INTERACTIVE 2D-3D IMAGE CONVERSION FOR MOBILE DEVICES

Yagiz Aksoy, Ozan Sener, Aydin Alatan

Electrical and Electronics Engineering, METU
06800, Ankara, Turkey
{yaksoy, ozansener, alatan@eee.metu.edu.tr}

Kemal Ugur

Nokia Research Center
Tampere, Finland
{kemal.ugur@nokia.com}

ABSTRACT

We propose a complete still image based 2D-3D mobile conversion system for touch screen use. The system consists of interactive segmentation followed by 3D rendering. The interactive segmentation is conducted dynamically by color Gaussian mixture model updates and dynamic-iterative graph-cut. A coloring gesture is used to guide the way and entertain the user during the process. Output of the image segmentation is then fed to the 3D rendering stage of the system. For rendering stage, two novel improvements are proposed to handle holes resulting from depth image based rendering process. These improvements are also expected to enhance the 3D perception. These two methods are subjectively tested and their results are presented.

***Index Terms*—** 2D-3D Conversion, Interactive Segmentation, Depth Image Based Rendering, View Synthesis, Mobile, Subjective Tests

1. INTRODUCTION

Although mobile 3D displays are ready for market, there is not enough 3D content for these displays. Therefore, conversion of 2D content to 3D is crucial in the intermediate period. For conversion of mono images to stereo, an interactive segmentation based method is proposed in [1]. The results of that method show that with accurate image segmentation, it is possible to generate 3D content with acceptable quality. However, fully automatic and accurate image segmentation is still a difficult problem, despite the advancements during recent years. When compared to fully-automatic methods, semi-automatic interactive segmentation algorithms have improved performance due to the user involvement. Recent interactive segmentation algorithms were shown to be successful in various applications [2, 3] but these methods are not very suitable for mobile touch-screen applications. Firstly, correction stage commonly used in these algorithms is not desired due to the tedious zooming operations. Moreover, the complexity of some of the algorithms make them not very suitable for mobile use-cases due to limited processing power of mobile devices. After extensive investigation of existing literature on interactive segmentation, it is argued that there is no candidate which satisfy all of these requirements. Therefore, an interactive segmentation algorithm based on dynamic and iterative graph-cut specifically tailored for mobile touch screen applications is proposed in [4] and utilized in our paper for our 2D-3D conversion.

After the image segmentation stage, depth image based rendering (DIBR) techniques are used in literature to obtain stereoscopic data from a single image and corresponding depth map [5]. Several algorithms, including using highly sophisticated image inpainting ones [6] are proposed for visually plausible hole filling. In this paper, we propose two novel hole handling techniques, namely foreground enlargement, covering holes by scaling the foreground to a slightly bigger scale, and background blurring, hiding the artifacts resulting from hole filling by imitating the blurry background effect in photographs having short depth of field.

This paper is organized as follows. In Section 2, we provide the details of our proposed algorithm. Firstly the details of the interactive segmentation algorithm is proposed, followed by the proposed hole filling techniques. In Section 3, we describe the conducted subjective experimental results of our algorithm. Section 4 concludes the paper.

2. PROPOSED METHOD

2.1. Interactive Segmentation

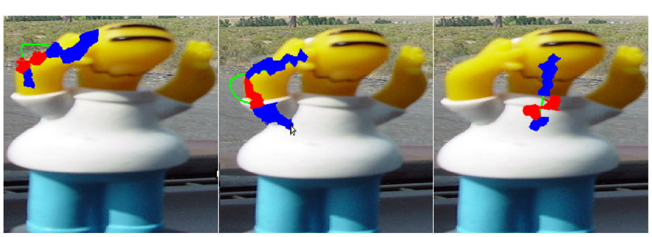
In order to decrease the computation time, input image is initially oversegmented via simple linear iterative clustering (SLIC) algorithm [7] in order to decrease the number of nodes in the graph. Then, grayscale version of the image is shown to user and dynamic-interactive image segmentation process takes place. User starts to colorize the object of interest by finger strokes on the screen. With each input stroke, a dynamic and iterative algorithm is executed, and the intermediate results of the algorithm is updated on the display in real-time. User continues to colorize the object of interest until he/she is satisfied with the result.

Color model used in the algorithm is mixture of Gaussians in RGB color space [3]. With each stroke of the user, foreground color model is learned from interacted super-pixels, while interaction is background free. Background is assumed to be the non-interacted super-pixels and background model is learned from these regions. As a result, there exist actual foreground regions in the background model and this makes the algorithm more conservative against expanding the foreground. Energy minimization problem is defined by using Gibbs energy and this energy is minimized by using min-cut/max-flow algorithm [8]. In order to apply min-cut/max-flow efficiently, the algorithm runs on the sub-image around the interaction instead of the whole image [4]. Size of this sub-image is determined using the method proposed in [4]. Some interactions and their corresponding results are shown in Figure 1.

Interaction | Segmentation Result

Fig. 1: User interactions and corresponding segmentation results

In order to conceal user interaction errors, a novel method is also proposed in [4]. While coloring the object of interest, user generally goes beyond the borders of the interested object. Moreover, most of the times, user returns back. In order to solve these type of interaction errors, a cost function based on color similarities is defined and minimized by using A* algorithm.



(a) Error - 1 color | (b) Error - 2 colors | (c) No Error

Fig. 2: Path correction algorithm for 3 main scenarios

In Figure 2, three main scenarios for the path correction is summarized. Blue superpixels are the accepted input, whereas green lines are the discarded input. Red superpixels are obtained by finding the minimum-cost path. So, only blue and red regions are used by the segmentation algorithm.

In Figure 2a, user first leaves the object erroneously and then returns back to foreground object pixels. In Figure 2b, user leaves the object accidentally from the yellow region, then comes back to a white colored region. Proposed algorithm also handles this case. In Figure 2c, user leaves the yellow region (it is not an error) and then continues along the white region. The algorithm first assumes an interaction error, but when minimum cost path is obtained, initial path is found to be also correct.

2.2. 3D Rendering

Following the interactive segmentation stage, user is asked to select from a different set of rendering options to create a stereoscopic image pair. A depth map is synthesized by taking foreground map from the interactive segmentation step and assigning corresponding pixels with minimum depth, while assigning background pixels with maximum depth. This depth map is smoothed for a better visual quality [9]. After processing several options detailed below, original image is kept as right view and left view is computed using conventional DIBR techniques [5].

There are two rendering options presented in our system: Foreground Enlargement and Background Blurring, while convergence distance is also controllable via the user interface.

2.2.1. Foreground Enlargement

Enlarging foreground is a novel contribution in order to handle holes in the depth image based rendered image and also to obtain a better depth feeling. The object selected in the interactive segmentation step is scaled to a larger size (in our system, scaling factor is 1.1) and depth map formation and DIBR processes are held accordingly. In order the enlarged foreground to look natural on the background, transition of intensity across the pixels along the edge is smoothed, imitating alpha mattes [10].

Enlarging foreground, especially if the foreground object is concave, results in occluded areas being also covered in the rendered left image and hence, without missing information from the background, left image can be synthesized (Figure 3). If any occlusions remained, we handled the occlusions by replacing the missing pixels by the closest available background pixel. When a 3D image is created from a single 2D still image, we also expect foreground enlargement technique to help give a better 3D depth feeling. This expectation is tested in subjective tests and results are presented in Section 3.2.

2.2.2. Background Blurring

Another novel approach we present to handle possible errors resulting from holes in the rendered image and to yield a bet-

Fig. 3: Top: Original and synthesized left view; holes are red. Bottom: Foreground enlarged right and left images.

ter 3D perception is background blurring.

In photography, high aperture size is sometimes preferred to emphasize foreground object. High aperture limits the effective focus area, in which the objects appear sharp. This way, object in foreground is highlighted by blurring other content. By blurring background, we are imitating this effect to emphasize foreground objects and hence to give a better 3D perception. Background blurring is also useful in handling holes in the rendered images as any possible visually implausible artifact resulting from attempt to fill the holes is also blurred out. An example can be seen in Figure 4.

Fig. 4: Left: Left view without background blurring. Right: Left view with background blurring.

3. SUBJECTIVE TESTS

3.1. Interactive Segmentation

We have proposed and practiced a subjective evaluation for interactive segmentation algorithms. We have compared 3 different interaction methods, namely intelligent scissors [2], grab-cut[3] and the proposed method [4]. Grab-cut uses only a rectangle around the object of interest as an input, whereas intelligent scissors requires the approximate boundary of the object. Our method requires coloring of the object using the approach detailed in Section 2.1.

3.1.1. Subjective Test Methodology

In order to evaluate the algorithms, four different evaluation metric is used: *performance*, *entertainment*, *easiness* and *overall*. A demo of three interaction methods is shown to the subject. After the demo, the subject is asked to segment four random images with all three algorithms in a random order. When finished, the subject is asked to rate each algorithm by the four aforementioned metrics. Rating is done by grading each test from 1 to 5. Grade 1 corresponds to the worst score within the metric, whereas grade 5 corresponds to the highest score. The tests were conducted in a regular office environment with capacitive touch screen tablet PC. 15 subjects, composed of undergraduate engineering students, participated in the tests.

3.1.2. Test Results

Performance: The proposed method has the best performance result, while intelligent scissors is the second best. This result is expected due to the clear performance difference between algorithms. Extensive analysis on the performance of each algorithm is presented in [4].
Easiness: Grabcut only requires a rectangle around the object. As a result, it scored to be the easiest one in the subjective tests. Our method becomes second since coloring is considered to be much easier than choosing a boundary.
Entertaining: Our method has the best entertainment result subjectively. This is possible due to the coloring gesture. Intelligent scissors was ranked third since selection of landmarks and moving around the boundary is a quite hard task.
Overall: All the subjects has selected the proposed algorithm as the one having the best overall experience.

Median ratings for each algorithm as well as inter-quartile ranges (IQR) and standard deviations (STD) are summarized below. Dependent ANOVA test is applied and resulting p-values are the same for each metric and equal to 0.0005.

Table 1: Interactive Segmentation Test Results (Median:IQR:STD)

	Perf.	Easiness	Entertain	Overall
Proposed Method	5:1:.45	4:0:.86	5:1:.74	4:1:.45
Grab-Cut[3]	3:2:.92	4:1:.75	2:1:.61	3:1:.77
Int. Scissors[2]	3:1:.51	2:1:.74	3:2:.89	2:1:.76

3.2. 3D Rendering

In 3D Rendering tests, user preferences for foreground enlargement and background blurring methods were evaluated.

3.2.1. Subjective Test Methodology

The double-stimulus continuous quality-scale (DSCQS) method [11] was adopted. 3D images from 12 different test images were created by the experimenter beforehand using the system presented. For each of the tests, two stereoscopic images are shown to the subject side by side and the subject is asked to choose the better one. One of the stereoscopic images is plain and the other one is generated using one or two of the algorithms presented in Section 2.2. The subjects are not told which one of the images are generated by the proposed methods, also the placement of the images are random.

The tests were conducted in a regular office environment. Each subject was shown 30 different image pairs. All pairs were shown twice at random times for a consistency check. 36 subjects, composed of undergraduate engineering students, participated in the tests. Images were shown on a 22” 3D computer screen with shutter glasses.

3.2.2. *Results*

Some of the images scored particularly well when foreground was enlarged, while others performed worse. It is observed that if the foreground object is actually closer to the camera in real 3D space (i.e. when the synthesized depth map is more realistic), foreground enlargement is preferred by subjects. On the other hand, if there exist some objects actually closer to the camera than the selected foreground object, foreground enlargement is not preferred due to complications in 3D perception. Moreover, if a visual inconsistency due to replacement of the foreground object occurs, the technique is unfavored. Several images belonging to these two subsets can be seen in Figure 5. Overall test results for these two subsets can be seen in Table 2.

Fig. 5: Images which scored high (top) and low (bottom) in foreground enlargement tests

Table 2: Foreground Enlargement Test Results

Test Image Subset	Positive	Negative	Neutral
FG Enlargement Compatible	%60	%35	%5
Not-FG Enl. Compatible	%41	%47	%12

Table 3: Background Blurring Test Results

Used Algorithm	Positive	Negative	Neutral
Background Blurring	%19	%74	%7
BG Blur & FG Enlargement	%22	%73	%5

Background blurring scored quite low in the tests. It can be argued that, the apparent degradation of image quality disturbs subjects. This also affected user preferences when foreground enlargement and background blurring were used together. Nevertheless, including an option to blur the background should be preferable in such an 2D-3D conversion system. Results are presented in Table 3.

4. CONCLUSION

This paper presents a novel 2D-3D image conversion system, optimized for mobile devices. The proposed system involves a novel interactive image segmentation algorithm and a 3D rendering stage. Proposed interactive segmentation algorithm is found out to be the most preferable through subjective tests. Among proposed rendering techniques, foreground enlargement appears to be preferable for some specific data, whereas background blurring should remain as a user preference.

5. REFERENCES

[1] H.E. Tasli and K. Ugur, “Interactive 2d 3d image conversion method for mobile devices,” in *3DTV-CON*, May 2011, pp. 1 –4.

[2] E.N. Mortensen and W.A. Barrett, “Intelligent scissors for image composition,” *Proceedings of the 22nd Annual Conference on Computer Graphics and Interactive Techniques - SIGGRAPH ’95*, vol. 84602, no. 801, pp. 191–198, 1995.

[3] C. Rother, V. Kolmogorov, and A. Blake, “Grabcut: Interactive foreground extraction using iterated graph cuts,” *ACM Transactions on Graphics*, vol. 23, pp. 309–314, 2004.

[4] O. Sener, K. Ugur, and A. Alatan, “Robust interactive segmentation via coloring,” in *In Proceedings of the Workshop on Vision Interfaces for Ground Truth Applications (in conjunction with AVI2012)*, May 2012.

[5] C. Fehn, “Depth-image-based rendering (dibr), compression, and transmission for a new approach on 3d-tv,” 2004, vol. 5291, pp. 93–104, SPIE.

[6] P. Ndjiki-Nya, M. Koppel, D. Doshkov, H. Lakshman, P. Merkle, K. Muller, and T. Wiegand, “Depth image-based rendering with advanced texture synthesis for 3-d video,” *Multimedia, IEEE Transactions on*, vol. 13, no. 3, pp. 453 –465, June 2011.

[7] R. Achanta, A. Shaji, K. Smith, A. Lucchi, P. Fua, and S. Susstrunk, “SLIC Superpixels,” Tech. Rep., EPFL, 2010.

[8] Y. Boykov and V. Kolmogorov, “An experimental comparison of min-cut/max-flow algorithms for energy minimization in vision.,” *IEEE PAMI*, vol. 26, no. 9, pp. 1124–37, Sept. 2004.

[9] W.J. Tam, G. Alain, L. Zhang, T. Martin, and R. Renaud, “Smoothing depth maps for improved steroscopic image quality,” 2004, vol. 5599, pp. 162–172, SPIE.

[10] D. Singaraju and R. Vidal, “Estimation of alpha mattes for multiple image layers,” *IEEE PAMI*, vol. 33, no. 7, pp. 1295 –1309, July 2011.

[11] “Itu-r recommendation bt.500-7 (1974-1997). methodology for the subjective assessment of the quality of television pictures,” .