2D TO 3D CONVERTION BASED ON EDGE DEFOCUS AND SEGMENTATION

Ge Guo[1,2,3], Nan Zhang[2], Longshe Huo[2], Wen Gao[1,2]

[1]Institute of Computing Technology, Chinese Academy of Sciences, Beijing 100080, China
[2]Institute of Digital Media, Peking University, Beijing 100871, China
[3]Graduate School, Chinese Academy of Sciences, Beijing 100039, China
gguo@jdl.ac.cn; {nzhang, lshuo, wgao}@pku.edu.cn

ABSTRACT

This paper presents a depth estimation method which converts two-dimensional images into three-dimensional data. Based on two-dimensional wavelet analysis of Lipschitz regularity for defocus estimation on edges, this method can effectively eliminate the horizontal stripes in the depth map resulted from traditional one-dimensional wavelet based approaches. Besides, we also propose several techniques such as edge enhancement, color-based segmentation, and depth optimization to obtain a more reliable and smoother depth map. The experimental results demonstrate the effectiveness of our proposed techniques.

Index Terms— 2D to 3D conversion, depth map, wavelet, Lipschitz exponent, color segmentation

1. INTRODUCTION

Recently an advanced 3-DTV system has been put forward based on the new technology called Depth Image-Based Rendering (DIBR) [1]. In this system, a new 3-D data representation is adopted, which includes the traditional 2-D images and their associated per-pixel depth maps. The depth maps can be used to describe the 3-D location of each point in the images. This representation is generally considered to be more efficient for coding, storage, transmission and rendering than traditional 3-D video representations.

One of the key problems rest in the above system is how to recover depth information from 2-D data, which is also a difficult problem in computer vision. Although there have been depth cameras which can directly obtain depth values along with the colorful 2-D images, the range of distance and other lighting conditions are quite restricted. The feasible methods for now are still the classical computer vision methods such as stereo or monocular vision. Depth from stereo match is the common method in this research area, which exploits the disparities between two slightly different images from the left and the right view [2]. Besides disparity as one kind of depth cues, there are also monocular cues such as gradients, defocus, occlusion, haze, etc. Estimating depth from monocular cues is a challenging and interesting problem, and particularly convenient for generating the DIBR based 3-D data representation.

In previous works of monocular cues based depth recovering, Harman et al. [3] used machine learning algorithms to estimate depths. Battiato et al. [4] proposed a classification method to label images as indoor, outdoor with geometric elements, or outdoor without geometric elements. For the first two classes of images, the depths were designated based on the detection of vanishing lines. In [5] a supervised learning approach and a discriminatively-trained Markov Random Field model were used. Researches on depth from defocus appeared in many literatures [6][7], which extracted the actual distance of a point in the scene by measuring the amount of blurring at the corresponding point in the image.

A method to obtain a relative depth map from a single image using wavelet analysis and edge defocus estimation based on Lipschitz exponents was proposed in [8]. Images were handled as series of 1-D row signals, with the resulting horizontal stripes in the depth map, as shown in Fig. 1. To address this issue, in this paper we present an incremental algorithm based on wavelet transform and edge focus analysis in two-dimensions, taking into account the direction of edges and the two-dimensional characteristics of images. The depth map is further optimized and smoothed based on color segmentation to obtain much more accurate and reliable results.

2. BACKGROUND AND RELATED WORK

2.1. Wavelet analysis

For images of limited depth of field (DOF), objects in the image may not be all in focus [6]. Usually the objects in the background are blurred and the textures are smoothed; whereas the main foreground objects are focused with sharp edges and textures in detail. In other words, the high frequencies are retained in the focused foreground, but greatly attenuated in the background. This suggests that the local spatial frequency is directly related with the degree of blurring, and thus the relative distance of the object from the camera.

(a) (b)

Fig. 1. Example of the results generated by [8]. (a) Original image; (b) Final depth map.

The high frequencies are described by the coefficients of the wavelet transform of the image. If there is larger energy in the wavelet bands of high frequency, it suggests that there are more details and less blurring in this region, where the 3-D location is nearer. The elementary relative depth can be estimated based on the values of wavelet coefficients in the high frequency bands.

Based on this, Valencia et al. [8] divided images into macro blocks whose size was 16-pixel by 16-pixel. A macro block wavelet transform which generated 256 wavelet coefficients was performed. Relative depth was estimated by counting the number of non-zero wavelet coefficients.

2.2. Edge defocus analysis

According to [9], the local regularity of signals can be characterized by the Lipschitz exponents, which can be considered as a measurement of how many times the signal is differentiable at a point. More "regular" intensity variation means the edge is more defocused.

In [8], an image was considered as a series of 1-D row signals. The Lipschitz regularity was computed for each row by measuring the decay of the wavelet transform from the coarser scale up to the finer scale, as the method demonstrated by Mallet et al. [9]. Each row was divided into several sections by edge points, and their depth was rectified according to the edge Lipschitz exponents. Finally the block-level depth map was converted to a final pixel-level depth map.

However, it is observed from the experimental results of [8] that there were many horizontal stripes appeared in the generated depth map. Besides, some of the object contours were also damaged. In order to overcome these problems, in the following section we present an incremental algorithm of depth estimation based on edge focus analysis in two dimensions and color-based segmentation.

3. THE PROPOSED APPROACH

The outline of our depth estimation algorithm is similar to that of [8]. However, some new techniques are reinforced. The main steps of our algorithm can be concluded as follows:

1) Initial pixel-level depth map creation based on local high frequency analysis. Since wavelet transform is performed on the local window of each point in the image instead of the block-divided method used in [8], blocky effects in the resulting initial depth map are avoided effectively.

2) Edge focus analysis based on the Lipschitz exponents in 2-D wavelet. Compared with the 1-D wavelet analysis used in [8], our method can retain the edge contour with a non-striped depth map.

3) Edge points connection. Edge enhancement helps to form complete edges and lessen the errors caused by edge discontinuities when refining the depth on the basis of Lipschitz exponents.

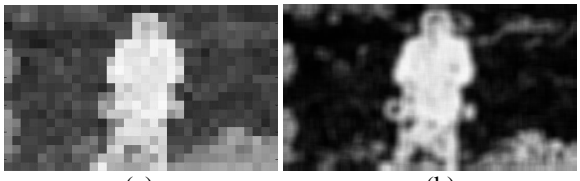
4) Depth map refining according to the Lipschitz exponents on edges, similar to the basic idea of [8] but in 2-D wavelet.

5) Color-based segmentation and depth optimization in each homogeneous color segment. This helps to optimize the depth of foreground regions with low frequency energy and smooth the depth map in each homogeneous color region.

3.1. Initial depth map creation

As discussed above, the depth of limited-DOF images can be measured by their local frequencies. In this step, we analyze the frequency energy of local regions based on the wave transforms in local block windows. For each point, its local window is created with the point as its center. The size of each window is N x N (N=16 in our experiments). The number of the nonzero coefficients in the high frequency wavelet bands (the LH, HL, and HH bands) shows how much the details are not blurred, and therefore gives a relative depth value. The range of depth is adjusted from 0 to 255 (0 denotes black and 255 denotes white in the depth map). More nonzero coefficients correspond to larger depth value, which indicates nearer in distance.

Fig. 2 compares the initial depth maps generated by the block-divided wavelet method of [8] and our pixel-level method. It can be seen that, compared with our method the depth map generated by [8] is blocky, with too many details lost.



(a) (b)

Fig. 2. Initial depth map comparison. (a) Result of [8]; (b) Result of the proposed method.

3.2. Edge defocus analysis based on 2-D wavelet

Mallet et al. [9] has demonstrated that the multi-scale wavelet modulus maxima detect all singularities. For images,

the 2D Gaussian function is used as the smoothing function. The wavelet transform has two components, which are proportional to the gradient vector of the corresponding smoothed images. Both the wavelet modulus and the angle are computed at each point (see also equation (45), (50), and (52) in [9]). Here each angle includes the information of the edge orientation and the direction along which the wavelet modulus maxima propagate across scales. In our algorithm, the angles are approximately classified into 8 orientations with an interval of $\pi/4$, as shown in Fig. 3.

We detect the wavelet modulus maxima along the angle direction across scales. These multi-scale maxima of a singular point form the modulus maxima curve. Among the points of modulus maxima, those with modulus smaller than a given threshold are discarded since they are possible noises. The Lipschitz exponent is computed from the slope of the wavelet modulus maxima curve in the logarithmic domain. If there are two maxima curves of a singular point, the one with smaller estimated slope is chosen.

The results of the Lipschitz value indicate the blurring degree of edges. Smaller value indicates less defocus. According to the theories mentioned above, Lipschitz $\alpha = 0$ indicates the step edge (normally focused edge). Therefore the edge points with Lipschitz exponent $-1 < \alpha < 0$ are considered to be focused, and those with $0 < \alpha < 1$ are defocused.

3.3. Edge enhancement

Because of the interaction of the wavelet modulus maxima during their propagation along different directions across scales, and the limitation of the image resolution, some of the edge points may not be detected. Edge discontinuities are yet caused by the computation errors of the Lipschitz exponents, as well as the non-selection of modulus maxima whose values are smaller than the threshold in the previous step. Therefore in order to create a precise pixel-level depth map, edge connection is necessary.

For each edge point that has been detected, we search for the possible edge points in its 8 neighborhoods. If the candidate point is a modulus maximum point and its modulus is close to that of the current edge point, then it is considered as an edge point. The Lipschitz exponent of the new edge point is estimated by its neighbor edge points which have already been computed. Therefore the edges, especially for the foreground objects, can be well connected. Fig. 4 shows the example of edge points before and after the operation of this step.

3.4. Depth map refining

With the complete edges and the Lipschitz exponents at all the edge points, each of the interior point can be classified into one of the following three categories according to its edges:

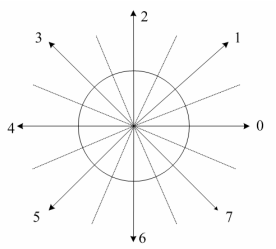
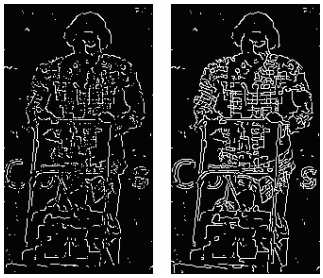


Fig. 3. Angle directions used in 2-D edge defocus analysis



(a) (b)

Fig. 4. Edges before and after connection. (a) Initial edges before connection; (b) The connected edges.

- If most of the Lipschitz values of the edge points satisfy $-1 < \alpha < 0$, the current interior point is marked as foreground point, and its depth is the maximum of the initial depths in the current region.
- If most of the Lipschitz values of the edge points satisfy $0 < \alpha < 1$ or even larger than 1 (the boundary of the image is set to any value more than 1), the current interior point is background whose depth is the minimum depth of the region.
- Points not satisfying the above two has the approximately equal numbers of edge Lipschitz in $-1 < \alpha < 0$ and $0 < \alpha < 1$, or none of the two has absolute advantage. The depths of these points are set to the initial depth value that appears most in the region.

After all points are processed with the above rules, the depth map is refined, as shown in Fig. 5.

3.5. Depth map optimizing

In limited-DOF images, usually the focused foreground with various textures has high energy and the number of nonzero wavelet coefficients is large. In this case the depth estimation has approximately accordant values. However in the interior regions of the focused foreground objects with uniform color and few textures, the local frequency may be very low. As a result, the estimated relative depth value is smaller than the truth. To rectify this error, this step tries to compensate the depth of such foreground regions.

The Mean Shift algorithm is used for color-based segmentation [10]. The image is divided into segments with homogeneous colors. Then we check in each segment the

number of the foreground pixels marked in the previous step. If the proportion of the foreground pixels in a segment is larger than a given threshold, the segment is considered as a foreground segment. If the average depth value of a foreground segment is smaller than that of the background, we rectify its depth D according to the mean value $\overline{\alpha}$ of the edge Lipschitz exponents. Let $\alpha_{\min}$, $\alpha_{\max}$ be the minimum and maximum Lipschitz of all the foreground respectively, $D_{\min}$, $D_{\max}$ be the minimum and maximum depth value of all the foreground, then

$$D = D_{\max} - \frac{D_{\max} - D_{\min}}{\alpha_{\max} - \alpha_{\min}}(\alpha_{\max} - \overline{\alpha})$$

Following the common premise that each homogeneous color region has approximately homogeneous depth value, it is necessary to smooth the depth in each homogeneous color segment. In our algorithm, the mean filter is selected to smooth the depth map. An example of the final result is shown in Fig. 6.

4. EXPERIMENTAL RESULTS

In our experiments, we first compare our proposed method with the one of [8] using the color image provided by [8] as shown in Fig. 1(a). The final results of these two methods are shown in Fig. 1(b) and Fig. 6 respectively. From the figures it can be observed that our result avoids the horizontal stripes and well retains the object edges. The foreground details such as the trees are also recognized with their relative depths. However, errors still happen in some background regions since they are surrounded by the focused edges. Segmentation parameters influence the final depth yet. Segments which coincide with the edges detected by the Lipschitz singularity have good results.

Fig. 7 demonstrates other images and their depth maps generated by the proposed algorithm of this paper. The left column shows the original 2-D images and the right column shows their corresponding depth maps. The results show that our method is robust and reliable.

5. CONCLUSIONS

This paper presents a depth estimation method from 2D images, which is quite enough for 3D rendering and stereo applications. Based on the edge defocus analysis and color-based segmentation, we can get relatively reliable and smooth depth maps. Future work is to add human interactions to depth generations.

ACKNOWLEDGMENT

This work was performed in Peking University, and was partially supported by the "863" Program of China under grant No.2007AA01Z315, the "985" construction project of Peking University, the China Postdoctoral Science Foundation, and NEC Research China.

REFERENCES

[1] C. Fehn, R.D.L. Barre, and S. Pastoor, "Interactive 3-DTV – Concepts and Key Technologies," *Processing of IEEE*, vol. 94, no. 3, March 2006.
[2] D. Scharstein and R. Szeliski, "A Taxonomy and Evaluation of Dense Two-Frame Stereo Correspondence Algorithms," *Int'l J. Computer Vision*, vol. 47, pp. 7-42, Apr. 2002.
[3] P. Harman, J. Flack, S. Fox, and M. Dowley, "Rapid 2D to 3D Conversion," *Proceedings of SPIE*, Vol. 4660, pp. 78-86, 2002.
[4] S. Battiato, S. Curti, M. La Cascia, M. Tortora, and E. Scordato, "Depth Map Generation by Image Classification," *Proceedings of SPIE*, vol. 5302, pp. 95-104, April 2004.
[5] A. Saxena, S. H. Chung, and A. Y. Ng, "Learning Depth from Single Monocular Images," *NIPS 18*, 2005.
[6] A. P. Pentland, "A New Sense for Depth of Field," *IEEE Trans. Patt. Anal. Machine Intelligence*, vol. 9, pp. 523-531, 1987.
[7] J. Ens, P. Lawrence, "An Investigation of Methods for Determining Depth from Focus," *IEEE Trans. on Pattern Analysis and Machine Intelligence*,vol. 15, no. 2, pp. 97-108, 1993.
[8] S. A. Valencia, R. M. Rodríguez-Dagnino, "Synthesizing Stereo 3D Views from Focus Cues in Monoscopic 2D images," *Proc. SPIE*, vol. 5006, pp. 377-388, 2003.
[9] S.G. Mallat and W.L. Hwang, "Singularity Detection and Processing with Wavelet," *IEEE Trans Info Theory,* vol. 38, no. 2, pp. 617-643, 1992.
[10] D. Comaniciu and P. Meer, "Mean Shift: A Robust Approach Toward Feature Space Analysis," *IEEE Trans. Patt. Anal. Machine Intelligence*, vol. 24, no. 5, pp. 603-619, 2002.

Fig. 5. The refined depth map based on edge defocus.

Fig. 6. The final depth map after optimization.

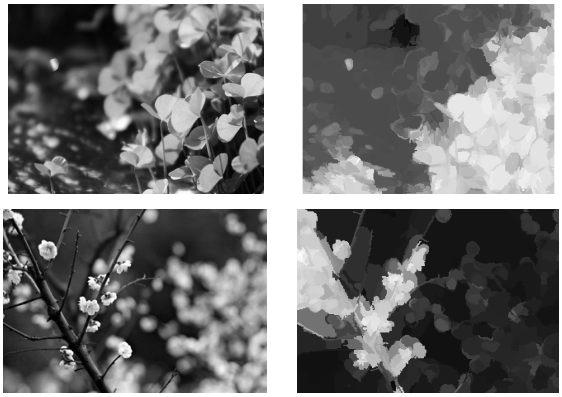


Fig. 7. Other images and their depth maps generated by the proposed method.