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
H5.2. User Interfaces.

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. In the manufacturing and marketing world, for example, even simple information like knowing how long someone spent looking at a certain part of a product could be very valuable since it would at least partially indicate the importance of that part in the overall product's design. 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. 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[1].

[1] For an example, please see the Toyota web site at http://www.toyota.com/vehicles/2004/camrysolara/ext360.html

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, it is indeed the potential treasure trove of information contained in these large quantities of data that make such an online collection process so valuable. Thus, it is critical that we develop appropriate techniques for quickly and accurately analyzing such viewing data.

In this paper we develop, implement, and evaluate techniques that distill 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 [10]. Ueda et al. [19] 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, 9, 15].

Time and interaction visualization. Stoev and Straber [17] 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 [11-14]. 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 [14] 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. [20] 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

The concept of creating Temporal Thumbnails can be explained with the metaphor of an individual examining a teapot by holding it in their hands and rotating and manipulating it to view its various parts. Imagine that while performing this examination, the individual concurrently breathes at a constant rate, and little droplets from their breath falls onto the teapot. Assuming a direct and consistent accumulation of droplets, after the examination is finished one may examine the density of droplets deposited on the teapot to determine the approximate amount of time the individual spent examining any particular part of the teapot.

Applying this idea to virtual 3D scenes, 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 1a, we can see the position of a viewer at different times during the examination. If the viewer dwells at the sme 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.

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 1b). 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

Our third technique, the temperature map, addresses the problems associated with the single point of interest of the View Intersection Glyph. It is based on the idea that the focus of the user's interest lies near the centre of the screen and diminishes toward the edges. We achieve this by associating an *attention score* with each triangle of the model, which represents how much interest each triangle received from the user's navigation. We also wanted to be able to maintain a clear view of the model without any glyph obstruction and yet still represent the areas of interest. Our solution is to color the model's geometry based on these attention scores. Although this coloration may obscure texture details, a simple toggle allows users to swap between the temperature map and the natural texturing.

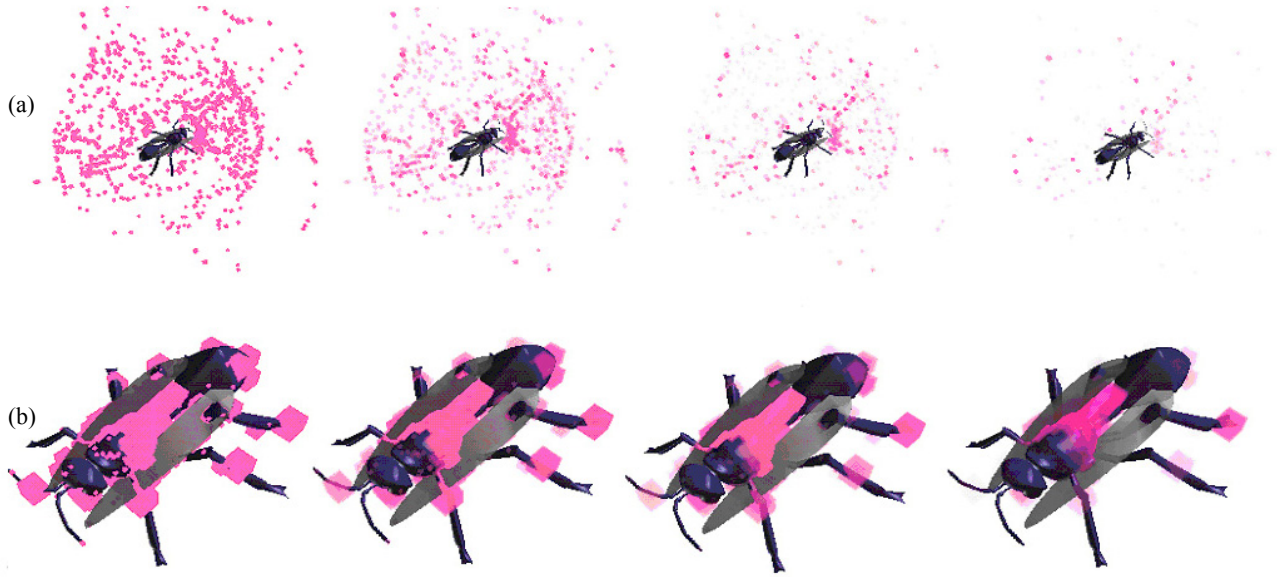

Figure 1. (a) Camera glyphs show the position of the viewer in 3D space as a function of time spent at that location. (b) View intersection glyphs represent via highlights on the model itself the relative time spent looking at that location. The transparency level for both techniques is increased in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

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 [13], 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).

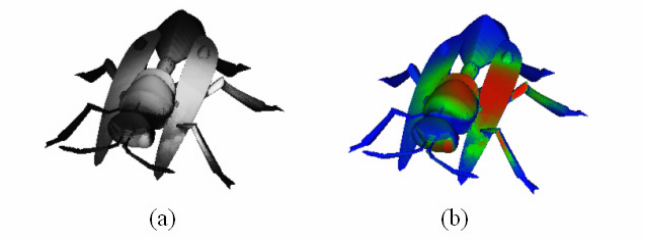

Figure 2. (a) An aggregation of virtual spotlights highlights the area of interest on the model . (b) The temperature map shows those same areas using a colored spectrum.

In order to map our attention scores to the color scale we normalized the attention scores with a sigmoid function, mapping the 95th percentile of accumulated attention score to a point on the sigmoid curve. The exact mapping of the curve was controlled by an interactive slider, allowing for an increase or decrease of the 'temperature' contrast setting.

3.3 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 this evaluation, we hope to differentiate the five different techniques in terms of speed, accuracy, and user confidence [8].

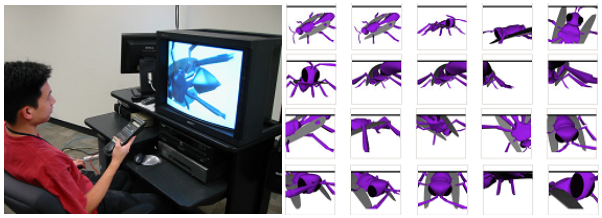

Figure 3. Traditional techniques: (left) Simple video analysis. (right) Storyboards.

4.1 Apparatus

All the visualization techniques used in the study were implemented in Java3D™ and OpenGL / C++, and ran on a Pentium 4, 2.4 GHz, processor running Microsoft Windows XP equipped with an ATI Rage128™ graphics card, connected to a 20" display. For the video analysis technique, we used a standard VCR with jog-shuttle controls connected to a 20" video monitor.

4.2 Participants

6 women and 5 men, recruited from the university community, participated in the study. Six were used in the data generation phase, and five in the analysis phase. None of the participants had prior experience in analyzing visualization data for 3D scenes.

4.3 Content Acquisition for Later Analysis

In order to conduct a user study we needed useful visualization data for our participants to analyze. This data was collected from six participants who were presented with a simple 3D model inspection interface with standard camera controls that allowed them to tumble, pan, and zoom in on the 3D model. Users were first asked to familiarize themselves with the camera navigation controls. Then, we presented them with a 3D model and told them to examine it for a minute. They were also told that after the one minute examination, they would be asked to recall certain details. This instruction ensured that participants inspected the model carefully; paying close attention to areas they thought could be important. This effectively simulates the real-world situation of, for example, someone examining a model of a product they were evaluating for possible purchase. After the minute elapsed, the model was then hidden and the participants answered questions based on the model. We then allowed them access to the model again, and they were asked to verbally correct their answers as we noted them down. This procedure created visualization data with a variety of styles, from overviews in the initial familiarization phase, to in depth examinations of various areas of the model when asked specific questions about the model in the second phase. Each participant examined the same model (Figure 4) which was chosen since it has the "boxy enclosure" shape of many consumer products.

Figure 4. Tractor model used in evaluation.

Throughout the process we recorded the users' interaction. We captured a videotape session of the examination sessions, including audio data, and recorded the virtual camera location and orientation every ten milliseconds. We also captured the session display screen as a bitmap every two seconds in order to produce a storyboard representation of the session. In total, 12 recordings were made: 6 participants x 2 inspection phases of the same model. After data collection was complete, we created 4 data streams as follows:

- *IndividualPhase1 (I1)*: the data from one of the 6 participants (randomly selected) in the first one minute overall examination of the model
- *IndividualPhase2 (I2)*: the data for the same individual, for the second more detailed examination of the model.
- *AggregatePhase (A1)*: the data from all 6 participants combined, for the first one minute examination.
- *AggregatePhase2 (A2)*: the data from all 6 participants combined, for the second examination phase.

For each of these data streams, we generated visualizations for each of the five visualization techniques.

4.4 Evaluation Procedure

In the analysis phase, five participants examined the data previously generated by participants in the content acquisition phase, using our five visualization techniques. A within-subjects design had participants using all five techniques; with between subjects presentation order counterbalanced using a balanced Latin square. At the start of each new technique session, participants were given a short introduction and allowed to familiarize themselves with the controls for that technique. We then presented each participant with the four sets of data for that technique, in a random order within each visualization technique.

For each dataset examined with each of the five visualization techniques, participants completed a questionnaire (Figure 5). They were asked to rank different parts of the model based on their understanding from the visualization of how much attention the users who generated the data paid to those parts. We asked them to rate their confidence in their answers, and recorded the time taken to answer the questions.

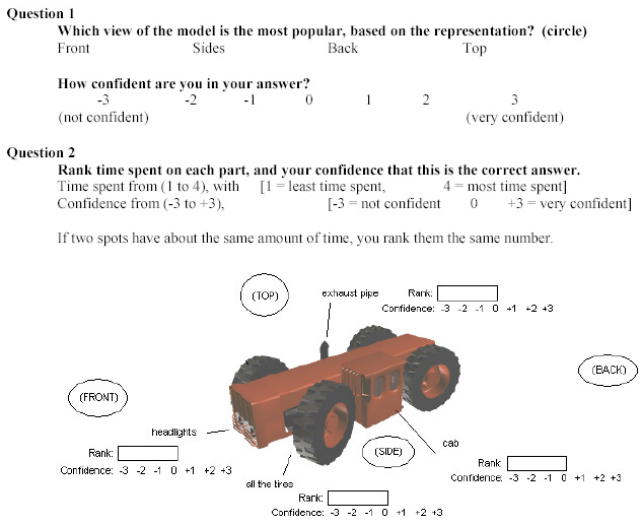

Figure 5. Questionnaire filled in for each of four data streams, for all five visualization techniques.

4.5 Hypotheses

Our hypotheses regarding the expected relative performance of the five techniques were as follows:

Video analysis: This technique inherently provides more information to the analyst user than other techniques. The audio narrative and the mouse movement give additional deictic information that is not available in the other visualizations. The video also gives a temporal ordering that is not available with other techniques; thus we expect the video to be more effective for short sessions. On the other hand, the linear time and serial nature of video analysis would reduce the effectiveness of video when large amounts of data were processed. We predict users would be overwhelmed by the data as time progressed, and forget details seen early in the video stream, requiring frequent rewinding and replaying of previously viewed content.

Storyboard: We expect the storyboard to exhibit similar properties to video analysis, but without the benefit of audio. The nature of the storyboard should allow people to quickly navigate the picture layout, with the number of pictures proportional to the recorded session length. On one hand we expect slightly better accuracy and user confidence in their results when compared to video, due to the ease of quick view summarization and navigation inherent in the parallel viewability of storyboard frames. However, since the storyboard frames were captured at a two second interval, we suspect that some details would be overlooked due to the sampling granularity.

Camera glyphs: We predict that users would have more difficulty mapping the camera glyphs to parts of the model. We also expected the glyphs themselves would be a problem in that they could obscure the user's view of the model. Users would likely have difficulty associating glyphs with the model since the projection is not immediately obvious, impacting on the confidence in this technique. On the other hand the camera glyphs were expected to provide a good indication of the preferred viewing positions, such as from the top, from behind, or from the sides of the model.

View intersection glyphs: We predict that the view intersection glyphs would be superior to the floating camera glyphs. First, view intersection glyphs obstruct the view less than the camera glyphs; and having glyphs placed directly on the model should provide the users with greater precision for identification tasks. The weakness of the view intersection glyph is that the single point of interest in the center of the screen may be misleading in some circumstances.

Temperature map: We expect the temperature map to be one of the most accurate techniques, as it addresses shortcomings of the view intersection glyph. We also predict that users would appreciate the simple format provided by this visualization, resulting in a high user confidence, as this technique is the least cluttered and has the least ambiguity. We predict that users would be able to analyze the visualization swiftly and accurately, even with large amounts of information.

Figure 6 summarizes our hypotheses, in terms of our measures of speed, accuracy, and user confidence.

4.6 Performance Measures

We used three performance measures. Speed was measured as the time taken to answer all parts of Question 2 in the questionnaire (Figure 5). Confidence was rated on a scale from -3 to +3 where -3 corresponded to 'not confident' and +3 to 'very confident'. Accuracy was represented by an error score for each of the rankings, computed by adding the time difference needed to reconcile the true and erroneous rank times for each place. For example, if the correct order is A,B,C and the ranked order is B, C, A the error attributed to A is |A-B| + |A-C|.

4.7 Results

Analysis of variance showed a significant main effect for speed (F = 15.59, $p < .0001$) and confidence (F = 28.34, $p < 0.0001$). Pairwise means comparisons indicated that our hypotheses for speed and confidence were correct, with all three Temporal Thumbnail techniques significantly faster than the video analysis and storyboard techniques. Participants were most confident of their results with the Temperature Map technique, followed in order by View Intersection Glyphs, video analysis, storyboards, and Camera Glyphs. Figure 7 illustrates these results.

Surprisingly, we did not find a significant difference in the accuracy of the techniques (F = 1.90, p = 0.12). We suspect that our input data streams were not of sufficient length to differentiate the techniques, as users were able to make multiple passes through the videotape for questions that were not obvious at first pass. In contrast, we believe the accuracy will be closer to our predictions as the length of recordings increase due to the fatigue of viewing linearly proportional information with the standard visualization techniques. Technique order and recording order did not statistically impact the measures.

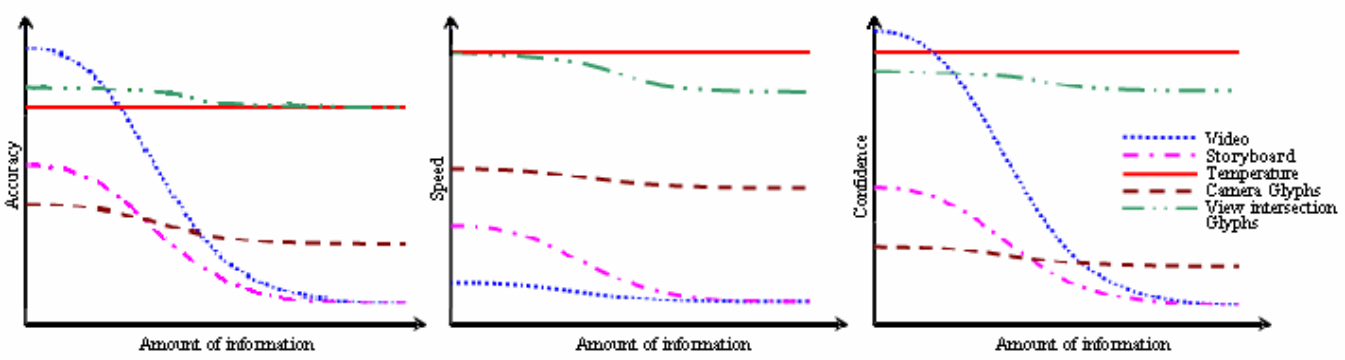

Figure 6. Predicted (from left to right) accuracy, speed, and confidence for the five visualization techniques, measured against the amount of viewing data being summarized.

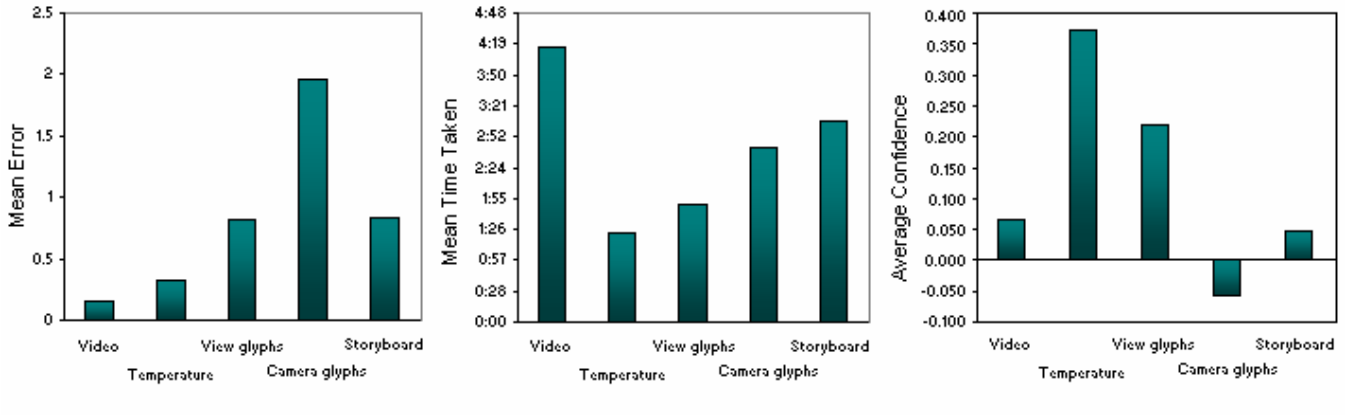

Figure 7. Evaluation results.

Qualitatively, the users reported that the video provided helpful audio cues. On the other hand, they noted it very difficult to judge the aggregate interest level, as it was difficult to recall past information for lengthy sequences, and thus felt that video was most useful for short sessions. The users also were frustrated having to estimate viewing information when using the storyboard representation. Some users tried to make this task easier by manually grouping sets of storyboard images, while others scanned through all the images multiple times, trying to get a sense of the overall viewing session. Most users felt that there was too much information in the storyboard making it difficult to determine aggregate information.

User's comments for the Temporal Thumbnail techniques differed depending on the technique. Camera Glyphs were reported to be useful for general overviews, but users had difficulty associating the Camera Glyphs with specific parts of the model. Consequently, as shown in Figure 7, the Camera Glyphs inspired the least confidence for our evaluation. In contrast, it was felt that View Intersection Glyphs were comparatively accurate, although not quite as effective as the Temperature Map technique, which was reported to be the most straightforward visualization technique for rapid evaluation. Users especially appreciated the automatic aggregation of the data in this technique.

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 [15], 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 [13]. 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.

5.1 Limitations

One assumption of our evaluation is that the viewer's interest is roughly focused at the center of the screen. Although this may seem to be a questionable assumption, our results suggest that the decision does not result in reduced accuracy when compared to the feature-rich video analysis. We acknowledge that the mouse navigation (rotate, pan, zoom) used in our study may not the best input choice to navigate around 3D objects and we believe that a more natural means of navigation around 3D objects such as in [18] would achieve even better results.

It is also important to acknowledge that we are not comparing our techniques against eye-tracking techniques. We believe that with minor adjustments, Temporal Thumbnails can be used in conjunction with eye-tracking techniques, which would result in more accurate analysis. However, we decided against using an eye-tracking solution since our system would be the most useful in the hands of the many users. For example, to aggregate data from a large number of users who may be off-site, such as over a standard internet browser with mouse input. It would be impractical to expect all the users to possess the necessary eye-tracking equipment.

5.2 Implications

Overall, the most efficient and confidence instilling techniques collapse time directly onto the model. This contrasts with the other representations that have a larger cognitive load imposed upon the user. For example, in the task of identifying the area with the most attention, the area of the interest must be extracted from each frame of a storyboard representation, and scores mentally computed by the user. This increased cognitive requirement contrasts with the temperature map which automatically aggregates the information and uses the pre-attentive processing property of color to identify areas of the most interest.

This suggests that the process of refining representations with respect to a specific task can be important. Unrefined representations such as the video may contain more 'raw' information, but refining these representations can distill the information towards a specific purpose. Although some information may be lost during refinement, the resulting visualization is left better suited for the task at hand.

In our example, the task is to estimate the amount of attention given to a specific area. For each successive technique: video analysis, storyboard, Camera Glyph, View Intersection Glyph, and Temperature Map; we increasingly refine the information towards this task. Mapping time to space via a concrete object on the scene increases the precision of the task. View intersection glyphs provide a tight coupling between task and visualization by projecting information directly on the model. The Temperature Map provides the strongest binding between task and representation by using the properties of the model itself and also eliminates the much of the visual clutter associated with having objects in the scene. It is apparent that refining information can result in a much more efficient representation, as the Temperature Map is the most refined towards the task. As long as the refining is accurate, this makes for the fastest, most accurate and confidence inducing visualization.

One can refine representations in a more general sense by determining what characteristics are most appropriate for the task and then work towards creating those characteristics in the representation. One way to accomplish this is by transforming one property to a different domain. When a property is mapped from one domain to another, affordances of the target medium are projected onto the information. For example in the Storyboard representation, time is represented as space and this can create problems associated with spatial content management, yet it also enables the quick navigation associated with spatially embedded content. Thus in refining representations for efficient visualization tasks, there must be a congruency between the representation and the task, as hinted in [21]. Simon [16] writes that the problem of representation is the task of making the solution salient. Congruency between task and representation helps accomplish this by creating a visualization which we are able to more easily understand since we offload processing from the user through representation refinement.

5 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.

The techniques presented here have potential in other domains, especially in document versioning where the user's attention is focused by the application tools. Using variations of the Temperature Map, one could easily allow users to grasp the areas of the document that differ between versions, for both 2D and 3D media assets.

One exciting avenue of exploration is the use of spatially-aware displays [7, 18] to generate Temporal Thumbnails. Spatially-aware displays enable easy navigation of 3D space using a very direct metaphor, and we believe that the efficiency and effectiveness of the Temporal Thumbnail concept will be increased dramatically if such methods of navigation are used to inspect the virtual models.

6 ACKNOWLEDGMENTS

We thank 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.

7 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., *et al.* (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. Fitzmaurice, G.W. (1993). Situated information spaces and spatially aware palmtop computers. *Communications of the ACM*. 36(7). p. 38-49.
8. Frøkjær, E., *et al.* (2000). Measuring usability: are effectiveness, efficiency, adn satisfaction really correlation? *ACM CHI Conference on Human Factors in Computing Systems*. p. 345-352.
9. He, L., & Gupta, A. (2001). Exploring benefits of non-linear time compression. *ACM International Conference on Multimedia*. p. 382-391.
10. He, L., *et al.* (1999). Auto summarization of audio-video presentations. *ACM International Conference on Multimedia*. p. 489-498.
11. Healey, C. (1996). Choosing effective colours for data visualization. *IEEE Visualization*. p. 263.
12. Healey, C. (1998). On the use of perceptual cues and data mining for effective visualization of scientific datasets. *Graphics Interface*. p. 177-184.
13. Healey, C., *et al.* (1996). High-speed visual estimation using preattentive processing. *ACM Transactions on Human Computer Interaction*. 3(2). p. 107-135.
14. Healey, C., & Enns, J. (1998). Building perceptual textures to visualize multidimensional datasets. *IEEE Visualization*. p. 111-118.
15. Omoigui, N., *et al.* (1999). Time compression: systems concerns, usage, and benefits. *ACM CHI Conference on Human Factors in Computing Systems*. p. 136-143.
16. Simon, H. (1969). The sciences of the artificial. MIT Press. .
17. Stoev, S., & Straber, W. (2001). A case study on interactive exploration and guidance aids for visualizing historical data. *IEEE Visualization*.
18. Tsang, M., *et al.* (2002). Boom chameleon: simultaneous capture of 3D viewpoint, voice and gesture annotations on a spatially-aware display. *ACM UIST Symposium on User Interface Software and Technology*. p. 111-120.
19. Ueda, H., *et al.* (1993). Automatic structure visualization for video editing. *ACM CHI Conference on Human Factors in Computing Systems*. p. 137-141.
20. Vertegaal, R., *et al.* (2002). GAZE-2: An attentive video conferencing system. *Extended Abstracts of ACM CHI Conference on Human Factors in Computing Systems*. p. 736-737.
21. Williams, D. (1996). Multimedia, mental models, and complex tasks. *Extended Abstracts of ACM CHI Conference on Human Factors in Computing Systems*. p. 65-66.