10-2012

3DME: 3D Media Express from RGB-D Images

Tam Nguyen
University of Dayton, tnguyen1@udayton.edu

Lusong Li
National University of Singapore

Jun Tan
National University of Defense Technology

Shuicheng Yan
National University of Singapore

3DME: 3D Media Express from RGB-D Images

Tam V. Nguyen, Lusong Li, Jun Tan, Shuicheng Yan
Department of Electrical and Computer Engineering, National University of Singapore, Singapore
{tamnguyen,eleyans}@nus.edu.sg;{lilusong, tanjun.nudt}@gmail.com

ABSTRACT

Considering the continuously increasing availability and accessibility of 3D media and the depth camera such as Kinect, we demonstrate an innovative 3D media system called *3DME*. The objective of this demo is three-fold. First, the demo exhibits the creation of 3D images from RGB-D images. Second, *3DME* allows a user to insert impressive effects to the produced 3D content. Last but not least, our demo is one of the first attempts towards advertising for 3D content which enables both the advertisers and content providers deliver more effective ads carried through 3D media.

Categories and Subject Descriptors

H.4 [**Information System Applications**]: Miscellaneous; H.5.2 [**User Interfaces**]: User-centered design

General Terms

Experimentation, Human Factors.

Keywords

3D media, 3D advertising, depth information.

1. INTRODUCTION

The beginning of this decade witnesses the 3D entertainment form thrilling the world succeeding the release of the 3D masterpiece *Avatar*. More and more 3D HDTV channels are in service, and 3D TV sets are getting popular in daily life. There are 3D videos increasingly published on the popular video sharing website YouTube. About 32,000+ videos came up when the keywords "3D videos glasses needed" are entered on YouTube. The emergence of 3D camcorder and 3D glasses display technology promises a large number of 3D contents in the near future. Recently, depth cameras have been rapidly gaining popularity with devices such as Kinect [1] becoming popular in the consumer space as well as the research space. Thus, we explore the mutual benefits between 3DTV and depth camera in order to generate and display 3D content. In addition, we think the inserted effects similar to Instagram [2] may attract the users' attention. Considering there will be more and more digital 3D media shared on the web, a new way for this advertising market is desired now. We believe that if people viewed a

MM '12, October 29 - November 2, 2012, Nara, Japan
ACM X-XXXXX-XX-X/XX/XX.

Figure 1: User interface of our 3DME system. The user selects the 3D display method, anaglyph or stereoscopic and effect options. "◯" button is for generating 3D media content.

commercial in three dimensions, they would pay more attention to the commercial, remember the product more sufficiently, have a better knowledge and understanding of the advertisement itself and be more willing to interact with the attractive advertisements. Li et al. conducted a user study and found that when watching 3D graphical advertisements, participants experienced stronger feelings of presence than watching 2D ones [7]. Although researches have provided some advertising methods for images such as ImageSense [9] and PageSense [8], yet there is no existing works about 3D advertising. To this end, we propose a friendly user interface shown in Figure 1. This demo is able to automatically generate and embed more effects into 3D images.

2. TECHNICAL DETAILS

In this section, we introduce the method to generate 3D images from the input RGB-D images captured by Kinect. Then we present how effects and additional 3D contents are embedded into the generated 3D images.

2.1 3D image generation

The difficulties of 3D image generation consist of the unalignment between color and depth images and the noisy depth map. Thus, we first perform calibration on both depth and color cameras to find the transformation between their images in a similar way as [3]. The second step is to smoothen the depth map. The color image is oversegmented into superpixels [6]. Each depth pixel, whose original depth value equal to 0, is assigned the average depth of the nearest neighbors in 8 directions in the same color superpixel. Finally, we apply a conventional Laplacian filter with a 3×3 kernel for pixels whose depth values equal to 0 until all missing depth pixels are filled. The stereoscopic image pair is produced by extracting parallax values from the smoothed depth map D and applying them to the left image I^l and right image I^r. For each pixel of the input color image I, the value of the parallax is obtained from its depth value. Considering the input image as a virtual central view, the left/right views are then obtained by shifting left/right the input image pixels by a value $\rho = parallax/2$, where $parallax$ is calculated as $parallax = \zeta \times (1 - \frac{V}{W})$, where ζ is the interaxial gap between two eyes, averaged as $60mm$, V is the screen depth or the distance from eyes to the screen and fixed as $150cm$ in our implementation, W is the vertex depth, equal to the summation of screen depth V and perceived depth P. For each pixel x in the image I, the perceived depth $P(x)$ can be calculated as $P(x) = D(x) \times \tau$, where τ ($\tau = 39.2$) is the ratio between the maximum depth distance captured by Kinect (10,000mm) and the maximum value in the depth image D (255). Figure 2 illustrates the framework of the 3D media generation. The stereoscopic images are then shown on a 3D monitor by using 3D Vision toolkit[4].

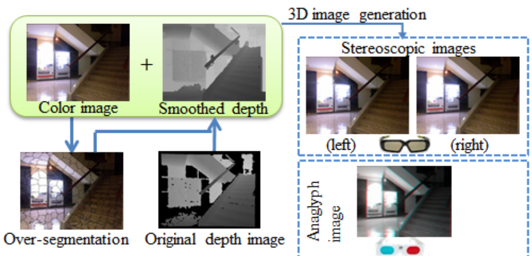

Figure 2: 3D image generation process.

2.2 Adding effects to 3D image

Similar to Instagram [2], we allow users to apply the digital filters to the taken 3D image. For the effects such as pencil sketching, grayscale, the filters are applied for the separate left and right images. For the effects of oil painting, the coordination between views is not maintained. Therefore, the effects are applied on the original color image. Then, the left and right images are generated following the method as aforementioned in the previous subsection.

2.3 Embedding 3D contents

Another distinctive feature of the demo, *3DSense*, is the insertion of 3D content such as advertisement logos, icons. The location to add the advertised information is based on the saliency map where each pixel indicates its informative value with respect to the human visual system. We applied saliency detection on the original color image [5]. Then, we consider depth information as another cue to detect salient regions. The closer depth regions are treated as the foreground, whereas the farther depth regions are considered as background. In the depth saliency map, the saliency value of one pixel is computed as $max\_depth - depth\_value$, which means the closer depth indicates the higher saliency. We also consider face area, obtained by a face detector [10], as the salient region. The final saliency map is computed as the average of all saliency maps. The final saliency map is divided into 6×8-cell grid. The ad information will be inserted

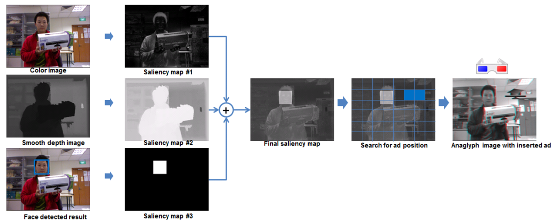

Figure 3: The procedure to embed additional content into 3D image.

into the least salient region as shown in Figure 3. Unlike [9], the additional information is not restricted to squared size, *i.e.* 1×1-cell. The resulted 3D image with embedded ad information is generated from depth and color images. Following our observation, the stereoscopic images create better 3D effect and do not hurt human eyes like red-cyan images. Figure 4 shows some results generating by 3DME .

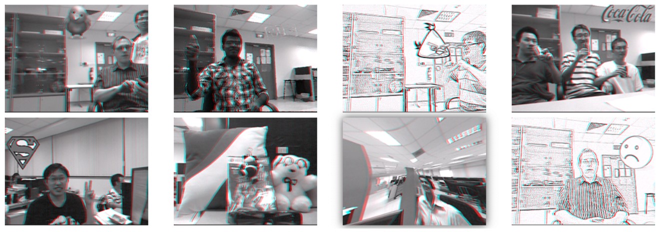
Figure 4: The exemplar inserted effects and contents to 3D anaglyph images.

3. DEMONSTRATION

We demonstrate the user-friendly 3DME in creating 3D media from a RGB-D camera. We also provide all the equipments consisting of a laptop with a special 3D screen coupled with active shutter glasses, and Kinect. Within this demo, the interested users are encouraged to generate 3D contents as well as try various additional effects. We hope our work would inspire more interesting research works in the near future.

4. ACKNOWLEDGEMENT

This work was supported by NExT Research Center funded under the research grant WBS. R-252-300-001-490 by MDA, Singapore.

5. REFERENCES

[1] http://www.xbox.com/kinect.
[2] http://http://instagr.am/.
[3] http://nicolas.burrus.name/index.php/Research/Kinect.
[4] http://www.nvidia.com/object/3d-vision-main.html.
[5] R. Achanta, S. S. Hemami, F. J. Estrada, and S. Süsstrunk. Frequency-tuned salient region detection. In *CVPR*, 2009.
[6] R. Achanta, A. Shaji, K. Smith, A. Lucchi, P. Fua, and S. Susstrunk. SLIC Superpixels. *EPFL Technical Report*, 2010.
[7] H. Li, T. Daugherty, and F. Biocca. Impact of 3-d advertising on product knowledge, brand attitude, and purchase intention: the mediating role of presence. *Journal of Advertising*, 2002.
[8] L. Li, T. Mei, X. Niu, and C.-W. Ngo. Pagesense: style-wise web page advertising. In *WWW*, 2010.
[9] T. Mei, X.-S. Hua, and S. Li. Contextual in-image advertising. In *Proceedings of ACM Multimedia*, pages 439–448, 2008.
[10] P. Viola and M. Jones. Robust real-time face detection. *IJCV*, 2004.