to less clutter, this method makes use of previous research results that show that hues can be pre-attentively processed [11]. The temperature method is fast, but there are many parameters such as the Gaussian 'spotlight' width and intensity that may be confusing if exposed to the user.

6 CONCLUSIONS AND FUTURE WORK

We have shown how aggregate time varying viewing data can be represented using an interactive Temporal Thumbnail which a user can rapidly interpret with reasonable accuracy. We have presented three different designs that implement the Temporal Thumbnail concept. A usability evaluation indicated that these designs enable faster analysis with greater confidence than existing video analysis and storyboard techniques. While accuracy was not improved by our techniques, there was no degradation either. The results suggest that we have succeeded in refining the representation for the task and offloading cognitive load associated in dealing with time-based data. We also discussed strengths and weaknesses of each technique, and suggest a method of projecting properties to other media in order to refine a representation towards a specific task.

7 ACKNOWLEDGMENTS

We thank Mike Wu, Glen Tsang and the members of the Dynamic Graphics Project lab at the University of Toronto for their assistance and advice throughout the course of this work, and to all those who participated in our user study.

8 REFERENCES

1. Arons, B. (1997). SpeechSkimmer: a system for interactively skimming recorded speech. *ACM Transactions on Computer Human Interaction*. 4(1). p. 3-38.
2. Baudisch, P., DeCarlo, D., Duchowski, A., & Geisler, B. (2003). Focusing on the essential: Considering attention in display design. *Communications of the ACM*. 46(3). p. 60-66.
3. Chi, E. (2002). Improving usability through visualization. *IEEE Internet Computing*. p. 64-71.
4. Crowe, E., & Narayanan, H. (2000). Comparing interfaces based on what the users watch and do. *Symposium on Eye Tracking Research and Applications*. p. 29-36.
5. Cugini, J., & Scholtz, J. (1999). VisVIP: 3D visualization of paths through web sites. *IEEE WebVis International Workshop on Web-Based Information Visualization*. p. 259-263.
6. DeCarlo, D., & Santella, A. (2002). Stylization and abstraction of photographs. *ACM SIGGRAPH Conference on Computer Graphics and Interactive Techniques*. p. 769-776.
7. He, L., & Gupta, A. (2001). Exploring benefits of non-linear time compression. *ACM International Conference on Multimedia*. p. 382-391.
8. He, L., Sanocki, E., Gupta, A., & Grudin, J. (1999). Auto summarization of audio-video presentations. *ACM International Conference on Multimedia*. p. 489-498.
9. Healey, C. (1996). Choosing effective colours for data visualization. *IEEE Visualization*. p. 263.
10. Healey, C. (1998). On the use of perceptual cues and data mining for effective visualization of scientific datasets. *Graphics Interface*. p. 177-184.
11. Healey, C., Booth, K., & Enns, J. (1996). High-speed visual estimation using preattentive processing. *ACM Transactions on Human Computer Interaction*. 3(2). p. 107-135.
12. Healey, C., & Enns, J. (1998). Building perceptual textures to visualize multidimensional datasets. *IEEE Visualization*. p. 111-118.
13. Omoigui, N., He, L., Gupta, A., Grudin, J., & Sanocki, E. (1999). Time compression: systems concerns, usage, and benefits. *ACM CHI Conference on Human Factors in Computing Systems*. p. 136-143.
14. Stoev, S., & Straber, W. (2001). A case study on interactive exploration and guidance aids for visualizing historical data. *IEEE Visualization*.
15. Ueda, H., Miyatake, T., Sumino, S., & Nagasaka, A. (1993). Automatic structure visualization for video editing. *ACM CHI Conference on Human Factors in Computing Systems*. p. 137-141.
16. Vertegaal, R., Weevers, I., & Sohn, C. (2002). GAZE-2: An attentive video conferencing system. *Extended Abstracts of ACM CHI Conference on Human Factors in Computing Systems*. p. 736-737.