International Journal on Recent and Innovation Trends in Computing and Communication ISSN: 2321-8169
Volume: 3 Issue: 6 3925 - 3929

Depth Estimation for 2D-to-3D Image Conversion Using Scene Feature

Shweta Patil
Dept. of Electronics and Tele-communication,
D. Y. Patil Collage of Engineering, Akurdi,
Pune, India.
shwetanjali.patil@gmail.com

Priya Charles
Dept. of Electronics and Tele-communication,
D. Y.Patil Collage of Engineering, Akurdi,
Pune, India.
prinnu@yahoo.com

***Abstract*—** In this modern era 3D supportive hardware popularity is increased but the demand for 3D contents and there availability is not matching. They are still dominated by its 2D counterpart hence there is need of 3D contents. While doing 2D-to-3D image or video conversion depth estimation is a key step and a bit challenging procedure. There are distinct parameters that can be considered during conversion like, structure from motion, defocus, perspective geometry, etc. Until now many researchers have been proposed different methods to close this gap by considering one or many parameters. In this paper for depth estimation, conversion using scene feature is used. Here color is chosen as a scene feature. Intensity information is used here to estimate depth image, hence RGB to HSV conversion is performed from which Value (V) deals with intensity information. RGB to HSV conversion is implemented on FPGA. The proposed method is prototyped on Spartan 3E FPGA based developing board and MATLAB.

Keywords- *3D hardware, 2D-to-3D conversion, depth estimation, scene feature*

*****

I. INTRODUCTION

Today there is an enhancement in 3D capable hardware such as 3D TVs, Blu-Ray players, gaming consoles, 3D cameras, 3D projectors and smart phones and many more. These 3D media gives feeling of immersion or more lifelike viewer experience. But the availability of 3D content is not matching with its production rate. There are two methods for generating 3D contents. First, capture the content directly with multiview method and other is to take 2D conventional footage and converts it to 3D. Multiview method gives best results but it is difficult and expensive as it requires specialized high resolution cameras and other costly equipments as well as production system should be strong [1]. The latter method is difficult but may be cost effective. Using this method large amount of available 2D data can be converted into 3D, instead of creating new [2].

A typical 2D-to-3D conversion process consists of two steps: first is depth estimation for a given 2D image and then depth based rendering of a query image in order to form a stereo pair images. The latter step of rendering is well realized and many algorithms are available that produce good quality results. The first step of depth estimation from a single image or video frame is a bit challenging, because it not having much information about depth [3]. Fig 1 shows 2D-to-3D image or video conversion procedure. 2D monocular image or video frame is given to depth generation block as an input. Using depth generation algorithm depth map is generated. Depth map is a monochromatic image, where a low intensity indicates a far distance from the camera, while a high intensity indicates a close distance. Depth map image is given to Depth Image Based Rendering (DIBR) for generating stereoscopic 3D images or multiple images.

Up till now many researchers have proposed different algorithms considering different approaches of image or video frame which includes structure from motion, defocus, shape from shading, perspective geometry, scene features, etc. 2D-to-

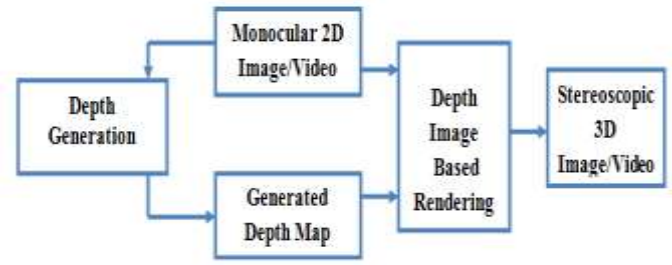


Figure 1.General 2D-to-3D image/video conversion procedure

3D conversion using scene features contain scene features like color, shape and texture, edges, segmentation and analysis of environment [4]. Pixel level attributes are considered for doing conversion and color and edge are significant attributes from them. Color is one of prominent feature that is used here because every pixel in the image is separated into a perceptual color space and hence these components can be used to determine a sense of depth. In some papers warm/cool color theory is used for conversion [5] [6].

The rest of paper is organized as follows. Section II reviews related work of 2D-to-3D conversion. Section III discusses the

proposed method. Section IV shows experimental results. Finally, the paper is concluded in Section V.

II. Related Work

Basically there are two approaches of 2D-to-3D conversion; one is automatic and another is semi-automatic. In semi-automatic method a skilled operator assigns depth to various parts of an image or video frame. Based on this marked sparse depth assignment, algorithm estimates dense depth over the entire image or video sequence. In an automatic method human intervention is not required; computer algorithm automatically does the whole estimation of the depth from a single image or video. Semi-automatic method is more successful but it is time consuming and costly [8]. Many films have been converted to 3D using this approach. For real time conversion automatic method is only preferable; but these methods can be applicable to restricted scenes i.e. algorithm for indoor images cannot be applied to outdoor images [8].

The depth estimation from a single 2D image or video frame, which is the key step in 2D-to-3D conversion, can be estimated in various ways, including depth from defocus [1] [9] [10] [11], depth from perspective geometry [12] [13], depth from models [14], depth from visual saliency [15], depth from motion [16], depth using scene features and so on. In the defocus based approach image defocus information is used to determine distance of objects. Previously for defocus information multi images or edge information or camera systems with limited class of point spread functions were used. Recently camera parameters are not used and hence it required two or three images for processing [9] [13]. Perspective geometry considers the fact that parallel lines in real world tend to converge at a point known as vanishing point. Normally, the vanishing point has the farthest distance, because it is the intersection of the projections of set of parallel lines in space on to the picture plane. Hence it is possible to assign depth, based on the position of the lines and the vanishing points [12]. The depth from models approach uses several depth models of typical 2D scenes and merges them to retrieve the depth. In [14] three depth models used are: spherical surface model, cylindrical and spherical surface model and plane and cylindrical surface model. Visual saliency is another type, based on the analysis of visual attention and a saliency map. This multiple cues acts directly as a depth map [17]. Depth from motion uses principle that near objects move faster across the retina than far objects for a moving observer, hence relative motion provides an important depth cue [18]. In this type motion vectors are used to generate depth information. Recently many researchers have been proposed algorithm using machine learning techniques to estimate the depth map from a single monocular image [1] [3] [8] [9]. Thus every algorithm proposed for conversion has considered either one or more cues for depth map estimation. Like in [18] multi cues are considered; visual saliency, defocus and color depth models. On the 2D image one of the method is applied considering scene in image.

In [19] RBG to HSV conversion is used as they want grayscale image. One method for that is directly use RGB values and eliminates one of the RGB components. But in this method object shape distortion is caused. Hence they have used HSV color model to maintain the original shape of the object. In [3] YUV to HSV conversion is used. They have used color transformation for that pixel attributes of hue (H) and value (V) is used. In [20] for 2D to 3D automatic conversion, they have used classification technique and edge based depth cue. For that first image scene is classified in two types, sky/ground type and normal type. To identify the pixel whether it belongs to sky or ground HSV color space is used. They have used threshold values of H, S and V to identify its category whether sky/ground type or normal type. In [21] application control using 3D gesture recognition is proposed. Here in image segmentation stage, segmentation of the image by removing background using HSV and YCbCr threshold algorithm are used. First RGB to YCbCr conversion is done and thresholds are for skin color identification. Though this YCbCr model detect skin color effectively but this YCbCr model fails when background is having flash or lighting effect, or shadow of hand etc. So to overcome this limitations HSV model is used. Though there are many algorithms available still each is having its own strengths and weakness.

III. PROPOSED METHOD

In this paper depth estimation using scene feature is proposed which will be used in 2D-to-3D conversion. Fig 2 shows block diagram of proposed method.

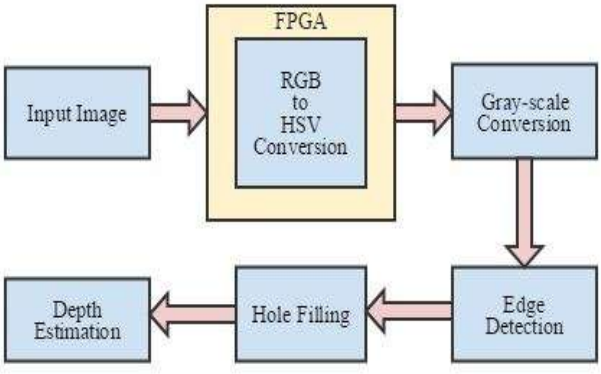


Figure 2. Block Diagram of Proposed Method

In image processing there are various color models: RGB model, CMY model, HSV model and YIQ model. RGB color model is used for color monitors and color video cameras, CMY model for printers, YIQ model is used in color TV broadcasting. Here RGB color model is converted into HSV color model. RGB (R-Red, G-Green and B-Blue) is a color model appears in primary colors and based on a Cartesian

coordinate system. HSV is one of the frequently used color model [7]. In HSV, H stands for hue which specifies the position of pure color on color wheel. Thus hue is related with dominant wavelength in a mixture of light wave. S is saturation, gives measure of the degree to which pure color is diluted with white. V is value called as lightness of color. Sometimes it is represented as I (intensity) or B (brightness). HSV model is having advantages over RGB model. HSV is strong model than RGB because it offers a more intuitive representation of the relationship between colors; it selects more specific color. RGB is costly in terms of computation time. The way in which human beings perceive color; hue and saturation components from HSV relates same way.

The interested 2D query image is read in MATLAB. As seen in fig 2, FPGA is used to convert RGB to HSV conversion step. Because of memory constraint of FPGA it is not possible to send whole image at a time to FPGA for computation. Hence row by row, serially pixels are sent to FPGA for computation. These processed pixels are sent back to MATLAB to generate output image which is HSV image.
The rest of operations are performed in MATLAB. For further operations the input image is converted in grayscale image.

Next significant step is canny edge detection. Canny edge detection is a standard edge detection method to detect edges in forceful manner. In canny detection as Gaussian filter is used any noise present in an image can be removed. Along with this advantage it enhances the signal to noise ratio. Once canny edge detection detects object, they are filled using *imfill* MATLAB command. But using only canny edge detection it is not possible to detect maximum whole objects. Since if the object is detected then only it is possible to fill it. So to overcome this problem morphological dilation operation is applied to an image. Once the objects are detected they are filled to estimate final depth values in image. Here to fill these objects the intensity values that are calculated in RGB to HSV conversion are used. Along with the intensity values, the min and max values from that object block is calculated and using them new depth value for that object block is estimated. Thus we are getting final depth map image.

The steps performed during the algorithm are given as follows:

Step1: First load interested query image in MATLAB using GUI.

Step2: Send image to FPGA for RGB to HSV conversion. From MATLAB pixels are sent row by row serially for computation on FPGA. The HSV image sent from FPGA is generated in MATLAB.

Step3: Convert input image to grayscale image.

Step4: Perform Canny edge detection to detect objects.

Step5: Perform whole filling operation.

Step6: Fill the detected objects with new values to estimate depth map image and generate final depth map image.

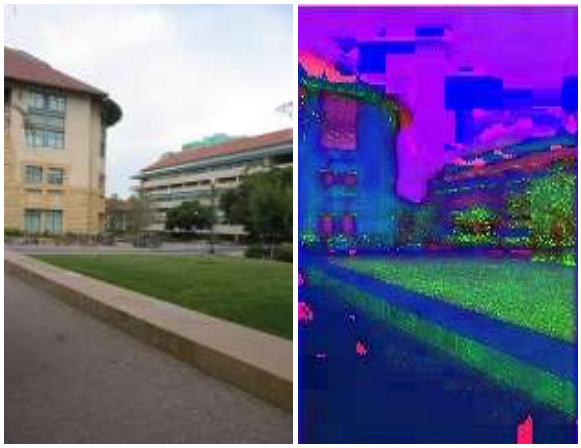
IV. EXPERIMENTAL RESULTS

The proposed hardware architecture of 2D-to-3D system is prototyped using Spartan 3E FPGA based Papilio developing board and MATLAB. The FPGA part has been coded in Arduino IDE using subset of C and rest of steps has been coded in MATLAB. Table I shows the result of what resources are used for a device.

Table I Xilinx Utilization

Xilinx XC3S500E			
Logic Utilization	**Used**	**Available**	**Utilization**
Number of Slice Flip Flops	908	9,312	9%
Number of 4 input LUTs	2,360	9,312	25%
Number of occupied Slices	1,561	4,656	33%
Number of RAMB16s	10	20	50%

As illustrated in literature review, various proposed methods by many researchers have considered different attributes of image and hence those algorithms are useful only for that type of images e.g. algorithm for indoor images cannot be applied to outdoor images. Our algorithm can work for state-of-the-art images. We have tested our approach on two databases: Make3D Dataset [22] and Middlebury [23]. Figure 3 shows the results of proposed method with the final output and the output images of intermediate steps.



(a) (b)

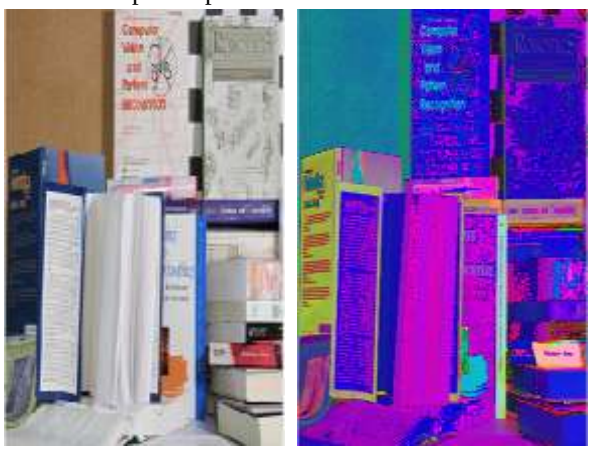
 ISSN: 2321-8169
3925 - 3929

(c) (d)

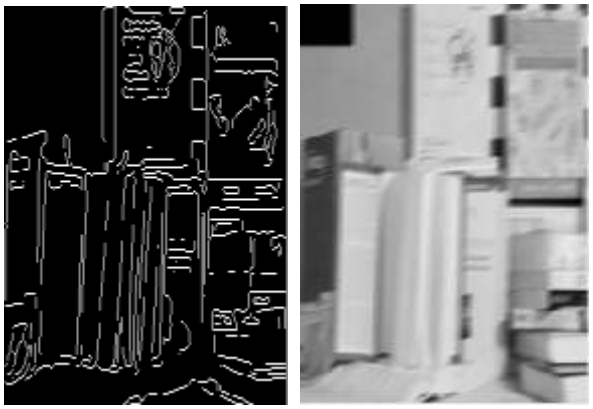
(e)

Figure 3. (a) RGB Input Image (b) HSV Output (c) Edge Detection Output (d) Final Depth Output (e) Output of Global Method

Fig 3(a) is RGB input image taken from Make3D dataset, fig 3(b) is HSV image computed using FPGA, fig 3(c) is output of canny edge detection, fig 3(d) is final depth image of proposed algorithm and fig 3(e) is depth image of global method proposed in [8]. Fig 4(a) is RGB input image taken from Middlebury dataset, fig 4(b) is HSV image computed using FPGA, fig 4(c) is output of canny edge detection, fig 4(d) is final depth image of proposed algorithm. Experimental results shows that the output from this method gives more clear and smooth depth map.



(a) (b)



(c) (d)

Figure 4. (a) RGB Input Image (b) HSV Output (c) Edge Detection Output (d) Final Depth Output

V. CONCLUSION

This paper presents algorithm of depth estimation which will be used in 2D-to-3D conversion based on scene feature. The depth is a very significant cue of a scene and hence it is necessary to extract it for many applications. So here color is used as a scene feature. Value (V) component from HSV which deals with intensity of image is used here to estimate depth. Along with color canny edge detection has used which is helpful for detecting objects. The experimental results show that proposed method works on any image and gives smooth depth map image.

REFERENCES

[1] M. Guttmann, L. Wolf, and D. Cohen-Or, "Semi-automatic stereo extraction from video footage," in *Proc. IEEE Int. Conf. Comput. Vis.*,Oct.2009,pp.136–142.

[2] R. Phan, R. Rzeszutek, and D. Androutsos, "Semi-automatic 2D to 3D image conversion using scale-space random walks and a graph cuts based depth prior," in *Proc.18 IEEE Int. Conf. Image Process.*,Sep. 2011, pp.865-868.

[3] J. Konrad, M. Wang, and P. Ishwar, "2D-to-3D image conversion by learning depth from examples," in *Proc. IEEE Comput. Soc. CVPRW*, Jun. 2012, pp. 16-22.

[4] Raymand Phan, Richard Rzeszutek and Dimitrios Androutsos "Literature Survey on Recent Methods for 2D to 3D Video Conversion", *Multimedia Image and Video Processing*, Second Edition. Mar 2012, 691-716.

[5] K. Yamada and Y. Suzuki, "Real-time 2D-to-3Dconversion at full HD1080P resolution", *the 13th IEEEInternational Symposium on Consumer Electronics*, 2009,pp.103-107.

[6] Shao-Jun Yao; Liang-Hao Wang; Dong-Xiao Li; Ming Zhang, "A Real-Time Full HD 2D-to-3D Video

Conversion System Based on FPGA," *Image and Graphics (ICIG), 2013 Seventh International Conference on* , vol., no., pp.774,778, 26-28 July 2013.

[7] Rafel C. Gonzalez, Richard E. Woods, Digital Image Processing, Prentice Hall, 2008.

[8] J. Konrad, M. Wang, and P. Ishwar,C.Wu,D. Mukharjee, "Learning based, automatic 2D-to-3D image and video conversion," in *Image Processing IEEE Trans on*,vol.22,no.9,pp.3485-96,Sept. 2013 Jun. 2012, pp. 16-22.

[9] K. Karsch, C. Liu, and S. B. Kang, "Depth extraction from video using non-parametric sampling," in *Proc. Eur. Conf. Comput. Vis.*, 2012, pp.775–788.

[10] J. Ens and P. Lawrence, "An investigation of methods of determining depth from focus," *IEEE Trans. Pattern Anal.Mach. Intell.*, vol. 15, no. 2, pp. 523–531, 1993.

[11] S. A. Valencia and R. M. Rodriguez-Dagnino,"Synthesizing stereo 3D views from focus cues inmonoscopic 2D images," in *Proc. SPIE*, 2003, vol. 5006, pp.377–388.

[12] S. Battiato, S. Curti, M. La Cascia, M. Tortora, and E.Scordato, "Depth map generation by image classification," in *Proc. SPIE*, Apr. 2004, vol. 5302, pp. 95–104.

[13] X. Huang, L.Wang, J. Huang, D. Li, and M. Zhang, "Adepth extraction method based on motion and geometry for 2D to 3D conversion", in *3rd Int. Symp. Intell. Inf. Technol.Appl.*, 2009, pp. 294–298.

[14] K. Yamada and Y. Suzuki, "Real-time 2D-to-3Dconversion at full HD1080P resolution", *the 13th IEEEInternational Symposium on Consumer Electronics*, 2009,pp.103-107.

[15] C. Huang, Q. Liu and S. Yu, "Regions of interestextraction from color image based on visual saliency",*Springer Science Business Media*, 2010.

[16] E. Imre, S. Knorr, A. A. Alatan, and T. Sikora,"Prioritized sequential 3D reconstruction in video sequencesof dynamic scenes," in *IEEE Int.Conf. Image Process. (ICIP)*,Atlanta, GA, 2006.

[17] J. Kim, A. Baik, Y. J. Jung, and D. Park, "2D-to-3Dconversion by using visual attention analysis," in *Proc. SPIE7524*, Feb. 2010, 752412.

[18] P.Ji,L.Wang,D. Li, M. Zhang, "An automatic 2D to 3D conversion algorithm using multi-depth cues," *IEEE Conf.Audio,Language and Image Processing*,pp.546-50,July 2012.

[19] Lim Zhao Yi; Mohamed, S.S., "3D model capture system," *Computer Applications Technology (ICCAT), 2013 International Conference on* , vol., no., pp.1,5, 20-22 Jan. 2013.

[20] Hao Dong; Shouyi Yin; Weizhi Xu; Zhen Zhang; Rui Shi; Leibo Liu; Shaojun Wei, "An automatic depth map generation for 2D-to-3D conversion," *Consumer Electronics (ISCE 2014), The 18th IEEE International Symposium on* , vol., no., pp.1,2, 22-25 June 2014.

[21] Ashutosh Verule, Shrivardhan Suryawanshi, Poonam Rajput, Rajashri Itkarkar, " Application control using 3D gesture recognition," *International Conference on Convergence of Technology In Press*, 2014.

[22] Make3D: http://make3d.cs.cornell.edu/data.html

[23] Middlebury: http://vision.middleburry.edu/stereo