2D-To-3D Conversion of Images using Edge Information

S. Bharathi
PG Scholar,
Kumaraguru College of Technology,
Coimbatore 641049, Tamil Nadu, India.

A. Vasuki
Professor,
Kumaraguru College of Technology,
Coimbatore 641049,Tamil Nadu, India.

ABSTRACT

The three-dimensional (3D) displays required the depth information which is unavailable in the conventional 2D content. This work presents a novel algorithm that automatically converts 2D images into 3D ones. The proposed algorithm first segments the image into object groups by choosing an effective grouping method. This is done with the help of the edge information. The grouping is based on the pixels having similar colors and spatial locality. Based on a hypothesized depth gradient model, a depth map is assigned. Next, the depth map is assigned by cooperating with a cross bilateral filter to diminish the blocky artifacts. Then the filtered image is processed using a depth image based rendering(DIBR) method.

General Terms

2D images, 3D displays, depth map, effective grouping, visually comfortable.

Keywords

Bilateral filtering, depth map generation, 2D-to-3D conversion, edge information.

1. INTRODUCTION

With the development of 3D applications, the conversion of existing 2D images to 3D images becomes an important component of 3D content production. The conversion process of existing 2D images to 3D is commercially viable and is fulfilling the growth of high quality stereoscopic images. The dominant technique for such content conversion is to develop a depth map for each frame of 2D material. When observing the world, the human brain usually integrates the heuristic depth cues for the generation of the depth perception. The major depth perceptions to be noted are binocular depth cues from two eyes and monocular depth cues from a single eye [4]. The disparity of binocular visual system helps human eyes to converge and also to accommodate the object at the right distance. Monocular cues which include focus/defocus, motion parallax, relative height/size, and texture gradient provide various depth perceptions based on human experience. Therefore, humans are able to perceive depth from the single-view image/video.

The key step in 2D to 3D conversion process is the generation of a dense depth map. In recent years, a number of depth map generation algorithms have been proposed according to the principle of human visual system. Every algorithm has its own strengths and weakness. Most depth extraction algorithms make use of a certain depth cue but few of them combines two or more depth cues to generate depth map. 2D-to-3D depth generation algorithms generally face two challenges. One of them is the depth uniformity inside the same object. A better grouping of pixels results in a better outcome for the depth uniformity inside the object. The other challenge involves retrieving an appropriate depth relationship among all objects [3], [8]. These methods result in the false depth information when the object is with different self motion vectors. The pixels can belong to the same object that may be assigned with different depth values. Thus the depth map generation from single 2D images is an ill-posed problem.

In order to overcome these two challenges, this work presents an algorithm that uses a simple depth hypothesis to assign the depth of each group instead of retrieving the depth value directly from the depth cue. At first, the proposed algorithm chooses an effective grouping method in which the grouping of pixels has similar colors and spatial locality. Then the depth values are assigned as per the hypothesis depth value. Cross bilateral filter is then used to enhance the visual comfort. Experimental results imply that the proposed algorithm generates promising results with slight side effects.

The paper is organized as follows. Section 2 describes the generation of depth map by means of an automatic 2D-to-3D conversion system with the help of edge information. Section 3 gives an analysis over the experimental results observed from four different images with the comparison over obtained depth map and the output images. This section is useful for the knowledge of computational complexity over the system and the visual quality obtained. Section 4 concludes with the overall remarks that were noted from the previous sections and summarizes the quality of the various output images and also the conversion system.

2. PROPOSED CONVERSION SYSTEM

The work here describes an efficient 2D-to-3D conversion method based on the use of edge information. Importantly, the edge of an image has a high probability as it can be the edge of the depth map. Once the pixels are grouped together, a relative depth value can be assigned to each region. Figure 1 schematically depicts the proposed conversion system. Initially, the block-based image is considered to segment it into multiple groups. Then the depth of each segment is assigned with the help of an initial depth hypothesis. Next, the blocky artifacts have to be removed using cross bilateral filtering. Finally, multi-view images are obtained by the method of DIBR. As a result, the input 2D image is converted into visually comfortable 3D image without the presence of artifacts enhancing the quality of the image in the display.

2.1. Block-Based Region Grouping

Computational complexity is reduced mainly by block-based algorithm. This implies each pixel in the same block has the same depth value. A 4-by-4 block is used as an example. Each
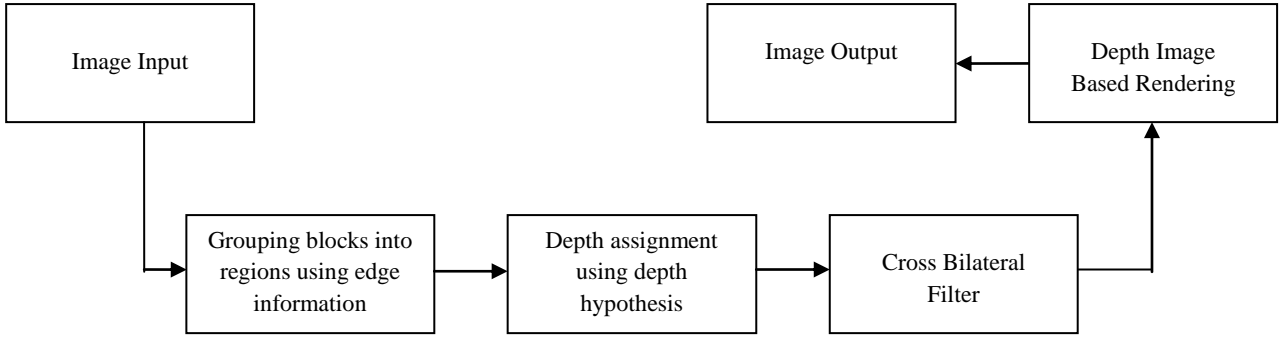
Figure 1: 2D-to-3D Conversion System

node is a 4-by-4 pixel block, and is four-connected. The value of each link is calculated by considering the absolute difference of the mean of neighboring blocks:

$$\text{Diff (a, b)} = \mid \text{Mean (a)} - \text{Mean (b)} \mid \qquad (1)$$

where a and b denote two neighboring blocks, respectively, and Mean (a) represents the mean color of a. A smaller value obtained implies a higher similarity between the two blocks. Following calculation of the absolute difference of the mean of the neighboring blocks, the blocks are then segmented into multiple groups by applying the minimum spanning tree (MST) segmentation. The flow of block-based region grouping is shown in the figure 2. Initially, a minimum spanning tree is constructed. And the multiple grouped regions are generated by removing the links of stronger edges. The MST algorithm is mainly used to identify the coherence among blocks with both the color difference and the connectivity without generating many small groups. This algorithm preserves the link connectivity and also has an excellent result in spatial locality [11].

A situation where MST algorithm possibly fails is when a series of regions, which differ slightly from each other, bridge a large gap in the feature space. As the edge weights are small, the graph will not be cut and so very different objects in terms of feature space characteristics will be merged. Recursive implementations of the MST algorithm, recalculating edge costs, can easily tackle the problem increasing complexity. With the efficient linkage preserving property, the MST segmentation method can generate excellent grouping results. The proposed depth generation method can also be substituted by other automatic or manual segmentation with satisfactory results.

2.2. Depth from Prior Hypothesis

The extraction of depth is the crucial one in the conversion process. The greatest difference between 2D and 3D image is the depth information. The object can jump out of the screen and look like a real life due to the depth information. If we extract these depth signals and integrate them together, we will build a strong foundation to make 3D images of better and higher quality. The depth generation algorithms are roughly classified into three categories which utilize different kinds of depth cues: the binocular, monocular and pictorial

depth cues. Each signal represents different depth information.

In this conversion process, following the generation of the block groups, the corresponding depth for each block is then assigned by the depth gradient hypothesis. The process includes the generation of gradient planes, depth gradient assignment, consistency verification of the detected region, and finally the depth map generation. When each scene change is detected, the linear perspective of the scene can be analyzed with the help of line detection algorithm using Hough transform [2]. The hypothesized depth gradient is derived based on the linear perspective information. The depth value of a given block group R is assigned by:

$$\text{Depth}(R) = 128 +$$

$$255\left\{\sum_{pixel\,(x,y)} W_{rl}\,\frac{x - \frac{width}{2}}{width} + W_{ud}\,\frac{y - \frac{height}{2}}{height}\right\} / pixel\_num(R)$$

$$(2)$$

where $\mid W_{rl}\mid + \mid W_{ud}\mid = 1$.

A larger value of the depth that is assigned implies a closer pixel to the user. The above equation suggests that the depth value is the gravity center of the block group that explains why each block group belongs to the same depth. The absolute value and sign of W_{rl} and W_{ud} can be adjusted to the left-to-right and top-to-bottom depth gradient weight. The orientation of depth gradient hypothesis can be derived from analysis of a geometrical perspective of the images. Analysis results indicate that the most important mode in the real world is the bottom-up mode. If the linear perspective information fails to detect the scene mode, then the bottom-up mode is the default mode to be selected.

2.3. Bilateral Filtering

The bilateral filter is non-iterative and also achieves satisfying results with only a single pass. This makes the filter's parameters relatively intuitive as their effects are not cumulated over several iterations. The bilateral filter has proven to be much useful although it is slow. It is nonlinear and also its evaluation is computationally expensive because the traditional accelerations like performing convolution after an FFT, are not applicable. Nonetheless, solutions have been proposed later in order to speed up the evaluation of the bilateral filter. Unfortunately, these methods seem to rely on approximations that are not grounded on firm theoretical foundations.

Among the variants of the bilateral filter, this conversion method has selected the cross bilateral filtering. In some
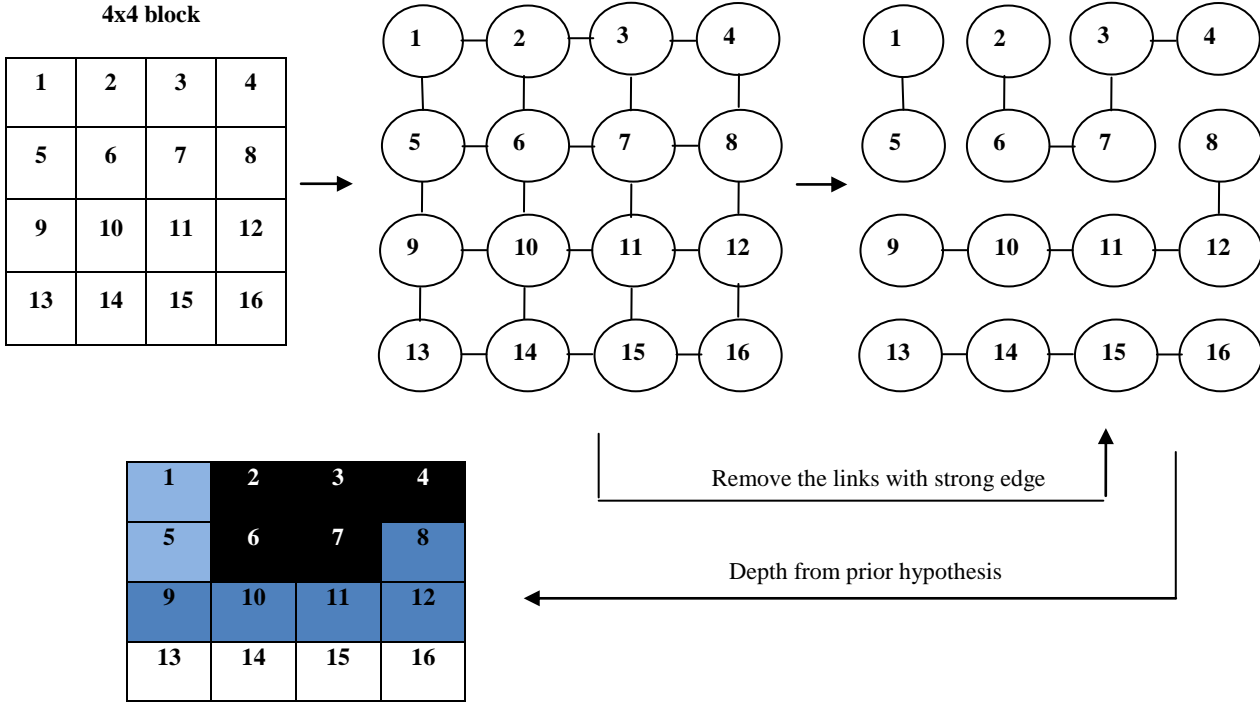
4x4 block

Figure 2: Flow of block-based region grouping

applications like computational photography, it is often useful to decouple the data to be smoothed to define the edges to be preserved. The chosen cross bilateral filter is a variant of the classical bilateral filter. This filter is used to smoothen the image to locate the edges to preserve. The depth map generated by block-based region grouping contains blocky artifacts. Here, the blocky artifacts are removed by using the cross bilateral filter, as expressed in the following equation:

$$\text{Depth}_f(x_i) =$$

$$\frac{1}{N(x_i)} \sum_{x_j \in \Omega(x_i)} e^{-0.5\left[\frac{|x_j - x_i|}{\sigma_s^2} + \frac{|u(x_j) - u(x_i)|^2}{\sigma_r^2}\right]} \text{Depth}(x_j) \quad (3)$$

$$N(x_i) = \sum_{x_j \in \Omega(x_i)} e^{-0.5\left[\sigma_x^2 |x_j - x_i| + \sigma_r^2 |u(x_j) - u(x_i)|^2\right]} \quad (4)$$

where $u(xi)$ denotes the intensity value of the pixel xi, $\Omega(xi)$ represents the neighboring pixels of xi, $N(xi)$ refers normalization factor of the filter coefficients and Depth_f is the filtered depth map. The cross bilateral filter here finely smoothens the depth map while preserving the object boundaries [9], [13]. The blocky artifact in the generated depth map is effectively removed while the sharp depth discontinuities along the object boundary are preserved.

2.4. Depth Image Based Rendering

The filtered depth map has a comfortable visual quality because the cross bilateral filter generates a smooth depth map inside the smooth region with similar pixel values and preserves sharp depth discontinuity on the object boundary. Following filtering by the cross bilateral filter, the depth map is then used for the generation of the left/right or multi-view images using depth image-based rendering (DIBR) [14] for 3D visualization.

Figure 3: DIBR system block diagram

DIBR for advanced 3D TV System can be illustrated by the following block diagram. This system includes three parts, pre-processing of depth map, 3D image Warping and Hole-Filling. Smoothing filter is first stage applied to smooth the depth image. Then the 3D image warping generates the left and right view according to the smoothed depth map and also intermediate view. If there are still holes in the image, hole-filling is then applied to fill color into these holes.

2.4.1. Pre-processing of depth image

Pre-Processing of depth image is usually a smoothing filter. Because depth image with that of the horizontal sharp transition may result in big holes after warping, smoothing filter is applied to smooth sharp transition to reduce the number of big hole. However, if we blur the depth image, we will not only reduce big holes but also degrade the warped view because the depth map of non-hole area is smoothed.

2.4.2. 3D Image Warping

3D image warping maps the intermediate view pixel by pixel to left or right view according to the pixel depth value. In the other words, 3D image warping transforms the location of pixels according to depth value. The 3D image warping formula is as following:

$$x_l = x_c + \left(\frac{t_x f}{2 Z}\right) \quad (5)$$

$$x_r = x_c - \left(\frac{t_x f}{2 Z}\right) \quad (6)$$

The x_l is the horizontal coordinate of the left view, and x_r is the horizontal coordinate of the right view. Besides, x_c is the horizontal coordinate of the intermediate view. Z is depth value of current pixel, f is camera focal length and t_x is eye distance. The formula shows that 3D warping maps pixel of intermediate view to that of the left and right view in horizontal direction.
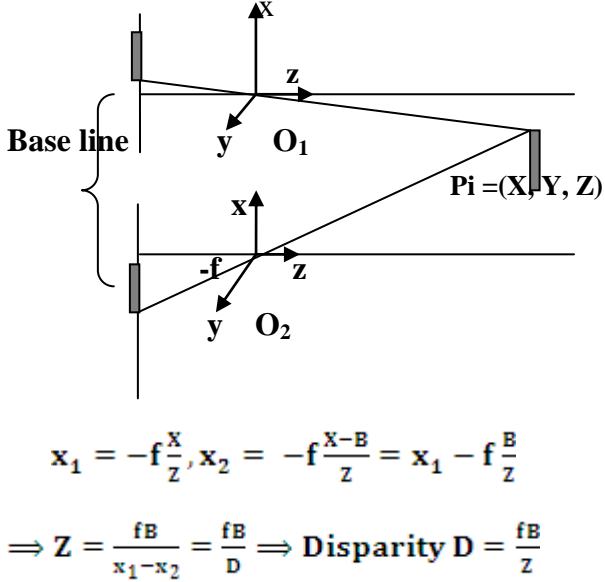

$$x_1 = -f\frac{X}{Z}, x_2 = -f\frac{X-B}{Z} = x_1 - f\frac{B}{Z}$$

$$\Rightarrow Z = \frac{fB}{x_1 - x_2} = \frac{fB}{D} \Rightarrow \text{Disparity } D = \frac{fB}{Z}$$

Figure 4: Equations of depth and disparity mapping for depth image based rendering

2.4.3. Hole Filling

Average filter interpolation method is a common method for Hole-Filling in DIBR. However, using average filter would result in artifacts at highly-textured areas. Besides, hole-size in DIBR is so huge such that it is needed to using average filter with large window size. At the same time, the average filter with large window size is unable to preserve edge information for the reason that edge information is blurred.

3. EXPERIMENTAL ANALYSIS

The proposed method generates excellent and comfortable visual quality since it is not restricted in horizontal line tracing. The proposed 2D grouping method has better results than the other single view based algorithm [7] especially in vertically standing objects. In contrast with the conventional motion-based algorithms [2], [3] and [8] in generating depth from multiple frames, this conversion method uses only a single image with only slight side effects. In finding the complexity of algorithm, a larger number of blocks in the frame imply a longer computational time since the MST segmentation algorithm has a highly sequential dependency. Thus, smaller the block size, shorter the computation time. However, a larger block size implies a lower depth map quality .This work also evaluated the visual quality of the proposed algorithm by comparing the four images. The block sizes are considered small for enabling the less computational
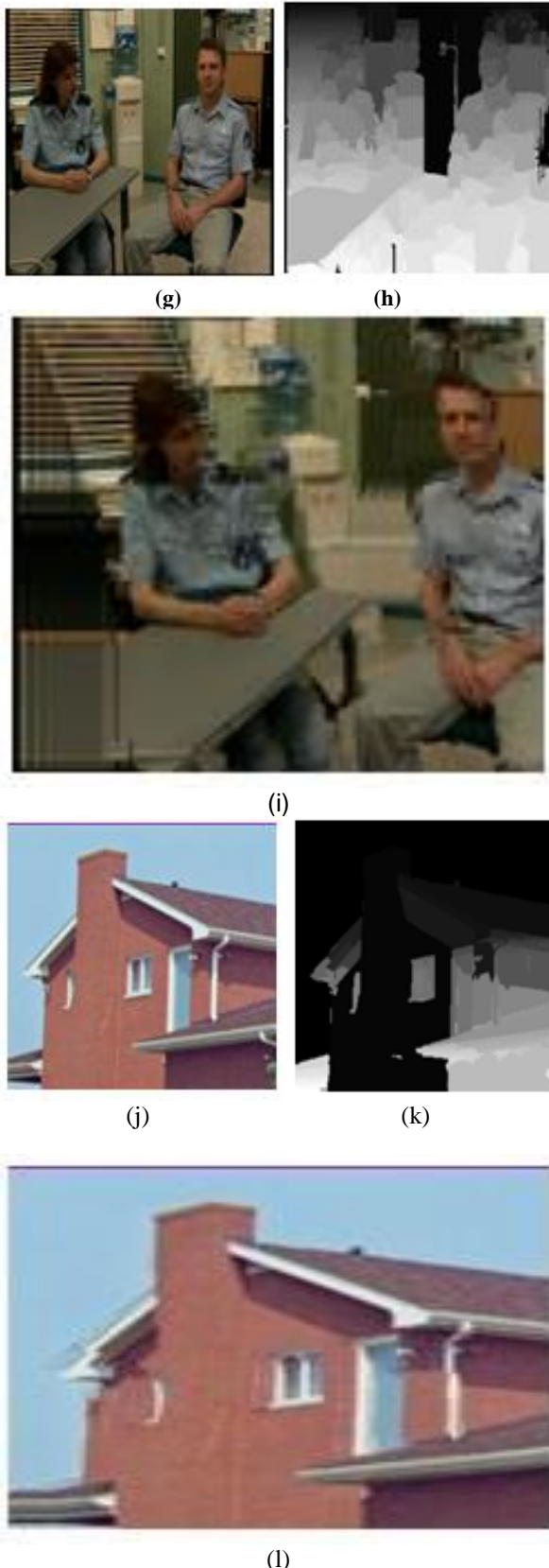

(a) (b)

(c)

(d) (e)

(f)

(g)   (h)

(i)

(j)   (k)

(l)
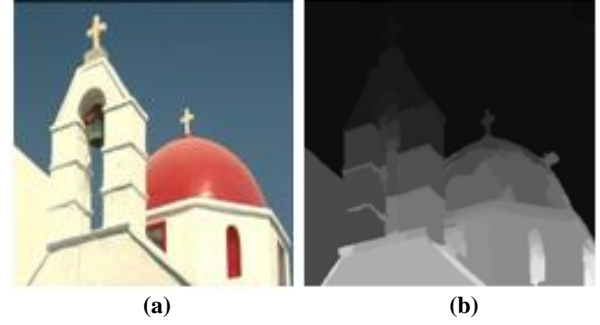
Figure 5: Experimental results showing four sets of images: 'church', 'image1', 'image2' , 'house'; (a), (d), (g), (j) are the original images; (b), (e), (h), (k) are the depth map assigned images; (c), (f), (i), (l) are the red cyan images.

The above results show some experimental analysis, including data on four sets of the original images, depth maps and red-cyan images. The results also indicate that the images violating the initial hypothesis still generate an acceptable quality. Thus the proposed method is assuring of high quality and comfortable visual images giving a 3D display from those data available in the input 2D images.

In figure 5, the input image 'church' gives a clear depth map regarding the nearest and the farthest objects. Thus this depth map enables greatly to give a clear and finely visualizing 3D display. The results are displayed in the first set of images. Considering the next image 'image1', the depth map is slightly blurred because of the distant objects in the image. So the display of the output image is also little blurred as if reproduced from that of the depth map of the image.

Similarly, the third image, 'image2' is also giving the clear view of the nearest and the farthest objects in its depth map and so the 3D display is clearly visualized when compared with that of the input image. Finally the image 'house' gives a clear depth map as if the initial image considered. So the output image is comfortably displayed preserving the edge information. The results are clearly shown in the fourth set of images in the experimental results.

4. CONCLUSION

This work has presented a novel 2D-to-3D conversion algorithm. The proposed algorithm utilizes edge information to group the image into coherent regions. A simple depth hypothesis is adopted here to assign the depth for each region and a cross bilateral filter is subsequently applied for removing the blocky artifacts. The proposed algorithm is quality-scalable as it depends on the block size. Smaller block size will result in better depth detail and that of the larger block size will have lower computational complexity. Capable of generating a comfortable 3D effect, this proposed algorithm is highly promising for 2D-to-3D conversion in case of 3D applications.

5. REFERENCES

[1] Chao-Chung Cheng, Chung-Te Li and Liang-Gee Chen, "A 2D-To-3D Conversion System Using Edge Information" in Proceedings of IEEE International Conference on Consumer Electronics, 2010

[2] C.-C. Cheng, C.-T. Li, P.-S. Huang, T.-K. Lin, Y.-M. Tsai, and L.-G.Chen, "A block-based 2D-to-3D conversion system with bilateral filter" in Proceedings of IEEE International Conference on Consumer Electronics, 2009

[3] Y-L. Chang, et al, "Depth Map Generation For 2D-To-3D Conversion By Short-Term Motion Assisted Color Segmentation" in Proceedings of ICME, 2007

[4] W. J. Tam, and L. Zhang, "3D-TV content generation: 2D-to-3D conversion," in Proc. ICME, pp. 1869-1872, 2006

[5] Sung-Yeol Kim, Sang-Beom Lee, and Yo-Sung Ho, "Three-dimensional natural video system based on layered representation of depth maps," in IEEE Transactions on Consumer Electronics, 2006

[6] Digital image processing using MATLAB by Rafael C. Gonzalez, Richard E.Woods, Steven L.Eddins

[7] Y.J.Jung, A. Baik, J. Kim, and D. Park, "A novel 2D-to-3D conversion technique based on relative height depth cue", in SPIE Electronics Imaging, Stereoscopic Displays and Applications XX, 2009

[8] D. Kim, D. Min, and K. Sohn, "A Stereoscopic Video Generation Method Using Stereoscopic Display Characterization and Motion Analysis", in IEEE Trans. On Broadcasting, Vol. 54, Issue 2, pp. 188-197, 2008

[9] S. Paris and F. Durand, "A fast approximation of the bilateral filter using a signal processing approach", in MIT Technical Report (MIT-CSAIL-TR-2006-073), 2006

[10] S. Battiato, S. Curti, E. Scordato, M. Tortora, and M. La Cascia,"Depth map generation by image classification", Three-Dimensional Image Capture and Applications VI, vol. 5302, pp. 95-104, 2004

[11] G. Economou, V. Pothos and A. Ifantis, "Geodesic distance and MST based image segmentation", in Proc. European Signal Processing Conf, 2004

[12] P. Harman, J. Flack, S. Fox, M. Dowley,"Rapid 2D to 3D conversion", in *Proc.SPIE Vol. 4660*, Stereoscopic Displays and Virtual Reality Systems *IX , 2002*

[13] C. Tomasi and R. Manduchi, "Bilateral filtering for gray and color images," in Proc. *ICCV,* pp. *839-846,* January 1998

[14] W.-Y. Chen and Y.-L. Chang and S.-F. Lin and L.-F. Ding and L.-G. Chen."Efficient depth image based rendering with edge dependent depth filter and interpolation," in Proc. ICME, pp. 1314-1317, 2005