# 3d Image Generation from Single 2d Image using Monocular Depth Cues

## B. Gnana Priya, M. Arulselvi

*Abstract: There has been a tremendous increase in the popularity of 3D hardware such as TV's, Smartphone's, gadgets for gaming, medical equipments, 3D printing and many more. 2D to 3D conversion is applied at various levels to get 3D content. In this paper, 3D image is generated from a single 2D image. we try to convert our own Karate and Bharathanatyam (KB) Dataset which contains both indoor and outdoor poses to 3D. Here, Watershed algorithm is employed to segment the image. Depth map is generated by sharpness and contrast as depth cues. The 3D image from single 2D image is created by depth image based rendering method.*

*Keywords: 2D to 3D conversion image conversion, depth cues, DIBR, KB Dataset, 3D Basics, Watershed algorithm.*

## I. INTRODUCTION

The humans have a stereoscopic vision which enables us to see the depth of an object. The depth perception is due to the binocular disparities of the two eyes. Similarly 3D effect can also be created by parallax effect by taking two images and layering them one over the other. A 3D image can be captured by a 3D camera or a 2D image can be converted to a 3D image. 3D camera comes in two flavours. In one type pair of lenses and image sensors is used to get a stereo pair. The other type uses a same camera and takes two pictures. In two lens camera the distance between the lenses are kept similar to the human eyes. When humans view things with their pair of eyes they perceive the depth information. Similar principle is employed here. The 3D processor present in the camera mixes the two images to form a single 3D image. The use of additional hardware makes it expensive and thus limited demand. The other type moves few inches left or right and takes picture with a single camera. This one is less expensive and more popular. Displaying a 3D image is more challenging task than creating them. The 3D display in a 3D camera (or) a 3D TV is used to view the 3D content taken.

Automatic and semi-automatic conversion techniques are available to convert 2D image to 3D. The different parts of image are assigned depths manually in semi automatic method. On the other hand an algorithm automatically calculates depth and assign for each part of image in automatic method.

**B. Gnana Priya,** Assistant Professor, Department of Computer Science and Engineering Annamalai University
**Dr. M. Arulselvi,** Associate Professor, Department of Computer Science and Engineering Annamalai University

Apart from pixel values, information regarding the geometry, scene and appearance are taken into consideration and stored as 3D file. Some of the 3D file formats that are commonly used are .3D, .3D2, .3D4,.3DM, .3DX, .ANIM, .BLEND, .C3D, .D3D, .MESH, .P3L ,etc.. Each one of the file format contains different types of attributes of a 3D image that they take into account. Therefore it becomes difficult to view or exchange 3D files between software. Some neutral formats like .STL and .DAE are available. We can convert our 3D file to this standard formats for interoperability. Open source software like Blender, Make3D, ConvertImage, AutoCAD, Etc., are available to create 3D models of 2D images.

## II. RELATED WORK

Namboodiri [1] proposed a model which finds depth from defocus. The blurring of each pixel is found and depth is assigned to it relatively. Many algorithms suggest the use of multiple depth cues for finding the depth. Han [3] uses Hough transform to find vanishing point. They propose the use of geometric cue to construct initial depth map and texture cue to refine the image after segmentation. Chiang [4] uses luminance from wavelet transform to detect foreground object which has high frequency. Higher depth values are assigned to nearer object and this method has better performance on close-up images. Liu [5] proposes a semantic based algorithm which segments image into categories like land, sky, building, etc. And assigns depth accordingly based on the category of the object. Jung [6] designed depth map using relative height that is based on depth refinement filter as well as line tracing method. Ideses [7] framed a depth map using the motion information given in H.264 bit streams. Yamada [8] uses colour theory and 3 depth scene models to produce a depth map. But, they are not suitable for different types of images and may produce temporal flicker since it ignores the temporal coherence in the depth maps that are between frames in the original 2D video.

Chang [9] Grouping of pixels plays a major role in depth estimation and maintaining uniformity of each and every one of the object. The major work is to find a correct depth relationship of all the objects in the given image. Tam[10]proposed that the major depth perceptions comes from binocular and monocular depth cues. Binocular cues are perceptions of both eyes whereas monocular cues are of a single eye.

856

Knorr [15] proposes structure from motion to analyze a frame sequence and using information from subsequent frames to generate a right view directly. No depth map is generated; instead the algorithm takes into consideration the camera properties and the desired disparity requirements of the generated view and then tries to find subsequent frames which match the requirements. They then expand on this idea in [16] by adding in information from camera position and object properties to improve the generated views.

The next group of methods relies on edge information and other colour cues in order to segment objects in a scene. Edge information is generally used to find complexity in objects, while colour information is used for segmentation purposes as well as a depth cue on its own as is seen in [20]. Chang et al. [21] use edge information along with a depth gradient hypothesis to create a depth map. The depth map is then filtered using a cross bilateral filter in order to smooth the edge boundaries of the depth map. DIBR is used to generate synthesized views for stereoscopic display.

## III.     PROPOSED WORK

The KB dataset which contains different 2D image poses from karate and Bharathanatyam dataset is used here. Fig [1] shows the overall framework of our system. Initially we segment 2D image using the watershed algorithm. The depth of each image is estimated using sharpness and contrast. We use monocular depth cues for estimating the depth of the images. Then cross bilateral filtering is used to preserve the edges and smoothing. All the karate poses were taken in outdoor and all the bharathanatyam poses taken indoor.
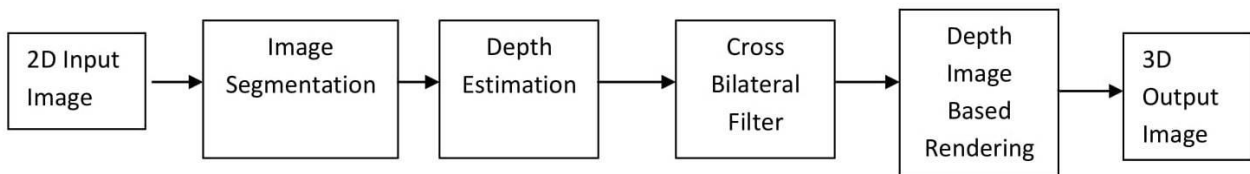


Fig(1): Architecture of the Proposed System

### A.  Image Segmentation with Watershed Algorithm

We use a marker based watershed algorithm. Since our dataset taken is very simple and have a single person per image, it is easy to identify the foreground and background. We label the foreground with one color and background labelled with other color. This marks the boundary regions of the object. Morphological opening and closing is used to remove small holes present.

### B.  Estimation of Depth

Inorder to generate a better 3D image, estimation of depth is very important. The depth map needs to be dense and accurate to produce a quality image. Depth cues that are derived from a single still image are called monocular depth cues. Here we use sharpness and contrast as the depth cues. Edges of a near object will have a sharp contrast than the far object. This will identify the near objects. The technique will make the centre of the image more closer than both the sides. Since all our images have a single person showing a pose that is centred, the method is providing a better result. Depth is expressed as a grayscale image composed of gray pixels in the range 0-255. The value 0 represents the 3D pixels are located far away in the 3D scene and 255 represent that the 3D pixels are located nearer in the scene.

### C.  Cross Bilateral Filter

To refine the depth map we use a cross bilateral filter. The bilateral filter is a weighted filter which evaluates the similarity of colors and distance between a current pixel and its neighbouring pixels. It then assigns the proper weight and calculates the weighted average.  A normal bilateral filter is used to perform edge smoothing. A special type of it called the cross bilateral filter is used in situations where we want smoothing based on color similarity or depth similarity. Here, we want to smooth one source based on the similarity of the other source. The kernel is a combination of weights based on the color similarity and depth similarity. The main function of the filter is, it will only smooth values with similar color and depth and keep the rest untouched. This filter provides efficient solution and also helps to prevent over segmentation.

### D.  3D image Construction

We use the already existing depth image based rendering scheme [26, 22] to convert the 2D image along with the generated depth map into a 3D image. Smoothing filter already applied to depth map to reduce the holes. Next step is image warping which maps left or right view pixel by pixel based on depth value. This changes the position of pixels along the horizontal direction, making the user to feel the 3D effect. Hole filling is employed to fill the holes that are caused by pixel warping. Most common method average filter interpolation is used here.

The warping formula is,

$$x_L = x_c + \left(\frac{t_x}{2} * \frac{f}{Z}\right)$$

857

$$x_R = x_c - \left(\frac{t_x}{2} * \frac{f}{Z}\right)$$

where,

$x_c$ - current pixel position of the image on x-axis

$x_L$ and $x_R$ - pixel position of left and right view on x-axis
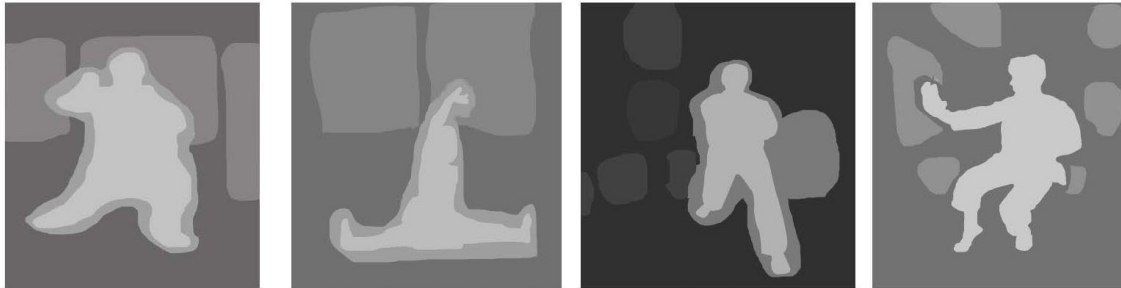
$t_x$ - baseline distance

$f$ - focal length

$Z$ - depth value of the current pixel.



Fig(2): Original Image from KB Dataset



Fig(3): Image after Watershed Segmentation



Fig(4): Depth Maps for the corresponding Images



Fig(5): 3D images obtained by DIBR method

## IV.   RESULTS AND DISCUSSIONS

Figure 2, shows some of the images from our KB dataset. Image Segmentation employed inorder to predict the edge information of the images and Figure 3, shows the results. In Figure 4 a clear depth map showing the near and far objects is given. Finally we use DIBR for 3D visualization of the 2D images shown in Figure 5.

| Category | Resultant 3D Image | Blurring Effect |
|---|---|---|
| Indoor Scenes (Less Color Tones) | Best & Clear 3D Effect | Minimal |
| Indoor Scenes (More Color Tones) | Good & Clear 3D Effect | Minimal |
| Outdoor Scenes (Less Color Tones) | Moderate Effect | Medium |
| Natural Sceneries (More Color Tones) | Less 3D Effect | Maximum |

**Table 1: Results on Various Categories of Images**

Since we use monocular depth cues sharpness and contrast the centre of the image gets highest attention. Figure 5 gives the actual result of our process. In all the images of the KB dataset we have got a single person showing different poses which is aligned centrally. Therefore for such pictures our algorithm works well than compared to natural sceneries. For images with variety of color tones such as natural scenes 3D output is produced with blurring effect.

## V. CONCLUSION

In this paper we propose a 2D to 3D image conversion scheme using monocular depth cues sharpness and contrast. We use watershed algorithm to segment the image and constructs a depth map using the depth cues. The cross bilateral filter is applied for smoothing. A general solution to convert any type of 2D image to 3D does not exist. This scheme provides better results for the KB dataset only. Natural scenes and more complex 2D images need enhancement of this model.

## REFERENCES

1. Namboodiri, , Chaudhuri,,” Recovery of Relative Depth from a Single Observation Using an Uncalibrated Camera”, Inf IEEE International Conference on Computer Vision and Pattern Recognition, 2008.
2. Burazerovic,, Vandewalle,, Berretty, ,” Automatic Depth Profiling of 2D Cinema - and Photographic Images”. IEEE International Conference on Image Processing, Cairo,2009.
3. Han, Hong,” Geometric and texture cue based depth-map estimation for 2D to 3D image conversion”, In: IEEE Inte. Con. on Consumer Electronics, 2011.
4. Chiang,., Tsai, T., Lin, Hsiao,” Fast 2D to 3D conversion based on wavelet analysis”, In: IEEE International Conference on Systems Man and Cybernetics,2010.
5. Liu,, Gould, Koller, “ Single image depth estimation from predicted semantic labels”, In: IEEE Conference on Computer Vision and Pattern Recognition, pp. 1253–1260,2010.
6. Jung, Y, Baik, Kim, Park, “ A Novel 2D-to-3D Conversion Technique Based on Relative Height-Depth Cue”, Proc. SPIE 2009.
7. Ideses, Yaroslavsky, Fishbain, “ Real-time 2D to 3D video conversion”,. Real-Time Image Process. 2007.
8. Yamada, Suzuki, ,”Real-Time 2D-to-3D Conversion at Full HD 1080P Resolution”, Proceedings of the IEEE 13th International Symposium on Consumer Electronics May 2009; pp. 103–106.
9. Chang, “Depth Map Generation For 2D-To-3D Conversion By Short-Term Motion Assisted Color Segmentation” in Proceedings of ICME, 2007
10. Tam, and Zhang, “3D-TV content generation: 2D-to-3D conversion,” in Proc. ICME, 2006.
11. Cheng, , C.T. Li, and Chen, “ A 2D-to-3D Conversion System using Edge Information”, Proc. IEEE Conf. On Consumer Electronics (ICCE), 2009.
12. Cheng. and Liang,” Depth Map Generation based on Scene Categories”, SPIE Jnl. Of Electronic Imaging, vol. 18, no. 4, October–December 2009.
13. Jung, and . Ho, “ Depth Map Estimation from Single-View Image using Object Classification based on Bayesian Learning”, Proc. IEEE Conf. 3DTV , 2010.
14. Agnot,, Huang, and Liu., “ A 2D to 3D video and image conversion technique based on a bilateral filter” , In Proc. SPIE Three-Dimensional Image Processing and Applications, volume 7526, Feb. 2010.
15. Knorr and Sikora, “ An image-based rendering (ibr) approach for realistic stereo view synthesis of tv broadcast based on structure from motion”, In Image Processing, 2007. ICIP 2007. IEEE International Conference on, volume 6, pages VI –572 –VI –575, Oct. 2007.
16. Knorr, Smolic, and Sikora, “ From 2d- to stereo- to multi-view video”, In 3DTV Conference, 2007, pages 1 –4, May 2007.
17. Durand, and Dorsey, “ Fast bilateral filtering for the display of high-dynamic-range images”, ACM Trans. Graph., 21:257-266, July 2002.
18. Konrad, Brown, Wang, Ishwar, Wu, and Mukherjee, “ Automatic 2D-to-3D image conversion using 3D examples from the Internet”, In Proc. SPIE Stereoscopic Displays and Applications, volume 8288, Jan. 2012.
19. Saxen ,Sun, and Ng, “ Make3D: Learning 3D scene structure from a single still image”, IEEE Trans. Pattern Anal. Machine Intell., , May 2009.
20. Tsai, Cheng, Li, and Chen,” A real-time 1080p 2d-to3d video conversion system”, In Consumer Electronics (ICCE), 2011 IEEE International Conference on, Jan. 2011.
21. Cheng, Li, and Chen, “ A 2d-to-3d conversion system using edge information”, In Consumer Electronics (ICCE), 2010 Digest of Technical Papers International Conference on, pages 377 –378, Jan. 2010.
22. Silva V, “ Depth image based stereoscopic view rendering for MATLAB”, available at http://www.mathworks.com/matlabcentral/fileexchange/27538-depth-image-based-stereoscopic-view-rendering, 2010.
23. Dejohn,., Seigle, “ A summary of approaches to producing 3D content using multiple methods in a single project”, Report, In-Three, 2008.
24. Graziosi, Tian , Vetro , “ Depth map up-sampling based on edge layers”, Signal Information Processing Association Annual Summit and Conference, CA, pp. 1–4, 3–6 December 2012.
25. Ideses, Yaroslavsky , Fishbain , “Real-time 2D to 3D video conversion” , Journal of Real-Time Image Processing, vol. 2, pp. 3–9, 2007.
26. Fehn ,” A 3D-TV Approach Using Depth image based Rendering” , in proceedings of SPIE, January 2004

## AUTHORS PROFILE

**B. Gnana Priya,** received M.E degree in computer science from Annamalai University, Tamil Nadu, India, in 2007. She is a Research scholar in Annamalai University and is working as Assistant Professor in the Department of Computer Science and Engineering, Annamalai University and is having experience of more than 15 years. Her research interests include Machine learning, Image Processing, computer vision and Data Analytics. She is a member of the Computer Society of India. She has publications in leading journals and international conferences/workshops.

**Dr. M. Arulselvi,** received PhD degree in computer science from Annamalai University, Tamil Nadu, India, in 2014. She is working as Associate Professor in the Department of Computer Science and Engineering, Annamalai University and is having a vast experience of more than 17 years. Her research interests include Image Processing, computer vision and machine learning. She is a member of the Computer Society of India. She has extensive publications in leading journals and international conferences/workshops.

859