Chapter 31
3D Thumbnails for 3D Videos with Depth

Yeliz Yigit
Bilkent University, Turkey

S. Fatih Isler
Bilkent University, Turkey

Tolga Capin
Bilkent University, Turkey

ABSTRACT

In this chapter, we present a new thumbnail format for 3D videos with depth, 3D thumbnail, which helps users to understand the content by preserving the recognizable features and qualities of 3D videos. The current thumbnail solutions do not give the general idea of the content and are not illustrative. In spite of the existence of 3D media content databases, there is no thumbnail representation for 3D contents. Thus, we propose a framework that generates 3D thumbnails from layered depth video (LDV) and video plus depth (V+D) by using two different methodologies on importance maps: saliency-depth and layer based approaches. Finally, several experiments are presented that indicate 3D thumbnails are illustrative.

INTRODUCTION

Today, the popularity of 3D media usage in computerized environment and the research on 3D content generation is increasing. 3D contents are frequently used in various applications such as computer games, movies and even in home environmental systems and this reputation leads visualization of 3D contents such as 3D videos and 3D images becoming more significant. For visualizing the 3D contents, thumbnail representation is used in order to provide a quick overview of multimedia files in order to allow a quick scanning over a large number of data. By using the traditional methods, the thumbnail generally shows the first frame of the video and for images, visual representation is generated by using shrinking, manual cropping or uniform scaling. However, these approaches do not preserve the important parts of the multimedia files and resulting thumbnails do not give the general idea of the content. Furthermore, in spite of the existence of

DOI: 10.4018/978-1-61350-326-3.ch031

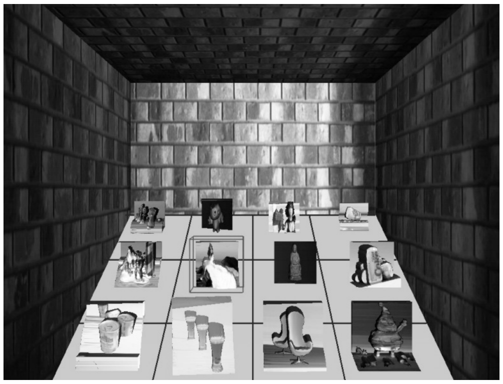
Copyright © 2012, IGI Global. Copying or distributing in print or electronic forms without written permission of IGI Global is prohibited.

3D content databases, there is no standardization on the thumbnail representation for 3D contents while their usage area is widespread. Thus, the thumbnail representation is very crucial to get a quick overview of the content rather than downloading from the database and processing it.

Therefore, we propose a thumbnail generation system that creates meaningful, illustrative visual representations of 3D video with depth contents without losing perceivable elements in the selected video frame by using saliency-depth and layer based methodologies. Moreover, in order to represent the 3D contents realistically and enhance depth perception, the resulting thumbnail should be in 3D. Thus, the framework constructs geometries of important objects as polygon meshes and adds 3D effects such as shadow and parallax mapping. Figure 1 illustrates a layout that holds resultant 3D thumbnails for 3D videos with depth.

While creating 3D thumbnails, it is required to select suitable 3D video formats since compression and coding algorithms of 3D videos show diversity according to the varieties of 3D displays: classical two-view stereo video (CSV), video plus depth (V+D), layered depth video (LDV) and multi-view video plus depth (MDV). Some of these formats and coding algorithms are standardized by MPEG, since standard formats and efficient compression are crucial for the success of 3D video applications (F.Institute, 2008). For our framework, V+D and LDV formats are eligible because of simplicity and the depth information they provide. V+D format provides a color video and an associated depth map that stands for geometry-enhanced information of the 3D scene. The color video is original video itself and the depth map is a monochromatic, luminance-only video. Besides, LDV is an extension of V+D format. It contains all information that V+D satisfies with an extra layer called background layer which includes foreground objects and the associated depth map of the background layer. By using the properties of V+D and LDV videos, we develop two different thumbnail generation methods based on the information they present. These proposed methodologies create meaningful thumbnails without losing perceivable visual elements in the selected original video frame.

Figure 1. 3D thumbnails on a 3D grid layout



In this chapter, the previous work on 3D video formats and thumbnail generation methods, the proposed framework that generates 3D thumbnails from video plus depth (V+D) and layered depth video (LDV), two 3D thumbnail generation methodologies based on 3D meshes and parallax mapping, and several experiments showing effectiveness and recognizability of 3D thumbnails, are presented.

BACKGROUND

We discuss 3D video formats and thumbnail generation approaches under two different subsections since our approach combines them.

3D Video Formats

Recently, several numbers of researches on 3D imaging and video formats are rapidly progressing. 3D video formats are roughly divided into two classes: *N*-view video formats and geometry-enhanced formats. The first class represents the multi-view video with *N* views. Conventional stereo video (CSV) is the least complex and most popular format of *N*-view video for stereoscopic applications.

Otherwise, geometry-enhanced information is provided for 3D video formats in the second class. Multi-view video + depth video (MDV), layered-depth video (LDV) and video plus depth (V+D) are examples of geometry-enhanced formatted videos. As it is referred from its name, MDV has more than one view and associated depth maps for each view. This depth data is used to synthesize a number of arbitrary dense intermediate views for multi-view displays (Gundogdu, 2010). LDV is one variant of MDV, which further reduces the color and depth data by representing the common information in all input views by one central view and difference information in residual views (Müller, 2008). Besides, foreground objects are stored on the background layer in the LDV format with associated depth information. Since geometry-enhanced formats are complex and more data is stored, the disadvantage of MDV and LDV is the requirement of the intermediate view synthesis. In addition to this, high-quality depth map generation is required beforehand and errors in depth data may cause considerable degradation in quality of intermediate views. On the other hand, the special case, V+D codes one color video and associated depth map and the second view is generated after decoding.

V+D and LDV formats are appropriate for creating effective thumbnails for 3D videos with depth in order to try the efficiency of our framework for both simple and complex 3D formats. Furthermore, our thumbnail generation system uses depth information that V+D and LDV formats satisfy for generating 3D thumbnails.

In addition to this, a video frame should be selected in order to create illustrative thumbnails. There are considerable number of researches on video summarization and frame selection that are based on clustering-based (Farin, 2002), keyframe-based (Mundur, 2006), rule-based (Lienhart, 1997) and mathematically-oriented (Gong, 2002) methods. However, for our work, video summarization and frame selection issues are out of scope. Thus, we apply a saliency-based frame selection. In this case, for each frame of the 3D input video, the saliency is computed and the frame that has the highest saliency value is selected.

Thumbnail Generation

Our goal is to create thumbnails from 3D videos with depth without losing perceivable elements on the selected original frame. Thus, it is essential to preserve the perceivable visual elements in an image for increasing the recognizable features of the thumbnail. Computation of important elements and performing non-uniform scaling to image are involved in the proposed thumbnail representation. This problem is similar to proposed methodologies for image retargeting (Setlur, 2005).

Manually by standard tools such as (Adobe, 2010) and (Gimp, 2101), image retargeting can be achieved by standard image editing algorithms such as uniform scaling and cropping. Nevertheless, important regions of the image cannot be conserved with uniform scaling and cropping. Moreover, when the input image contains more than one important object, it leads contextual information lost and quality of the image degrades.

In addition to this, automatic cropping techniques based on visual attention have been proposed (Suh, 2003) which can be processed by saliency maps (Itti, 1998) and face detection (Bregler, 1998; Yow, 1998). Nevertheless, the main disadvantage of this technique is that it only performs for a single object and this leads loss of multiple features.

Another way to generate thumbnails is by using epitomes. Epitome is the miniature and the condensed version of the input image which contains the most important elements of the original image (Jojic, 2003). Despite the conservation of important elements, this method is suitable when the image contains repetitive unit patterns.

The main work behind our approach is Setlur (2005) image retargeting algorithm since it works for multiple objects by preserving recognizable features and maximizes the salient content. This method segments the input image into regions by using mean-shift algorithm, identifies important regions by a saliency based approach, extracts them, fills the resulting gaps, resizes the filled background into a desired size and pastes important objects onto it by using computed aspect ratios according to importance values of objects.

MAIN FOCUS OF THE CHAPTER

The objective behind this work is to create helpful and demonstrative 3D thumbnails for various types of 3D video formats. Since the proposed methods for generating thumbnails do not preserve the important features and do not give the idea of the content, we suggest a new thumbnail format, 3D thumbnail.

Moreover, today 3D is popular in computer games, movies and home environment applications. Besides, everything included user interfaces will be in 3D soon. Thus we have generated the resulting thumbnail in 3D because by using 3D layouts, more objects can be illustrated on the thumbnail with a realistic look. 3D contents which are represented by multiple thumbnails can be epitomized with a single thumbnail by preserving the important objects in a 3D layout. Lastly, in spite of the popularity and widespread usage of 3D content databases, there is no standardized thumbnail representation for 3D contents such as 3D videos.

In 3D content databases, the 3D contents are signified in 5 or 6 images in order to help users identify the 3D content. Instead of using several numbers of 2D thumbnails for giving information about the content, it is sufficient to use single 3D thumbnail that satisfies geometry-enhanced information.

The inputs of our system are V+D and LDV formatted 3D videos. V+D and LDV formats are suitable for our system since the associated depth maps are essential for generating 3D thumbnails. On the other hand, with the purpose of trying different thumbnail generation methods based on saliency-depth and layer information, and the efficiency of our framework over both simple and complex 3D formats, V+D and LDV formats are appropriate.

In order to create 3D thumbnails for V+D formatted videos, the first step is to segment the selected frame of the input color video into regions with the aim of finding the important regions. Then, by a saliency-depth based approach, importance map is obtained and important objects are extracted from the original frame. The resulting frame with gaps are filled by reconstructing the blanks with the same texture as the given input color frame by successively adding pixels and the frame is resized to a standard thumbnail size. Next, with the intention of generating the saliency-depth based retargeted image, the aspect ratios and positions of the important objects are determined by a constraint-based algorithm and scaled important objects are pasted on to the resized background. Finally, 3D mesh that represents the 3D thumbnail is created by using the retargeted color frame and the associated depth map.

On the other hand, the thumbnail generation algorithm for LDV is similar to the proposed approach for V+D except some steps. Firstly, as well as the input color video, the associated background layer is also segmented into regions. Secondly, instead of finding salient regions and classifying them as important objects, foreground objects on the background layer are assumed to be important objects. Apart from these steps, the remaining procedure is same as the one for V+D.

The 3D thumbnail generation methodology for V+D and LDV are explained in detail in the next section.

METHODOLOGY

System Overview

The input of our framework is the specific video frame of a 3D video with depth (Either LDV or V+D) as RGB color map and the associated depth map. For LDV, besides the input color video, the associated background layer is additional essential input for importance map extraction.

Firstly, the input color map is segmented into regions. Then, the importance map is extracted by using a saliency-depth approach for V+D formats. This step is different when the type of the content is LDV since the background layer is utilized with the purpose of importance map generation. Thus, while stating salient and foremost objects as important for V+D formats, the foreground objects on the background layer of LDV formats are important. After mapping the importance values, important objects are extracted, later to be exaggerated and the resulting gaps are filled with same texture as the given input color frame. Afterwards, the background is resized to the standard thumbnail size which has 192x192 resolutions. Then, important objects are pasted onto the resized background by a constraint-based algorithm. Finally, we apply two different methods for the resultant 3D thumbnail: 3D mesh-based and parallax-mapped techniques.

Image Segmentation

In order to find the objects on the color map and assign their importance values, it is necessary to segment the color map into regions. There are three proposed image segmentation methods: mean-shift (Comaniciu, 2002), graph-based (Felzenszwalb, 2004) and hybrid segmentation (Pantofaru, 2005). These three approaches are evaluated in the work of Pantofaru (2005) by considering correctness and stability of the algorithms. According to the results, both the mean-shift and hybrid segmentation methods create more realistic segmentations than the graph-based approach with a variety of parameters and both of the methods are stable. Since the hybrid segmentation algorithm is the combination of mean-shift and graph-based segmentation, it is more computationally expensive. Thus, we have preferred the mean-shift algorithm for its power and flexibility of modeling.

In Computer Vision, the mean-shift segmentation has a widespread usage. This algorithm takes spatial radius h_s, color radius h_r and the minimum region area M as parameters with the input color map. In this algorithm, the first step is to convert RGB color map into $L\alpha\beta$ color space since the method uses CIE-Luv color space which has Gaussian smoothed blue-yellow, red-green and luminance planes. The next step is to determine and label the clusters by neighboring pixels within a spatial radius h_s, and color radius h_r. As the parameters are set by users, we set h_s as 6, h_r as 5 and M as 50 after some trials for our system.

Note that this step is also applied to the background layer of LDV formats in order to identify the foreground objects. Thus, the color map and background layer are segmented into regions by using mean-shift algorithm for LDV.

Importance Map Extraction

The importance map extraction approach works differently for V+D and LDV. Saliency-depth based method is applied to V+D, while a layer based importance map extraction is used for LDV.

Saliency-Depth Based Importance Map

For V+D videos, importance map extraction is based on the saliency and depth information. After the segmentation of the color map, three steps are achieved for generating the importance map: computation of saliency based on color map, computation of saliency based on depth map and computation of overall saliency map.

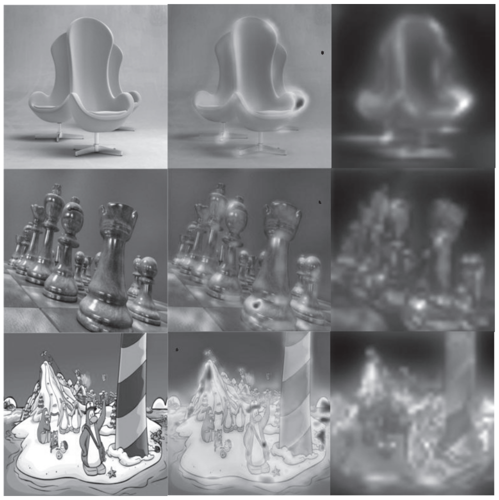
Most of the physiological experiments verify that human vision system is only aware of some parts of the incoming information in full detail. In order to locate the points of interest, the saliency concept is proposed. The graph-based visual saliency image attention model is used for saliency computation (Harel, 2007). It is a bottom-up visual saliency model that is constructed in two steps: Constructing activation maps on certain feature channels and normalization.

Graph-based visual saliency method contains three steps: *Feature extraction*, *activation* and *normalization of the activation map*. In the *feature extraction* step, the features such as color, orientation, texture, intensity are extracted from the color map through linear filtering and the calculation of center-surround differences for each feature type is completed. In the *activation step,* single or multiple activation maps are extracted by using feature vectors and subtracting feature maps at different scales such as henceforth, center, surround. Finally, the *normalization of the activation map* is performed. The goal of this step is to concentrate mass on activation maps by normalizing the effect of feature maps and summing them into the final saliency value of the pixel based on the color map.

Figure 2 shows several numbers of results that are based on graph-based visual saliency image attention model. The detailed explanation of the algorithm can be found in the work of Harel (2007).

After calculating saliency for each pixel on the color map, the depth saliency is computed. It is observed that depth is another factor to decide whether an object is important or should be ignored. In other words, closer objects should be more essential than the ones that are distant. Thus, we add the depth saliency for each pixel on the color map by using the associated depth map. A simple equation that is adapted from the work of Longurst (2006) is used in order to calculate the depth importance. The equation uses a model of exponential decay to get a typical linear model of very close objects.

Figure 2. Graph-based visual saliency image attention model. (a) Original Image; (b) Salient parts of the image (red – most salient); (c) Resulting saliency maps.



The last step is to compute the overall saliency. For each region that is segmented by mean-shift algorithm, the calculation of the overall saliency of the region is processed by averaging the sum of the color-based and depth-based saliency of pixels belonging to the region.

Layer Based Importance Map

For LDV videos, we follow a layer based approach with the aim of importance map extraction since foreground objects on the background layer are assumed to be important. In other words, the closer objects should be more salient than the distant ones. Thus, the segmented regions on the background layer are extracted from the color map at the end of this step. This approach is simpler than the saliency-depth based approach which is applied for V+D, because we use the features of LDV as our basis.

Background Resynthesis

After extracting important objects from the original color map, the background resynthesis step takes place. In this case, resynthesis refers to filling gaps of the extracted area with information from the surrounding area. This step is based on Harrison (2002)'s inpainting method. The algorithm reconstructs the gaps with the same texture as the given input color map by successively adding pixels that are selected. The procedure has two stages: pixel analysis and filling. In the first stage, relationships between pixels on the color map are analyzed and the value of each pixel that can be obtained by neighboring pixels is established. In the second stage, until all blank locations on the color map are filled, pixels are added by using the results of the pixel analysis stage. The procedure is capable of reproducing large features from the color map, even though it only examines interactions between pixels that are close to neighbors. Then, the color map is resized to 192x192 (standard thumbnail size).

Pasting of Important Objects

The next step is to paste important objects onto the new background. The constraint-based algorithm is utilized with the aim of pasting each object according to their importance values from the most important to least (Setlur, 2005). The goal is to preserve the relative positions of the important regions in order to keep the resized images layout similar to the original color map. For this algorithm, there are four constraints: positions of the important objects must stay the same, aspect ratios of the important objects must be maintained, the important objects must not overlap in the retargeted background if they are not overlapping in the original color map, and the background color of the important objects must not change.

From the most important object to least, this step reduces the change in position and the size of the important objects and the algorithm seeks whether the four conditions are satisfied or not. The aspect ratio and the position of the important objects are calculated according to the original and the retargeted color map.

3D Thumbnail Generation

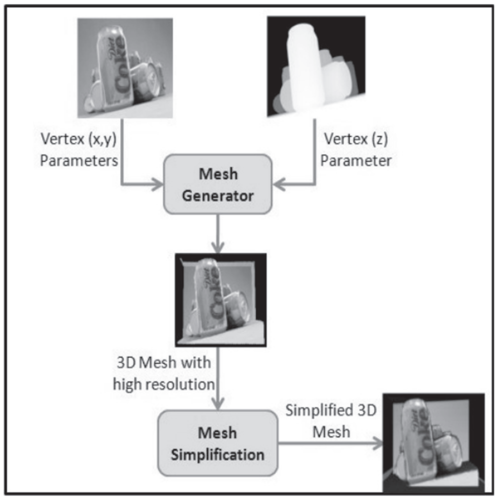
In the 3D thumbnail generation stage, we apply two different approaches to get the 3D visual effect on the retargeted color map: 3D mesh generation and parallax mapping techniques.

3D Mesh Generation Technique

With the purpose of the generation of a 3D mesh from the retargeted color map by using the associated depth map, we follow a simple algorithm as illustrated in Figure 3.

The inputs of the 3D mesh generation algorithm are the retargeted color map and the corresponding depth map. In order to create the geometry of the resulting 3D mesh, the vertices that describe points and corner locations of the mesh in 3D space should be extracted. Thus, positions of vertices on the x-y coordinate are obtained from the retargeted color map and the depth values of

Figure 3. The flow order of the 3D mesh generation technique



the corresponding vertices are acquired from the depth map. After obtaining all vertices, the construction of triangular faces between vertices to form the actual 3D mesh is achieved. Next step is to compute the texture data and face normals. Since the thumbnails should be in a simple format and a several numbers of thumbnails should be displayed in applications, it is necessary to consider the performance. Therefore, the constructed 3D mesh should be simplified because for a retargeted image which has 192x192 resolutions, there exist 36864 vertices and 73728 faces without simplification and this makes simultaneous rendering of multiple thumbnails impossible. For achieving the simplification, we use an edge collapse algorithm based on the quadric metric approach (Garland, 1998). This method produces high quality approximations of polygon models rapidly. In order to process simplification, iterative contractions of vertex pairs are used and the surface error approximations are maintained by using quadratic matrices. In addition to this, the algorithm joins unconnected regions by reducing arbitrary vertex pairs. Thus, after simplification it is guaranteed to have meshes that contains up to 4000 faces.

Parallax Mapping Technique

Parallax mapping is a shading technique and the enhancement of bump mapping or normal mapping which is proposed by Tomomichi (2001). It is applied to textures in 3D rendering applications such as 3D games, virtual environment applications etc. and also known as offset mapping or virtual displacement mapping.

Parallax mapping is a simple method to give motion parallax effects on a polygon. In other words, 2D textures have more apparent depth when this approach is applied. The combination of traditional normal and height mapping creates 3D effect without the use of additional vertices. By adding depth to 2D textures, the final render appears to have a much higher polygon count than it actually has. Finally, it is a per-pixel shape representation and can be accomplished using the current generation of 3D hardware.

USER EVALUATION

In order to evaluate the performance of the resultant 3D thumbnails from 3D videos with depth and the proposed 3D thumbnail generation methods (3D mesh vs. parallax-mapped based), we have performed two experimental studies.

Subjects

15 voluntary subjects participated: 13 males and 2 females with a mean age of 25.17. Two of the subjects were novice and others were experienced users with a computer science background.

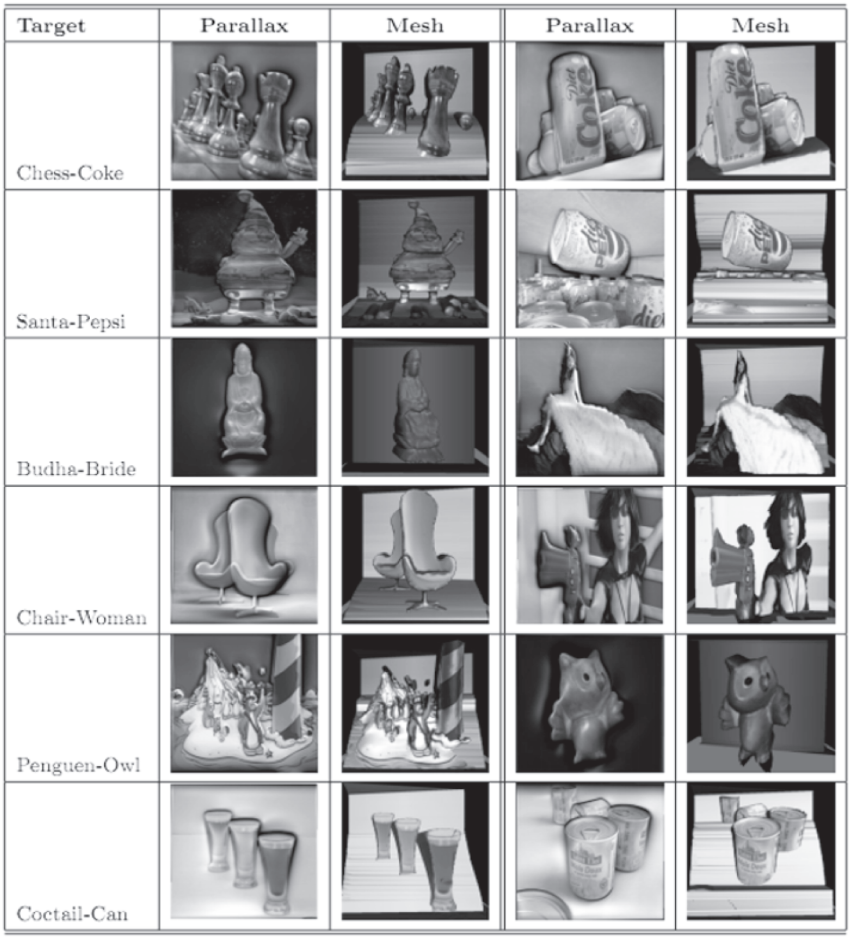
Equipment

All experiments were performed on the Sharp Actius RD3D with 15-inch XGA (1024-by-768) autostereoscopic color display. For interaction, mouse and keyboard were used. Subjects did not wear any special glasses.

Tasks

In the first experiment, the task that participants should accomplish was to select the correct thumbnail from a large set of 2D and 3D thumbnails for a given content name in a reasonable time. This user study had 60 steps. For the first 30 step, 3D thumbnails were randomly located on the 3D environment and displayed in a 3D grid layout. In addition to this, 2D thumbnails were randomly positioned on the 3D grid layout for the rest. For each step, a target thumbnail with a given content name was asked to be browsed by subjects. The second experiment was similar to the first, but in this case, subjects accomplished the selection of the target thumbnail from a set of 3D mesh or

Figure 4. The target 2D, 3D and parallax-mapped thumbnails used in experiments



parallax-mapped based thumbnails. Other than this, the structure of the experiment was similar with the first one.

For all tests, 12 different target content names as illustrated in Figure 4, were asked to be browsed and no text labels were satisfied for thumbnails.

Results and Discussion

For the first experiment, our hypothesis *was 3D thumbnails are illustrative than 2D thumbnails*. In order to prove this, we recorded the search time and the number of clicks performed to reach the target thumbnail. The comparison results are illustrated in Figure 5. By using 3D thumbnails, subjects accomplished 30 experiment steps in 142.138 seconds with 78.92 clicks, while they performed 87.304 clicks in 159.184 seconds with 2D thumbnails. From the figure and total results, it is clearly occurred that recognition time for 3D is shorter than 2D. With the aim of better indication of the statistically significant difference of 3D thumbnails, we have also performed a *paired samples t-test* on the experimental data. The mean error of each test case of 2D thumbnails was compared to the mean error of 3D thumbnails, and it showed that the difference between 3D thumbnails and 2D thumbnails is statistically significant with $p < 0.05$.

Figure 5. 2D thumbnails vs. 3D thumbnails (based on search time)

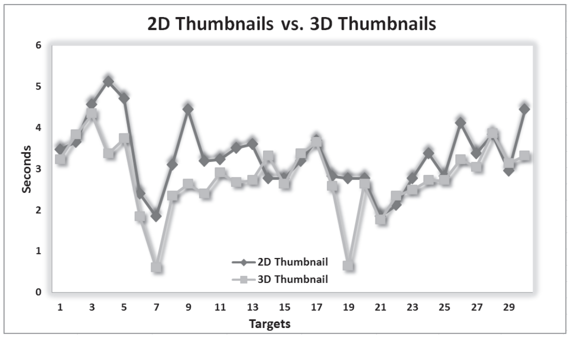
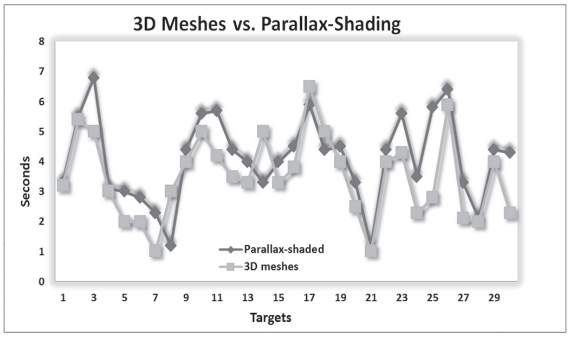


Figure 6. 3D thumbnails based on 3D mesh generation vs. parallax-mapped technique (based on search time)



The second experimental study was based on *3D meshes are illustrative than parallax-mapped images* hypothesis. By using 3D mesh-based thumbnails, subjects completed 30 experiment steps in 139.476 seconds while they performed the test with parallax-mapped thumbnails in 142.965 seconds. Moreover, 96.17 clicks were acquired to complete tasks for finding targets with 3D mesh-based thumbnails and 101 clicks were obtained with parallax-mapped thumbnails. Figure 6 and total result show that thumbnails that are based on 3D mesh generation technique are more recognizable than parallax-mapped thumbnails. However, from the *paired samples t-test* results, this difference is not significant and our hypothesis was rejected because the statistically significant difference between the two methods was not acceptable ($p > 0.05$).

FUTURE RESEARCH DIRECTIONS

For the future work, 3D thumbnails should be generated for CSV and MDV formats and a comparison between 4 methodologies should be accomplished. Thence, the suitable format that provides a fast 3D thumbnail creation approach can be determined. Moreover, as our video frame selection is based on a saliency-based approach, a stronger and efficient video summarization technique should be applied to our 3D thumbnail generation system in order to get the most meaningful frame that represents the entire video. As a result, our thumbnail generation methodology can be improved and more illustrative thumbnails for all kinds of 3D videos can be generated. Finally, additional user studies should be performed with the aim of the proof for the efficiency of 3D thumbnails.

CONCLUSION

A framework that generates 3D thumbnails for 3D videos with depth is proposed in this chapter. The goal of the framework is to create meaningful, illustrative and efficient thumbnails by preserving the important parts of the selected frame of the input video. The inputs of the system are two different video formats: V+D and LDV, and the generation approaches are different for each format. For V+D, the important objects are extracted by a saliency-depth based method. In other words, the visually salient and closer objects are important and necessary to be preserved. However, the foreground objects that are on the background layer are assumed to be important for LDV formats. This method is called layer-based. After determining the important objects, remaining steps are same for all formats. Important objects are extracted from the color map and resulting gaps are filled. After that, newly created background image is resized to a standard thumbnail size and important objects are pasted on it by a special algorithm that has 4 constraints: aspects ratios of the important objects are maintained, the background color and the positions of the important objects should be same as the original color map and objects should not be overlapping if they are not overlapping in the original image. In the final stage, two techniques are used to generate the resultant 3D thumbnail: 3D mesh and parallax mapping methods.

Finally, we have performed two user experiments in order to test the efficiency of 3D thumbnails. In the first experiment, we compared 2D thumbnails and 3D thumbnails. The experiment results show that 3D thumbnails are statistically *($p < 0.05$)* illustrative than 2D thumbnails. Moreover, the second experiment's aim is to test the efficiencies of the two proposed methods for generating 3D thumbnails: 3D mesh and parallax mapping techniques. From the experiment results, it is indicated that there is no significantly difference between two techniques since $p > 0.05$. Thus, either the 3D thumbnail which is generated by 3D mesh or parallax-shading technique can give the 3D visual effect and enhance the depth perception.

ACKNOWLEGMENT

We gratefully acknowledge support from the "All 3D Imaging Phone" project under contract FP7-213349.

REFERENCES

Adobe. (2010). *Website.* Retrieved August 12, 2010, from http://www.adobe.com

Bregler, C. (1998). Tracking people with twists and exponential maps. *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, (pp. 8-15).

Comaniciu, D., & Meer, P. (2002). Mean shift: A robust approach toward feature space analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *24*(5), 603–619. doi:10.1109/34.1000236

Farin, D., Effelsberg, W., & De, H. N. P. (2002). Robust clustering-based video-summarization with integration of domain-knowledge. In *International Conference on Multimedia and Expo (ICME): Vol. 1,* (pp. 89-92). Lausanne, Switzerland.

Felzenszwalb, P., & Huttenlocher, D. (2004). Efficient graph-based image segmentation. *International Journal of Computer Vision*, *6*(2), 167–181. doi:10.1023/B:VISI.0000022288.19776.77

Garland, M., & Heckbert, P. (1998). Simplifying surfaces with color and texture using quadric error metrics. In *IEEE Conference on Visualization*, (pp. 263-269).

Gimp. (2010). *Website*. Retrieved August 12, 2010, from http://www.gimp.com

Gong, Y., & Liu, X. (2000). Generating optimal video summaries. In *International Conference on Multimedia and Expo (ICME): Vol. 3* (pp. 1559-1562).

Gundogdu, R. B., Yigit, Y., & Capin, T. (2010). 3D thumbnails for mobile media browser interface with autostereoscopic displays. *Springer Lecture Notes in Computer Science, Special Issue on IEEE Multimedia Modeling 2010*. Chongqing, China.

Harel, J., Koch, C., & Perona, P. (2007). Graph-based visual saliency. *Advances in Neural Information Processing Systems*, *19*, 545–552.

Harrison, P. (2001). A non-hierarchical procedure for re-synthesis of complex textures. In *Proc. WSCG*, 190-97.

F Institute. (2008). *D5.1 – Requirements and specifications for 3D video*. All 3D Imaging Phone.

Itti, L., Koch, C., & 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. doi:10.1109/34.730558

Jojic, N., Frey, B., & Kannan, A. (2003). Epitomic analysis of appearance and shape. *IEEE International Conference on Computer Vision*, (pp. 34-41).

Kaneko, T., Takahei, T., Inami, M., Kawakami, Y. Y. N., Maeda, T., & Tachi, S. (2001). *Detailed shape representation with parallax mapping* (pp. 205–208). ICAT.

Lienhart, R., Pfeiffer, S., & Effelsberg, W. (1997). Video abstracting. *Communications of the ACM*, *40*(12), 55–62. doi:10.1145/265563.265572

Longhurst, P. (2006). A GPU based saliency map for high-fidelity selective rendering. In *International Conference on Computer Graphics, Virtual Reality, Visualization and Interaction*, (pp. 21-29). Africa.

Müller, K., Smolic, A., Dix, K., Kauff, P., & Wiegand, T. (2008). Reliability-based generation and view synthesis in layered depth video. In *Proc. IEEE International Workshop on Multimedia Signal Processing (MMSP2008)*, (pp. 34-39). Cairns, Australia.

Mundur, P., Rao, Y., & Yesha, Y. (2006). Keyframe-based video summarization using Delaunay clustering. *International Journal on Digital Libraries*, *6*, 219–232. doi:10.1007/s00799-005-0129-9

Pantofaru, C. (2005). *A comparison of image segmentation algorithms*. Robotics Inst., Carnegie Mellon University.

Setlur, V., Takagi, S., Raskar, R., Gleicher, M., & Gooch, B. (2005) Automatic Image retargeting. In *International Conference on Mobile and Ubiquitous Multimedia*, (pp. 59-68). New Zealand.

Suh, B., Ling, L., Bederson, B., & Jacobs, D. (2003). Automatic thumbnail cropping and its effectiveness. *ACM Symposium on User interface Software and Technology*, (pp. 95-104). Vancouver, Canada.

Yow, K. (1998). *Automatic human face detection and localization.* Unpublished doctoral dissertation, University of Cambridge, Cambridge.

ADDITIONAL READING

Comaniciu, D., & Meer, P. (2002). Mean shift: A robust approach toward feature space analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *24*(5), 603–619. doi:10.1109/34.1000236

Gundogdu, R. B., Yigit, Y., & Capin, T. (2010). 3D thumbnails for mobile media browser interface with autostereoscopic displays. *Springer Lecture Notes in Computer Science, Special Issue on IEEE Multimedia Modeling 2010*. Chongqing, China.

Harel, J., Koch, C., & Perona, P. (2007). Graph-based visual saliency. *Advances in Neural Information Processing Systems*, *19*, 545–552.

Harel, J., Koch, C., & Perona, P. (2007). Graph-based visual saliency. *Advances in Neural Information Processing Systems*, *19*, 545–552.

Itti, L., Koch, C., & 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. doi:10.1109/34.730558

Setlur, V. (2005). *Optimizing computer Imagery for more effective visual communication.* Unpublished doctoral dissertation, Northwestern University, Illinois.

Talton, J. O. (2004). *A short survey of mesh simplification algorithms. University of Illinois*. Urban-Campaign.

KEY TERMS AND DEFINITIONS

3D Video: Kind of a visual media that satisfies 3D depth perception that can be provided by a 3D display.

Auto-Stereoscopic Display: A 3D display that helps user to see the content without help of any 3D glasses on the flat screen.

Depth Map: A grey-scale map that includes depth information of the content for every pixel. In this map, the object that is nearest is the bright one.

Grid Layout: A type of a layout that divides the container into equal-sized rectangles and each rectangle holds one item.

Mean-Shift Segmentation: A powerful image segmentation technique that is based on non-parametric iterative algorithm and can be used for clustering, finding modes etc.

Parallax Shading: A shading technique that displaces the individual pixel height of a surface, so that when the resulting image is seen as three-dimensional.

Saliency: Refers to visual saliency which is the distinct subjective perceptual quality that grabs attention.

Thumbnail: A small image that represents the content and used to help in recognizing and organizing several numbers of contents.