3-D Model Generation from Single-view Image Using Object Classification

Jae-Il Jung and Yo-Sung Ho
Gwangju Institute of Science and Technology (GIST)
261 Cheomdan-gwagiro, Buk-gu, Gwangju 500-712, Korea
E-mail: {jijung, hoyo}@gist.ac.kr

***Abstract*- In order to generate a three-dimensional (3-D) scene from a single-view image, we need to estimate the depth map of the single-view image. In this paper, we propose a new method for object classification and depth assignment to generate a mesh-based 3-D model from a single-view image. After we classify video objects based on vanishing lines, we assign depth values to the objects. For objects hidden by other objects, we define a depth baseline and assign depth values from the baseline. After estimating the depth map, we generate a mesh-based 3-D model. Experimental results show that the proposed method generates a 3-D model successfully, reflecting properties of video objects pretty well.**

I. INTRODUCTION

Although two-dimensional (2-D) images are successfully exploited in various multimedia services nowadays, interest on three-dimensional (3-D) images is increasing gradually and 3-D image processing techniques attract more attention. In order to acquire a 3-D scene, we need special equipments, such as stereo or multi-view cameras and a depth camera.

Even if some 3-D image contents are available, the amount of 3-D contents is not enough to satisfy the user demand yet. On the other hand, there are abundant 2-D image contents captured by conventional single cameras. Hence, conversion of the 2-D contents to 3-D images can be an alternative solution to overcome the current discrepancy and fill up the lack of 3-D image contents.

However, it is not straightforward to convert a 2-D image to 3-D because we do not have any distance information between image objects and the camera. We call the distance information of each pixel in the 2-D image from the camera as the depth value, and the matrix of depth values for all the pixels as the depth map of the 2-D image, respectively. The depth map plays an important role in 2-D to 3-D conversion, and the accuracy of the depth map strongly affects the performance of the conversion operation.

Figure 1 shows a 2-D image and its corresponding depth map. As we can observe in the depth map, an image object close from the camera viewpoint has larger depth values than far objects. Although the depth map in Fig. 1 looks very reasonable, it is very challenging for us to find an accurate depth map. If we have multi-view images captured by stereoscopic or multiple cameras, we can estimate the depth map using stereo matching algorithms. However, it is much more difficult to estimate a depth map from a single-view image because there is no additional information, such as camera parameters and disparity information.

(a) (b)

Figure 1. (a) Single-view image and (b) corresponding depth map

Recently, there are several proposals to estimate the depth map from a single-view image. S. Batiato *et al.* generated a depth map in the following steps: generation of gradient planes, depth gradient assignment, consistency verification of detected region, and final depth map generation [1]. J. Ko *et al.* proposed an automatic conversion method based on the degree of focus of segmented regions and generated a stereoscopic image [2]. They utilized higher-order statistics to check the degree of focus. S. A. Valencia *et al.* presented a depth estimation method by measuring focus cues, which consists of a local spatial frequency measurement using multi-resolution wavelet analysis and a Lipschitz regularity estimation of significant edges [3]. Tam *et al.* found that the most critical depth information tends to be concentrated at object boundaries and image edges [4]. They generated the depth map in a single-view image using the Sobel edge detector.

However, those research works did not consider different types of image objects and simply assigned depth values to all the image objects using the same algorithm. This is not appropriate because any image can contain different types of objects. In this paper, we propose a new method for depth extraction and depth assignment using object classification in order to generate a more accurate depth map and produce a mesh-based 3-D model from a 2-D image.

II. BACKGROUND

In order to comprehend the proposed algorithm, we should understand several techniques. In this section, we explain the background about depth cues and an image segmentation mainly used in this algorithm.

A. Linear Perspective

The monocular depth cues can make people perceive depth in a 2-D image. There are many kinds of monocular cue, such as occlusion, relative size, texture gradient, and linear perspective. Among them, we use linear perspective as a main depth cue, in this paper.

Figure 2 shows the fundamental notions of linear perspective in the 2-D image. Parallel lines in the environment that are not parallel to the retinal image plane do not remain parallel in the 2-D image, as shown in Fig. 2.

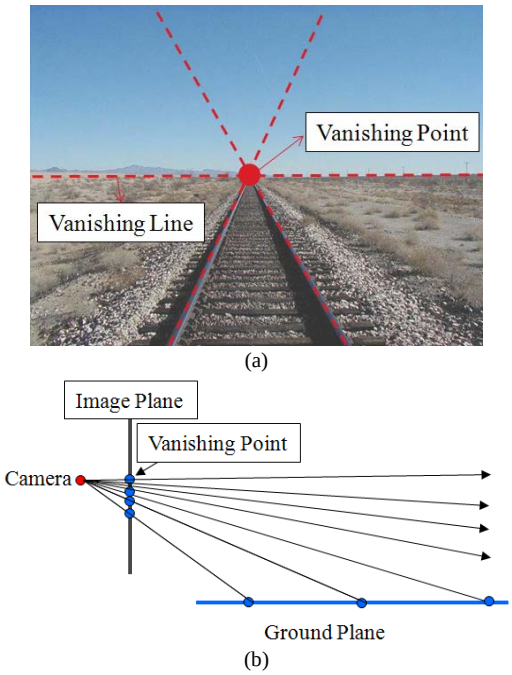


Figure 2. (a) The example of linear perspective depth cue, and (b) the fundamental notions of linear perspective

(a). With a depth, a distance between such lines become smaller and smaller and they seem to vanish at infinity. This infinity point is called as vanishing point. The vanishing lines are defined lines which pass through the vanishing point. This is the basis of projective geometry. This linear perspective helps us give depth cues. Texture surface in particular shows such a gradient due to linear perspective that provides additional cues about the depth of scenes. We use Canny edge detection [5] and Hough transform [6] to find vanishing lines and point.

B. Mean Shift Segmentation

An image can include various objects having different depth values. Before assigning proper depth values to these objects, we should divide the image into segments. Therefore, image segmentation which is the process of partitioning an image into multiple regions to collect pixels having similar property is an essential part for 2-D to 3-D conversion. The performance of the segmentation strongly influences the result of the conversion. In this paper, we use mean-shift algorithm [7] to segment a input image.

The process of mean shift segmentation is following. The image is converted into tokens according to color, gradients, and texture measures. The initial search window locations are found uniformly in the data. Then, the mean shift window location for each initial position is computed. The data are merged on the same peak or mode and clustered together. By this process, the input image is divided into several segments. These segments are utilized as a basic unit,

Figure 3. The segmented image by mean shift algorithm

for our algorithm. Figure 3 shows the segmenting results of the Fig. 2. (a).

III. PROPOSED ALGORITHM

In this section, we introduce our proposed algorithm. Figure 4 shows the flowchart of the proposed algorithm.

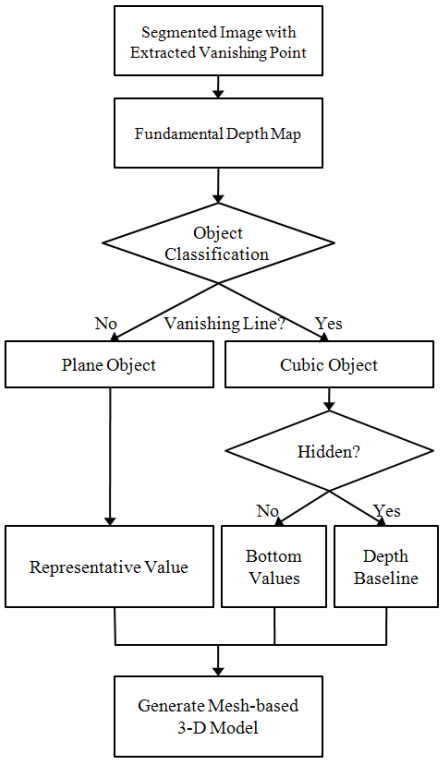


Figure 4. The flowchart of proposed algorithm

Firstly, we make a fundamental depth map on the basis of the linear perspective cue in a 2-D input image. Segmented objects then are classified into two types: plane object and cubic object. For the classification, we consider the inclusion relationship between the object and vanishing lines. The depth of plane objects is filled with a representative value of the fundamental depth map. The depth assigning methods for the cubic objects has two different types, according to whether the object is hidden by other objects or not. Finally, we generate a mesh-based 3-D model by considering the input 2-D image and the estimated depth map.

In the following part of this section, we would introduce our proposed algorithm in detail. For easy comprehension, we explain the proposed algorithm by using a simple example image shown in Fig. 5.

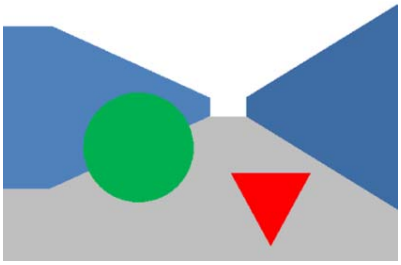
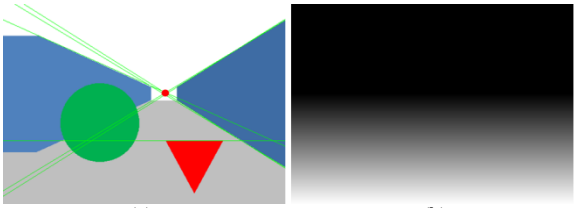


Figure 5. The example input 2-D image

A. Fundamental Depth Map Generation

As we mentioned in Sec. 2, we have segmented the input image by mean shift algorithm, and extracted vanishing lines and regarded the most overlapped point as a vanishing point. Figure 6. (a) depicts this process. The red point and green lines represent a vanishing point and vanishing lines, respectively.



(a) (b)

Figure 6. (a) Input single-view image, and (b) fundamental depth map

Then, the fundamental depth map is estimated on the basis of the vanishing point. Zero value is assigned to the upper area than the vanishing point, because only ground is referred when we assign depth values to the objects on the ground. In order to generate the fundamental depth map, we assign the depth values for bottom area by Eq. (1).

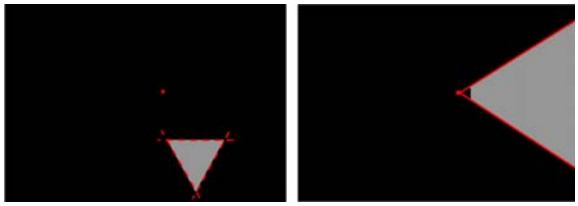
$$depth_y = \frac{255(y - Y_{VP})}{height - Y_{VP}} \quad (1)$$

In Eq. (1), y, Y_{VP}, and *height* represent the current row position, the row coordinate of the vanishing point and the height of the input image, respectively. By this equation, we can obtain the fundamental depth map shown in Fig. 6. (b). As you can see, the fundamental depth map reflects the real depth of the ground well and will be utilized as a reference depth during the following process.

B. Object Classification

After generating the fundamental depth map, we classify the objects into two types: plane object and cubic object. The plane objects are regarded as the object facing the perpendicular direction to camera ray, and have a singular depth value. The cubic objects are regarded as the object having the different depth values according to the distance from the vanishing point.

In order to classify the objects in the input 2-D image, we consider the inclusion relationship between the object and the vanishing lines. Figure 7 gives an assistance to understand the method of the classification.



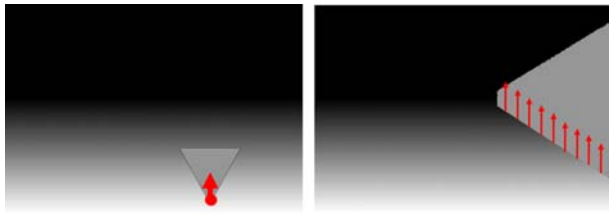
(a) (b)

Figure 7. (a) Plane object, and (b) cubic object

We move each object to individual binary image plane and detect the edge boundary of the object, as shown in Fig. 7. This movement can help the edge detecting process accurate. The solid and dotted line mean the vanishing line and a normal edge boundary. The triangular object in Fig. 7. (a) is classified to the plane type because it does not include vanishing lines. Contrary to the triangular object, the wall in Fig. 7. (b) includes vanishing lines. Therefore, it is classified to the cubic object. We apply this classification to whole objects in the input image.

C. Depth Assignment

In order to generate the final depth map, we assign the proper depth values to the objects. For the objects of plane type, we copy a singular depth value of the fundamental depth value, which is located at bottom position of the object and fill the whole region of the object with this value, as shown in Fig. 8. (a).



(a) (b)

Figure 8. Depth assignment method for (a) Plane object, and (b) cubic object

Contrary to plane objects, we copy the fundamental depth value and fill the columns. This process is repeated per every column line for the objects of cubic type. By this method, the objects have the gradient depth value according to distance from the vanishing point.

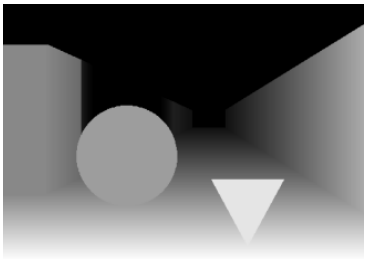


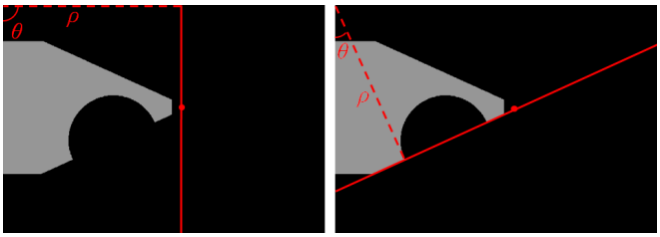
Figure 9. Depth map using simple assignment method

Figure 9 shows the depth map generated by this method. The depth values of the objects are successfully assigned according to the type of the objects. The circular and triangular objects have the singular depth value, and the wall has the gradient depth value by the distance from the vanishing point. However, this simply depth assignment for the cubic object has critical problem.

As you can see the region behind the circular object in Fig. 9, when the cubic object is hidden by other objects, wrong depth values are assigned. It is caused by the wrong information of the bottom boundary of the cubic objects. In order to overcome this problem, we define a depth baseline for these areas. This line acts as a guide line when we copy the depth value from the fundamental depth map. Limited Hough transform is used to find the depth baseline:

$$X_{VP} \sin\theta + Y_{VP} \cos\theta = \rho \qquad (2)$$

where X_{VP} and Y_{VP} represent the coordinates of the vanishing point. ρ and θ stand for the distance and the angle between the origin of the image and the candidate line, as shown in Fig. 10.

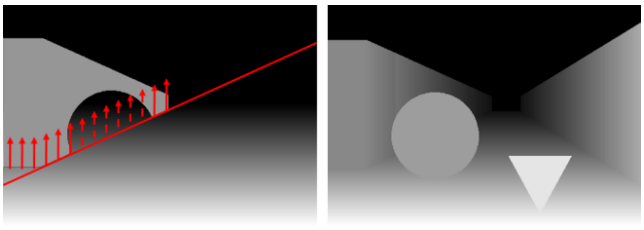


(a) (b)

Figure 10. Finding base line of depth:
(a) $\theta = 90°$ and (b) $\theta = 30°$

This equation means every line passing though the vanishing point. The candidate line rotates by changing θ and ρ values, until the line contacts the cubic objects, as shown in Fig. 10. Then, θ and ρ values are saved in the buffer memory for each cubic object.

After finding the depth baseline, we fill the hidden parts of the cubic object with the values of the fundamental depth values on the depth baseline. As can be seen in Fig. 11. (a), we control that this method is only applied to the hidden part. Figure 11. (b) is the final depth map by estimating this method. As you can observe, the depth value of the region behind the circular object is excellently modified. We generate 3-D model by using this depth map in the next section.



(a) (b)

Figure 11. (a) Depth assignment method using the depth baseline, and (b) final depth map

D. Mesh-based 3-D model generation

After estimating the final depth map, we generate the mesh-based 3-D model by considering the input 2-D image and the final depth map. Before generating the mesh model, we warp the 2-D image to the 3-D space to check the appropriateness of the final depth map. Figure 12 depicts the result of the warping.

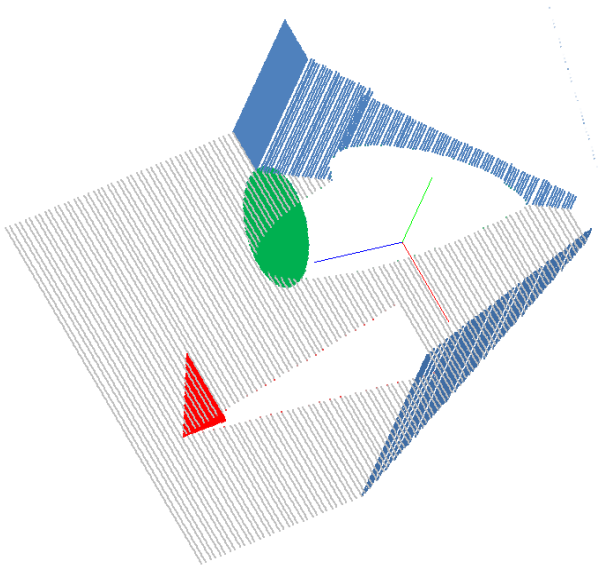
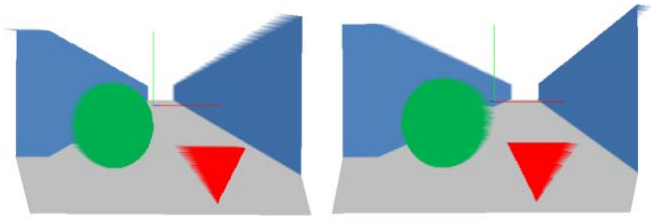


Figure 12. The warping result based on input 2-D image and the final depth map

As you can perceive from the input 2-D image, the triangular object is located in front of circular objects and walls. The warping result is well-matched to our perception. We generate mesh by binding three neighboring pixels with the final depth map. Figure 13 shows the mesh-based 3-D model captured at different camera position.



(a) (b)

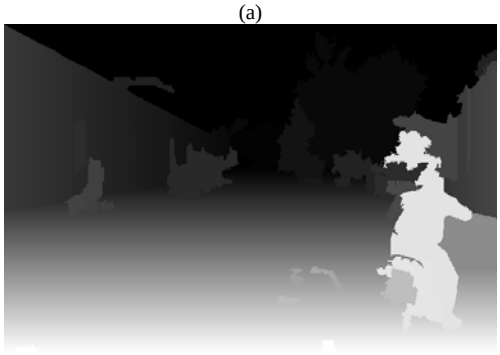
Figure 13. Mesh-based 3-D model
(a) left view and, (b) right view

IV. Experiment and Results

In order to evaluate the performance of the proposed algorithm, we take tests with two outdoor pictures having a linear perspective cue. Before applying our proposed algorithm, we have segmented the input image and extracted the vanishing point by the method mentioned in Sec 2. The first input image is shown in Fig. 14. The result of image segmentation and the final depth map are shown in Fig. 15. In the final depth map, the region of the person who rides a bike has a high singular depth value and the building has gradient depth values.

Figure 14. The first test image

(a)



(b)

Figure 15. (a) the result of image segmentation, and (b) the final depth map

We warp the input 2-D image to 3-D space as shown in Fig. 16.

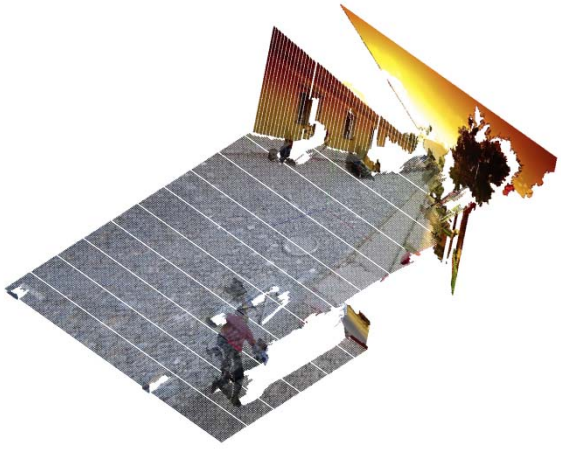


Figure 16. The warping result of the first test image

The warping result shows that the final depth map has the similar depth values to our perception. Based on this result, we generate 3-D model and capture this model at different viewpoints as shown in Fig. 17.

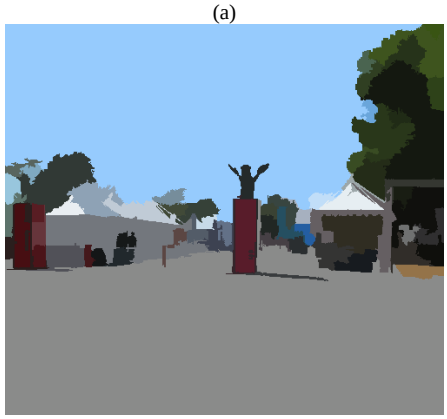
(a)

(b)

Figure 17. (a) The mesh-based 3-D model : (a) left view and (b) right view

The following figures show the experimental result of the second test image.

(a)



(b)

Figure 18. (a) The second test image, and (b) the result of image segmentation

Figure 18 shows the second test image and the result of image segmentation.

Figure 19. The final depth map of the second image

Figure 19 depicts the estimated depth map of the second input image. A pavilion in the image is classified to the cubic object and filled with gradient depth values.

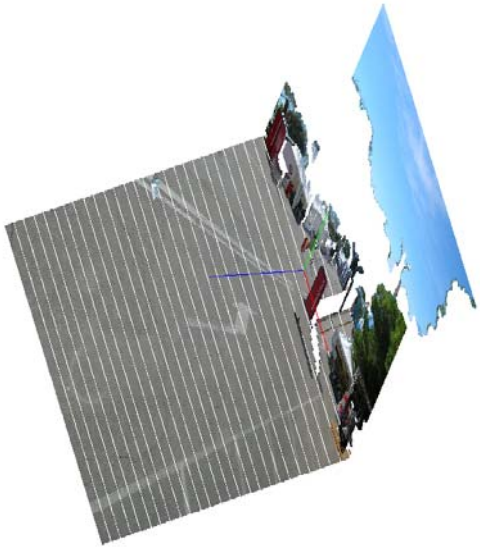


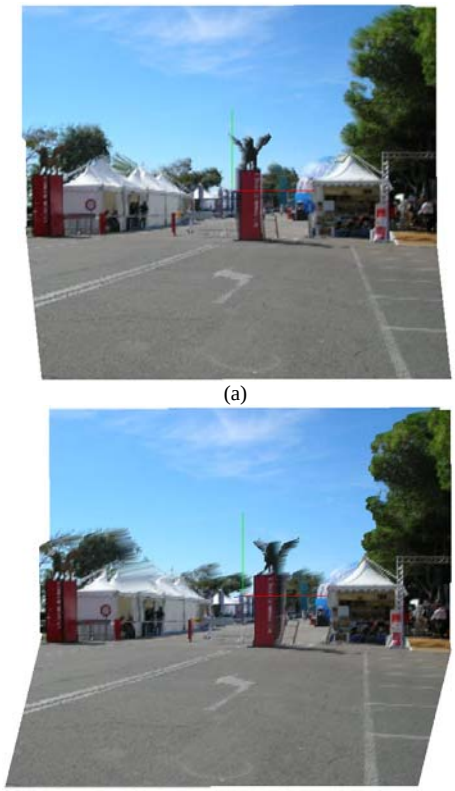
Figure 20. The warping result of the second test image

Through the warping result in Fig. 20, it can be confirmed that the positions of the objects are reasonable. Figure 21 is the mesh-based 3-D model generated by using Fig. 18 and Fig. 19.

As you can see in the experimental results, our proposed method can generate similar mesh-based 3-D model to our perception from the input image.

V. Discussion

In this paper, we propose object classification and depth assignment methods to generate mesh-based 3-D model from single-view image. We classify objects into two types, and assign depth values to objects. In order to overcome the hidden object problem, we define the depth baseline and assign depth values according to this line. After estimating the depth map, the mesh-based 3-D models are successfully generated. By using this model, we can see the single-view image at any viewpoints. We hope that the lack of 3-D contents can be solved by using our proposed algorithm.



(a) (b)

Figure 21. The mesh-based 3-D model : (a) left view and (b) right view

Acknowledgment

This work was supported in part by ITRC through RBRC at GIST (IITA-2008-C1090-0801-0017).

References

[1] S. Battiato, A. Capra, S. Curti, and M. La Cascia, “3D Stereoscopic Image Pairs by Depth-Map Generation,” International Symposium on 3D Data Processing, Visualization, and Transmission, pp. 124-131, Sept. 2004.

[2] J. Ko, M. Kim, and C. Kim, “2D-To-3D Stereoscopic Conversion:Depth-Map Estimation in a 2D Single-View Image,” in Proc. of the SPIE, vol. 6696, pp. 66962A, Aug. 2007.

[3] S. A. Valencia and R. M. Rodriguez-Dagnino, "Synthesizing Stereo 3D Views from Focus Cues in Monoscopic 2D mages,” in Proc. of the SPIE, vol. 5006, pp.377-388, Oct. 2003.

[4] W. J. Tam, F. Speranza, L. Zhang, R. Renaud, J. Chan, and C. Vazques, “Depth Image Based Rendering for Multivew Stereoscopic Display: Role of Information at Object Boundaries,” in Poc. of the SPIE, vol. 6016, pp.601609, Nov. 2005.

[5] J. Canny, “A Computational Approach to Edge Detection,” IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 8, no. 6, Nov. 1986.

[6] J. Illingworth and J. Kittler, “A survey of the Hough transform,” Computer Vision, Graphics, and Image Processing, vol 44, pp.87-116, Oct. 1988.

[7] D. Comanicu and P. Meer, "Mean shift: A robust approach toward feature space analysis," *IEEE Transaction on Pattern Analysis and Machine Intelligence*, vol. 24, pp. 603-619, May 2002.