2D to 3D Conversion for Outdoor Scenes

Chien Wu, Chung-Te Li, Cheng-Yuan Ko, Chen-Han Chung and Liang-Gee Chen

DSP/IC LAB, Graduate Institute of Electronic Engineering, National Taiwan University
{mpemial, ztdjbdy, kevin, penal, lgchen}@video.ee.ntu.edu.tw

Abstract

This paper provides 2D to 3D conversion system based on dark channel prior. This method has low computation and high performance on outdoor scene. However traditional dark channel results suffer from noise that generated from wrong assumption to monocular cues. We use scene estimation and cubic-spline interpolation to reduce these artifacts from finding the global depth map. By applying HHT object-based filtering, the algorithm can smooth the depth map and increase quality of viewing experience.

1. Introduction

As 3D technology matured, there are much more needs in 3D contents than before. People can use stereo matching to get binocular cues and rendered the corresponding 3D video [1]. However there are still lots of video sequences that lack of stereo view and cannot applied stereo matching to get the depth information. Therefore some research focus on how to estimate depth information from monocular cues. In [2] they proposed a conversion system based on relative size, which is a strong depth cue during small viewing distance. While in [3], Cheng et al. proposed a 2D to 3D conversion system considering edge information to group the image into regions. [4] presented an algorithm to refine depth map with dark channel prior and color saturation.

Although 2D to 3D conversion is an ill-posed problem in signal processing, we can still using some monocular cues for depth map estimation. The proposed method uses dark channel prior as overall estimation. However lack of boundary message and scene information, the quality of dark channel based depth map is poor to watch. We then apply scene estimation and HHT filter to refine the depth map and generate comfortable 3D content.

2. Proposed Method

The system first uses dark channel prior for coarse depth estimation. Scene estimation is then applied in order to refine the depth map. We provided object-based depth filtering to refine the edge of depth map. In Fig. 1, we show the system diagram of our proposed method.

2.1. Dark Channel Prior

We first adopt dark channel prior to estimate depth information from perspective monocular cue. The main purpose of applying dark channel prior is to estimate the atmosphere light, which is an important cue in outdoor scene. In [5] it states that in the most of the region that do not cover the sky, there is at least on color channel that have some really low pixel intensity. And in these images, these low intensity pixel contribute to the air light.

$$J^{dark}(x) = \min_{c\in\{r,g,b\}}\left(\min_{y\in\Omega(x)}(J^c(y))\right), \quad (1)$$

where J^c is a color channel of **J** and $\Omega(\mathbf{x})$ is a local patch centered at x. J^{dark} the *dark channel* of **J.** The patch size here is very critical to the depth map quality. If we use improper patch size, the result may be too coarse to synthesis or may have many high frequency noises which can contribute to uncomfortable viewing experience.

2.2. Scene Estimation

Traditional dark channel prior is very sensitive to noise signal, therefore may cause cue confliction on some region. We use scene estimation to analyze global depth intensity and can remove so local noise caused by dark channel. According to dark channel prior, we choose the highest value and lowest value for scene

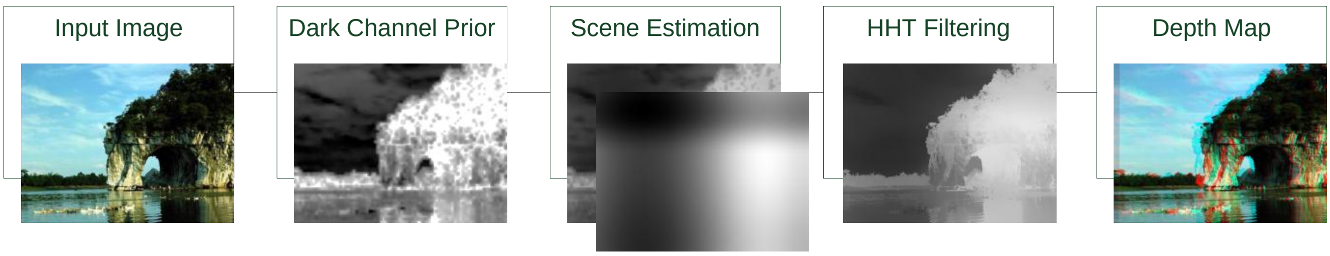


Figure 1. Proposed 2D to 3D conversion system

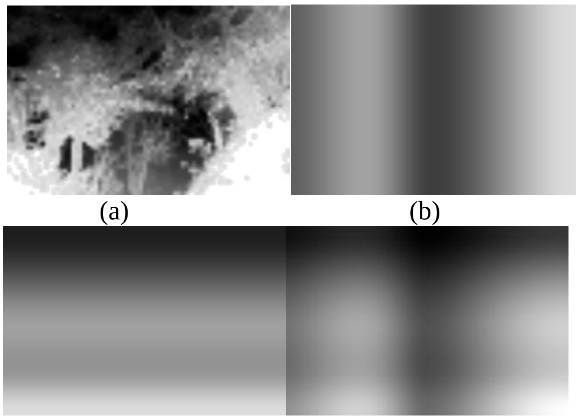
TABLE I
THE SCENE ESTIMATION PROCESS

Iterative process a) to g):

a) select local maxima or minima a
b) select local maxima or minima b
c) assign four vertex intensity from 0~255.
d) use cubic_spline interpolation to synthesis global depth map from four vertex point, boarder on the left(top), point a, point b, boarder on the right(bottom)
e) compare original depth and interpolate depth map, output the total difference D_i.
f) *if* $D_i < D_{min}$ update the new parameter and local maxima, minima point a, b.
g) repeat a) to f) to find the difference minima

estimation. And apply cubic-spline interpolation to generate depth plane. The pseudo code is shown as in Table I.

The proposed method repeat Table I. for finding both x-direction and y-direction minima, and then combined two result to synthesis the final scene depth. Fig. 2(a) is the result of depth prior, and Fig 2(b) is the x-direction assumption. Fig. 2(c) shows the y-direction estimation and in Fig. 2(d) is the combined image for overall scene estimation.



(a) (b) (c) (d)

Figure 2. The results of scene estimation.

2.3. Object-Based Depth Filtering

By applying Hilbert Huang decomposition, we can refine the object boundary which is very essential while watching 3D [6]. Hilbert Huang transform (HHT) [7] can decompose non-stationary and nonlinear stochastic signals such as tides or electrocardiography into several intrinsic mode functions (IMF). The algorithm first separates the edge of the textures and edge of the objects. And then depth map is refined by the regrouping pixels. (2) and (3) shows that image *I(x,y)* is separated into row signals and column signals.

$$Row_{I,k}(x) = I(x,k), \quad (2)$$

$$Col_{I,l}(y) = I(l,y). \quad (3)$$

The generic function h(x) is the sum of different IMF.

$$h(x) = \sum_{j=0}^{\infty} IMF(j, h(x)). \quad (4)$$

By using (4), we can set a threshold p to separate row signals that represent the textures (5) and objects (6).

$$Tex(Row_{I,k}(x)) = \sum_{j=p}^{\infty} IMF(j, Row_{I,k}(x)), \quad (5)$$

$$Obj(Row_{I,k}(x)) = \sum_{j=0}^{p-1} IMF(j, Row_{I,k}(x)), \quad (6)$$

Also for column signals, we use the same way to get textures part (7) and objects part (8).

$$Tex(Col_{I,l}(y)) = \sum_{j=p}^{\infty} IMF(j, Col_{I,l}(y)), \quad (7)$$

$$Obj(Col_{I,l}(y)) = \sum_{j=0}^{p-1} IMF(j, Col_{I,l}(y)). \quad (8)$$

The final energy map *E(x,y)* is defined by (9), the higher the energy map the higher the possibility that is belong to boundary region. We can therefore refined the depth map according to these information, since human are very sensitive to boundary in stereo vision

$$En(x,y) = \max\{\frac{\partial Obj(Row_{I,y}(x))}{\partial x}, \frac{\partial Obj(Col_{I,x}(y))}{\partial y}\}. \quad (9)$$

2.4. Depth Image-Based Rendering

In order to synthesis red-cyan image, we adopt a 3D voxel image completion method [8]. The layer based image depth rendering can provide better visual quality and less computation overhead.

3. Experiment Result

Fig 3. shows the result of our proposed method. The first row is original color image. The second row is the depth prior assumption. The third row is our scene estimation with both x-direction and y-direction. The object-based depth filtering is shown in the fourth row, and the results of DIBR are shown in the final row.

4. Conclusions

We proposed a 2D to 3D conversion system which use aerial perspective cue to estimate the depth information. At first stage dark channel prior is applied to acquire coarse depth map. Then we use scene estimation to generate the approximate depth plane. Hilbert Huang decomposition is used to refine the depth map, we can divide the frequency into two parts including object the textures. The overall system can provide better viewing experience than only use dark channel to guess monocular cues. However our method is mainly affected by the original assumption, which is the image have strong aerial perspective cue. Therefore it is restricted to only outdoor scenes.

5. Acknowledgement

This work is partially supported in part of Himax Technologies, INC, Taiwan, R.O.C..

6. References

[1] Yen-Jei Lai, Chung-Te Li, Chien Wu, Chao-Chung Cheng and Liang-Gee Chen, "Trilateral-Filter-Based Depth Interpolation for Occlusion Handling in Stereo Vision", in International Conference on 3D systems and Applications (3DSA 2011),Seoul, Korea, June 2011.

[2] Yong Ju Jung, Aron Baik, Jiwon Kim, and Dusik Park, "A novel 2D-to-3D conversion technique based on relative height depth cue" in SPIE Electronics Imaging, Stereoscopic Displays and Applications XX, 2009

[3] Chao-Chung Cheng and Chung-Te Li and Po-Sen Huang and Tsung-Kai Lin and Yi-Min Tsai and Liang-Gee Chen, "A Block-based 2D-to-3D Conversion System with Bilateral Filter", in International Conference on Consumer Electronics (ICCE) , 2009, Las Vegas, NV, USA, published by IEEE.

[4] Tien-Ying Kuo, Yi-Chung Lo, Yun-Yang Lai "Depth Estimation from a Monocular Outdoor Image," ICCE, Jan. 2011.

[5] He, K., Sun, J., and Tang, X., "Single image haze removal using dark channel prior," CVPR , 1956–1963 (2009).

[6] Chung-Te Li, Yen-Chieh Lai, Chien Wu, Sung-Fang Tsai and Liang-Gee Chen, "3D Image Correction by Hilbert Huang Decomposition," ICCE, Jan. 2012.

[7] N. E. Huang, Z. Shen, S. R. Long, M. Wu, H. H. Shih, Q. Zheng, N. C. Yen, C. C. Tung, and H. H. Liu, "The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis," Proc. of The Royal Scoiety, vol. 454, no. 1971, pp. 903–995, 8 Mar. 1998.

[8] Chung-Te Li and Pei-Kuei Tsung and Kuan-Yu Chen and Chao-Chung Cheng and Liang-Gee Chen, "Three-Dimensional Image Completion For Depth Image Based Rendering", in 3DSA, 2010, Tokyo, Japan.

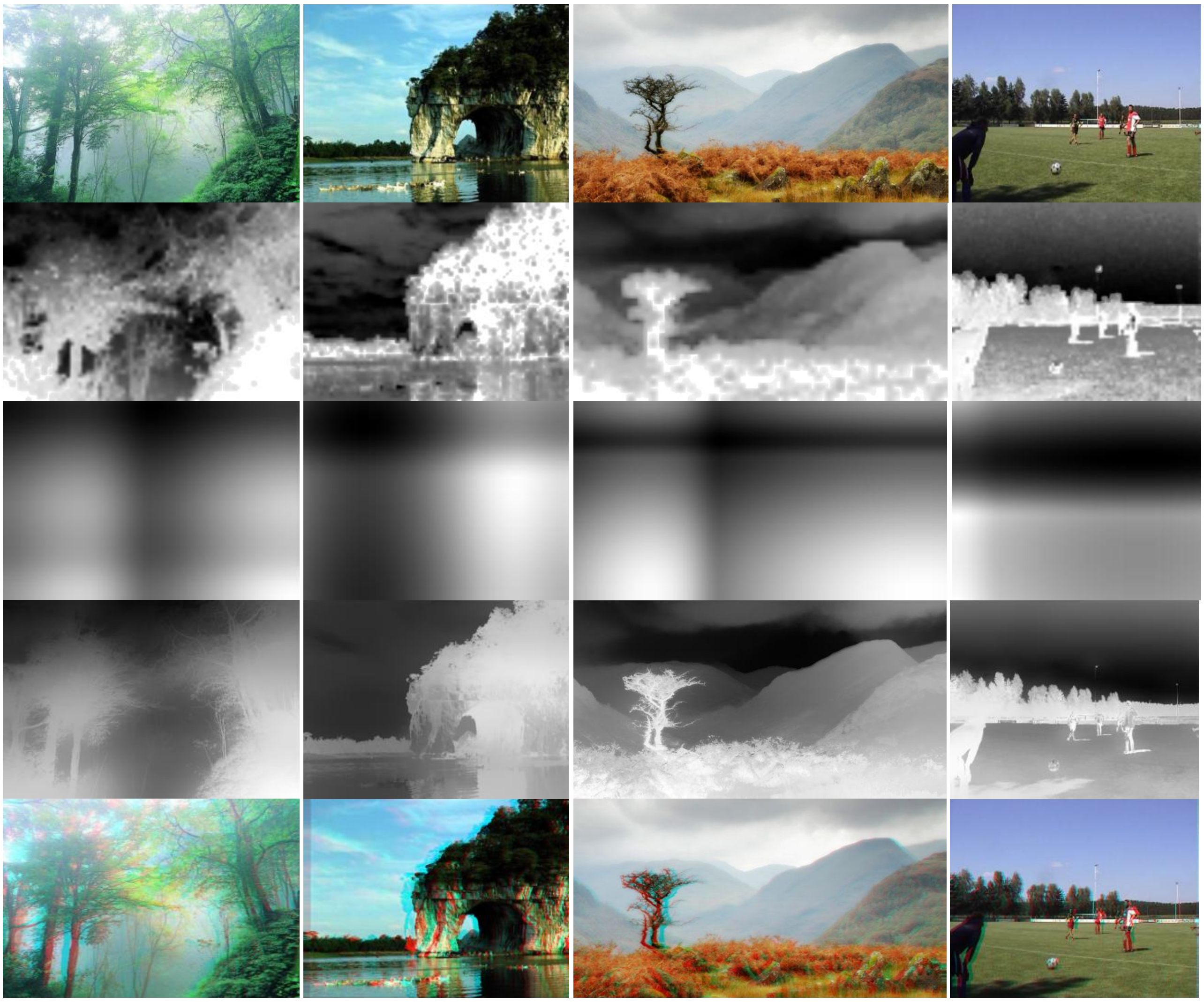


Figure 3. Experimental result of our proposed system.