RESEARCH ARTICLE OPEN ACCESS

Automatic 2D to 3D Stereoscopic Conversion

Pallavi R. Kapse
Department of Electronics and Telecommunication
Solapur Univercity
N.B.N. Sinhgad College of Engineering, and Kegoan
India

ABSTRACT

With increasing amount of 3D content there is huge amount of growth in past many years, the availability of 3D content is still below the normal size by that of its 2D corresponding closely to one another. To remove this distance many 2D to 3D image conversion techniques has been proposed. It is important to estimate relative depth map in a single view image for 2D to 3D image conversion techniques. Semiautomatic image conversion method which makes use of human operators has been most successful but it takes many times to proceed and costly too. Hence automatic image conversion methods come in result which reduces complexity. But have not still achieved same level of quality. The stereoscopic image provides information on details of each object in the picture in the three dimensions and helps us to feel the picture realistic one. The main step of 2D to 3D image conversion process consists of depth estimation for given 2D image. So this paper is focused is depth recovery. In this paper proposed 2D to 3D image conversion method is based on globally estimating the entire depth map of query image directly from the repository of 3D images which contains images and their depth pairs using nearest neighbour regression type idea.

Keywords:- 3D images, stereoscopic images, conversion of 2D to 3D image, nearest neighbour classification, bilateral filtering.

I. INTRODUCTION

Three-dimensional pictures contain yet another dimension: depth. This type is the most realistic one, as the depiction of objects or environments resembles the way we see them through our own eyes. Painters use the technique of perspective, drawing distant objects smaller and depicting angles as visible through one's point of view, while 3-D movies use two images superimposed on the same screen. However, such pictures give only the illusion of depth, as the canvas or screen always remains flat. The today's life 3D system are used such as 3D TV, smart phones, 3D movies and gaming zones are not yet achieved 3D content production. The most important and difficult problem in 2D-to-3D conversion is how to generate or estimate the depth information using only a single-view image. Since there is no 3D information, we should estimate relative depth differences for each region in a single-view image. Several methods have been proposed to estimate the depth information from a single-view image. In [3], Feldman *et al*. generate the depth map which is made by manual method using some adjusting tools offered in Photoshop. As an automatic estimation method, S. Battiato *et al.* utilize the colour information to estimate the relative depth map in a 2D image [4].

A standard 2D-to-3D image conversion process consists of two steps: depth estimation for a given 2D image and second one is depth-based rendering of a new image in order to form a stereo pair. There are two basic approaches to 2D-to-3D conversion: one that requires a human operator's help and one that does not. The semi-automatic methods have been proposed where a skilled operator assigns depth to various parts of an image or video. Based on this sparse depth assignment, a computer algorithm estimates dense depth over the entire image or video sequence. [12] The involvement of a human operator may vary from just a few scribbles to assign depth to various locations in an image to a precise delineation of objects and subsequent depth assignment to the delineated regions. In the other case of automatic methods, no operator intervention is needed and a computer algorithm automatically estimates the depth for a single image (or video). To this effect, methods have been developed that estimate shape from shading, structure from motion or depth from defocus. Although such methods have been shown to work in some restricted scenarios they do not work well for arbitrary scenes. In an attempt to equip 3D TVs, Blue-Ray players and gaming consoles with real-time automatic 2D-to-3D conversion, consumer electronics manufacturers have developed simpler techniques that rely on various heuristic assumptions but such methods fail on more challenging scenes. Recently, machine-learning-inspired methods have been proposed to automatically estimate the depth map of a single monocular image by applying image parsing. Although restricted to architectural scenes, these methods opened a new direction for 2D-to-3D conversion.

The developed method is based on globally estimating the entire depth map of a query image directly from a repository of 3D images (image and depth pairs or stereo pairs) using a nearest-neighbour regression type idea. Early versions of our learning-based approach to 2D-to-3D image conversion either suffered from high computational complexity [6] or were tested on only a single dataset [7]. Here, we introduce method evaluate the qualitative performance and the computational efficiency of local against those of the Make3D algorithm [8] and a recent method proposed by Karsch *et al.* [5].

This paper is organized as follows. The section II describes the related work. In section III we describe the conversion method based globally estimating depth map and in section IV we show numerous experimental results and we conclude the paper in Section V.

II. RELATED WORK

Lots of work has been done regarding 2D to 3D image conversion. A one of the novel 2D to 3D conversion method is based on edge information. Edge of the depth map has a high probability. First pixels are grouped together then a relative depth value can be assigned to each region. [1] Initially the block based image is segmented into multiple groups. The depth of each segment is then assigned by using initial depth hypothesis. Next, the blocky aircraft is removed using cross bilateral filtering. Finally multi view images are rendered by depth image-based rendering and display on a display 3D.

The algorithm proposed is quality-scalable depending on the block size. Smaller block seize will result in better depth details and large block size will have lower computational complexity. [2] Automatic monoscopic video to stereoscopic 3Dconversion scheme is presented using block-based motion depth from motion estimation and colour segmentation for depth map enhancement. The colour based segmentation provides good boundary region. This paper presents a robust 2D to 3D stereoscopic video conversion system for off-line automatic conversion application.

III. PROPOSED METHOD

The proposed method is based on globally estimating depth map of query image directly from repository of the 3D image which contain image and their respective depth pairs also called stereo pairs using nearest neighbor search. The method proposed here is based on key observation and assumption. Observation is in case that there are lots of stereo images available on-line or internet whose contents is match to our 2D query image which we wish to convert to 3D image. Assumption made that two images that are photo metrically similar also have similar 3D depth map.

The block diagram of algorithm is shown in fig. 1. The 3D dataset can be constructed by many methods. Here dataset is is made up of the make 3D dataset [9] composed of many outdoor images with depth fields captured by a laser range finder and NYU Kinect dataset [10] contain may indoor RGB image and corresponding depth fields.

Let I= {(i1, d1), (i2, d2), (i3, d3)….. (ik, dk)} Indicates training dataset composed of k pairs (ik, dk), where ik is a color image and dk is its respective depth image. Here Q is monocular query left image that we will convert to stereo image. Qr is the right image which is being sought for each query image Q and dQ is query depth image. The aim is to find a depth estimate and then a right image estimate Qr form given 2D query image Q and 3D dataset I. Following steps provides description of each step of proposed algorithm.

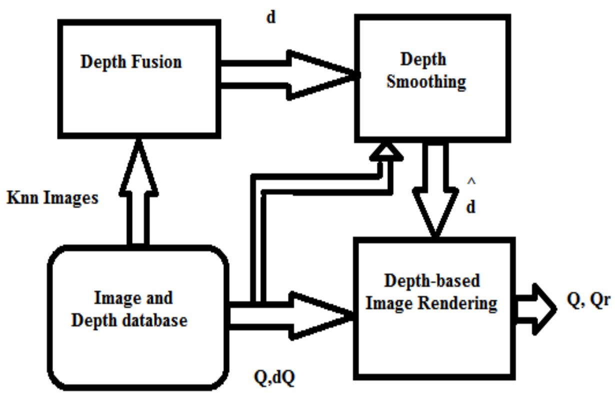

Fig. 1 Block diagram of algorithm

A. *Knn Search*

The dataset contains image and their depth pair's two types of images: those that are relevant to estimated depth map of 2D query image and those that are irrelevant. So images that are not photo metrically similar to the 2D query image need to be rejected because they are not useful for estimating depth map. It is very important to select only k nearest neighbor depth relevant images form repository, and this accomplished using distance function called Euclidean distance norm. This distance function measures that difference between histogram computed form two images. It will search for top k images form dataset closely match to query images and this search returns with images and their depth pairs that are photo metrically similar to 2D query images.

B. Depth Fusion

The knn images and depth pairs not completely match to 2D query image. Like location of the object and part of the object may consistent with those in respective query image. If similar object present at similar location in many knn search images then it is same as that such object also appear in query image and depth images too. Finally depth image is estimated from knn search image and depth pairs by applying median operator at each spatial location x as follows:

$$d[\mathbf{x}] = \text{median}\{di\ [\mathbf{x}]\ \forall i \in K\}.$$

These depth fields are overly smooth, globally consistent estimated from several knn depth images. Examples of such depth fused images are shown in fig.2.

C. Depth Fusion

The fused depth field is overly smooth and locally inconsistent with query image due to edge misalignments between depth fields of knn and the query image. There is lack of edges of fused depth fields and lack fused depth smoothness occur in depth fused image. To correct this we apply cross bilateral filtering (CBF). CBF is a variant of bilateral filtering, an edge-preserving image smoothing method that applies anisotropic diffusion controlled by the local content of the image itself [4]. In CBF the diffusion is not controlled by the local content of the image under smoothing but by an external input. We apply CBF to the fused depth d using the query image Q to control diffusion. This allows us to achieve two goals simultaneously: alignment of the depth edges with those of the luminance Y in the query image Q and local noise/granularity suppression in the fused depth d.

D. Depth based image rendering

The estimation of the right image Qr form from monocular query image is generated by calculating disparity depth δ from the estimated depth d. We assume that stereo pairs Q and Qr was captured by parallel cameras with baseline B and focal length f. The maximum value of depth sensor is 255 and disparity δ is calculated as follows:

$$\delta = B^*f/\text{depth (d)}$$

This depth rendering generate the right image of the stereo pair using the 2D query (left) image and Smoothed median depth field followed by suitable processing of occlusions and newly-exposed areas.

IV. EXPERIMENTAL RESULT

The proposed method is tested on dataset the Make3D dataset [21] composed of many outdoor images with depth fields captured by a laser range finder and the NYU Kinect dataset [22] composed of many pairs of RGB images and corresponding depth fields.

Fig.2 Knn images and their respective depth pairs

Above figure 2 shows 3 nearest neighbour images and their depth pairs searched form 3D repository that are closely matched to given 2D query image. This can be done with help of Euclidean distance function. Next step is to fuse all knn depth images with help of median operator. In fig. 3 shows 2D query image which we wish to convert to 3D image and middle one shows fused depth image. Since such estimated depth image contain much misalignment at edges and it is not smoothed. Also fused depth field is overlay inconsistent. So remove these errors we can apply cross bilateral filtering.

CBF is strong edge preserving smoothed corss bilateral filtering. The filtered depth preserves the global properties captured by the unfiltered depth field *d*, and is smooth within objects and in the background. At the same time it keeps edges sharp and aligned with the query image structure. In fig. 3 right image shows result of smoothed depth image of given 2D query image.

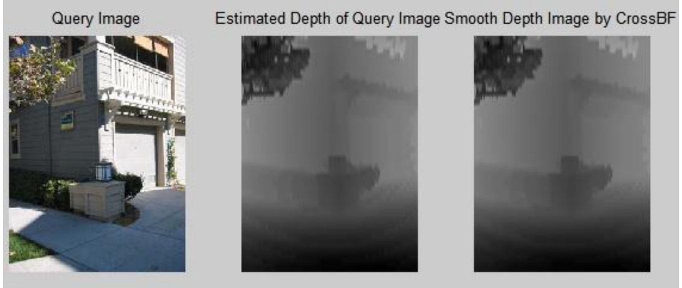

Fig. 3 Estimated Fused depth image and smoothed depth image.

Fig. 4 shows Original image and rendered image.

Above fig.4 shows 2D query image in original form and image obtained after depth based image rendering. In rendering step is well understood and formula exist that gives us rendered image. Baseline and focal length need to estimate the right image of fictitious stereo pairs. Newly exposed areas calculated with disparity value and get rendered image. Finally anaglyph image of 2D query image is generated as shown in figure 5. An anaglyph is an image made up of two colour layers, typically red and cyan that observed by the use of a particular type of glasses, shows depth properties of the image giving the idea of 3D.

Fig. 5 Anaglyph image of 2D query image.

The picture contains two differently filtered coloured images, one for each eye. When viewed through the colour coded "anaglyph glasses", they reveal an integrated stereoscopic image. The visual cortex of the brain fuses this into perception of a 3D scene or composition. To see the better result of above image can use red cyan anaglyph glasses.

V. CONCLUSION

This work has presented a novel 2D to 3D automatic image conversion. The proposed algorithm based on globally estimating entire depth map of query image directly from the repository 3D image and depth pairs or dataset using nearest neighbour search. This knn search provides top k matches of image and depth pairs which closely match to 2D query image. These knn depth images get fused with help of median operator and get smoothed with cross bilateral filtering. Finally very well known technique applied on estimated smoothed depth image and 3D image get generated. This will result in fast and accurate algorithm by comparing with other 3D algorithm. But still there is room for improvement as such 3D image may contain void aid of distortion.

ACKNOWLEGMENT

The author would like to acknowledge Prof. S. S. Hippargi, department of electronics and telecommunication, N. B. Navale Sinhgad College of engineering, Kegaon, Solapur University, for their valuable support and discussion regarding this project.

REFERENCES

[1] Chao-Chung Cheng, Chung-Te Li, and Liang-Gee Chen, "A novel 2D-to-3D conversion system using Edge information," 2010.

[2] Lai-Man Po, Xuyuan Xu, Yuesheng Zhu, Shihang Zhang, Kwok-Wai Cheung and Chi-Wang Ting," Automatic 2D-to-3D video conversion technique based on depth-from-motion and color segmentation," 2013.

[3] L. Angot, W.-J. Huang, and K.-C. Liu, "A 2D to 3D video and image conversion technique based on a bilateral filter," Proc. SPIE, vol. 7526, p. 75260D, Feb. 2010.

[4] T. Brox, A. Bruhn, N. Papenberg, and J. Weickert, "High accuracy optical flow estimation based on a theory for warping," *in Proc. Eur.* Conf. Comput. Vis., 2004, pp. 25–36.

[5] 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

[6] J. Konrad, G. Brown, M. Wang, P. Ishwar, C. Wu, and D. Mukherjee,"Automatic 2D-to-3D image conversion using 3D examples from the Internet," Proc. SPIE, vol. 8288, p. 82880F, Jan. 2012.

[7] 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.

[8] A. Saxena, M. Sun, and A. Ng, "Make3D: Learning 3D scene structure from a single still image," IEEE Trans. Pattern Anal. Mach. Intell., vol. 31, no. 5, pp. 824–840, May 2009.

[9] *(2012). Make3D [Online]. Available*: http://make3d.cs.cornell.edu/data.html

[10] F. Durand and J. Dorsey, "Fast bilateral filtering for the display of high-dynamic-range images," ACM Trans. Graph., vol. 21, pp. 257–266, Jul. 2002.

[11] Janusz Konrad, Meng Wang, Prakash Ishwar, Chen Wu, and Debargha Mukherjee, " Learning-Based Automatic 2D-to-3D Image and Video conversion," vol. 22, no. 9, September 2013.