Fast 3D Image Depth Map Estimation Using Wavelet Analysis

Yu-Hong Lin[1], Te-Wei Chiang[2], Tienwei Tsai[3], Mann-Jung Hsiao[4*]
[1]*Dept. of Multimedia Design*, *Chihlee Institute of Technology*
[2]*Dept. of Accounting Information Systems*, *Chihlee Institute of Technology*
[3]*Dept. of Information Management*, *Chihlee Institute of Technology*
[4]*Dept. of Computer Science and Engineering, Tatung University*
yhlin@mail.chihlee.edu.tw

Abstract

This paper is to propose a method for depth estimation in a 2D image using wavelet analysis. According to our observations, the high frequency components represent focused objects in images of limited depth of field (DOF) and their high frequency wavelet subbands contain high energy. In our approach, each image is first transformed to YUV domain and the Y component is extracted for further analysis. Afterwards, the high frequency bands are derived with wavelet analysis. Through two stages of smoothing and scale manipulation, the depth map data with less error can be used for some 3D display. The experimental result shows that the proposed approach is effective and efficient for depth map estimation.

1. Introduction

Due to the huge popularity of James Cameron's Avatar, 3D movies have been guaranteed box office earnings for Hollywood movie industry in recent years. Not just in the States, even in China, the number of 3D screen has increased dramatically. However, theaters are eagerly searching for proper 3D contents to compensate their investment. There are mainly two concerns: time and money. On the contrast to the huge expense of making new 3D movies with real people action, it is much more economic to do 2D to 3D conversion from the numerously existing 2D movies. The most prompt solution is from 3D computer animations, which are ready to be converted to 3D. Therefore, Hollywood plans to convert some 2D hit movies into 3D and replay them in 3D theaters, such as Toy Story and Matrix series.

There are several ways of constructing stereoscopic images. The most affordable solution is anaglyph, which is also the technology once popular in 1950s when 3D movies were first introduced. However, due to image quality and distribution problems, the "golden age" of 3D movies just lasted 5 years. Not until 2000, digitization technology first addresses all the issues regarding 3D movie production and distribution. The audiences now have fewer problems than before, like headache, while watching 3D movies. Also, new stereoscopic image construction methods, such as shutter or polarization systems, are introduced. Philips took another approach using 2D image plus depth map to composite stereoscopic images on an autostereoscopic display[1]. The WOWvx technology provides solutions of 2D-plus-depth format content creation for 3ds Max software. In addition, BlueBox service was developed to generate 3D content from 2D videos in a semi-automatic process.

2. Previous Works

To create 2D-plus-depth format content from images of limited depth of field, the focused objects in the foreground are sharp-edged and defocused objects in the background are blurred[2]. That is, the focused area contains high frequency components and blurred area contains low frequency components. The degree of blurring can be directly related to the spatial frequency described by wavelet transformation. It suggests there are more details and less blurring in the high frequency bands with more energy, where the objects are closer to the camera. According to this observation, relative depth of each pixel can be estimated based on the values of wavelet coefficients.

Valencia and Dagnino[3] proposed a method to derive a depth map from a single image based on wavelet analysis and edge estimation based on Lipschitz exponents. The image was divided into macro blocks of 16-pixel by 16-pixel and derived 256 wavelet

coefficients by wavelet transform. The depth information associated with each pixel was estimated according to a threshold. The results came with horizontal strips because the image was handled as series of 1-D row signals. Guo et al[4] invented an incremental algorithm with the edge direction and two-dimensional characteristics considerations, which reduces the strip effects. Their approach also considered color segmentation and derived more optimal and smooth depth map with details.

3. The Proposed Approach

In this section, the proposed depth estimation algorithm in summary is described as below:

1.Y component extraction based on YUV color space. This step is to derive the luminance component of the image.

2.Wavelet-based edge detection. We use the wavelet transform to derive the horizontal, vertical, and diagonal edges. Afterwards, they are combined into one image.

3.First round of smoothing. This step is served to defocus the edges and lessen the errors caused by noises.

4.Separate the map into N-level according to a threshold to find the major focused objects.

5.Second round of smoothing. In this step, the contrast is doubled to enhance the edges of the focused objects.

3.1. Y Component Extraction

Based on the characteristics of high frequency components of focused objects in an image, the first step is to find out the focused objects using texture analysis[5]. The Y component in YUV domain of an image represents the luminance and texture of the original image. Therefore, the image is transformed from RGB domain to YUV domain at first, and then wavelet analysis is performed subsequently. See Figure 1.

3.2. Wavelet-based Edge Detection

The frequency components of an image are derived here in this stage to distinguish the edges of focused objects in the foreground. The pixels with larger value wavelet coefficients contain larger energy. At a given relative depth value, in the range of 0 to 255, the sums of the coefficients in the high frequency subbands(the H-component, V-component, and D-component) reveal the focused object details and directional information[6][7]. Larger values mean closer to the camera. The results of the components and their sum are shown in Figure 2.

3.3. First Stage of Smoothing

In order to get rid of errors and noises, the average of pixels around a pixel is calculated, which acts like a spatial rectangular filter, but still maintains the high frequency components. The smoothing algorithm used in our paper is called the rectangular smooth; it simply replaces each point in the signal with the average of adjacent points. Let m denote a positive integer called the smooth width, which is the number of points in each direction to the central point, and n represent the total number of points involved in the smoothing, then

$$S_{i,j} = \sum_{x=(i-m),\, y=(j-m)}^{x=(i+m),\, y=(j+m)} Y_{x,y} \Big/ n \quad \text{and} \tag{1}$$

$$n = (2m+1)^2, \tag{2}$$

where $S_{i,j}$ is the point in the smoothed signal and $x_{x,y}$ is the point in the original signal. Some experiments are made to show the effects of m (or n). Figure 3(a) shows the outcome of a 25-point smooth, where $m = 2$ and $n = 25$, and Figure 3(b) shows the outcome of a 49-point smooth, where $m = 3$ and $n = 49$. It is observed that the greater the parameter, the greater the degree of smoothing and, hence, the greater the suppression of the higher frequencies. The advantage of neighborhood averaging is that pixels with outlying values are forced to become more like their neighbors, but at the same time edges are preserved.

3.4. N-level Scaling

The image is then processed and the values are scaled into N-level according to a threshold. The purpose is to distinguish the focused objects. The optimal result is shown in Figure 4(b).

3.5. Second Stage of Smoothing and Double Contrast

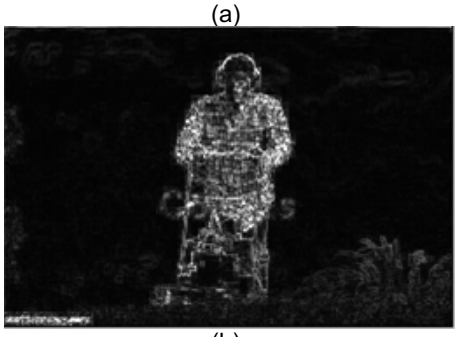
To further remove errors and noises, the second stage of smoothing is performed. Here, we use a uniform rectangular filter like the first stage smoothing with different number of pixels. Double contrast is to put the focused objects of the image into more stand-out representation. The optimal result with m=2 is shown in Figure 5.

(a)

(b)

Figure 1. (a) The original image (b) Y component of the image

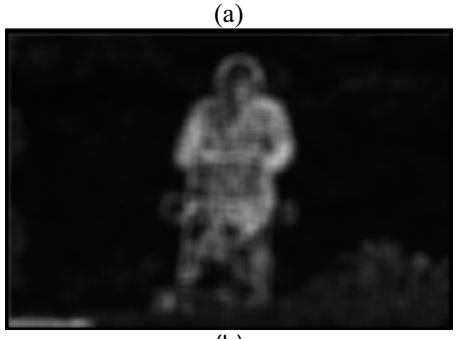
(a)



(b)

Figure 2. The wavelet analysis components (a) and the sum (b)

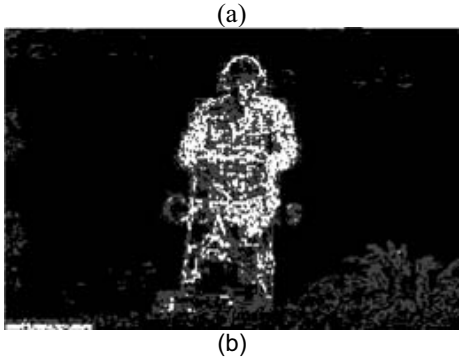
(a)



(b)

Figure 3. First smoothing results: (a) m =2 and (b) m =3

(a)



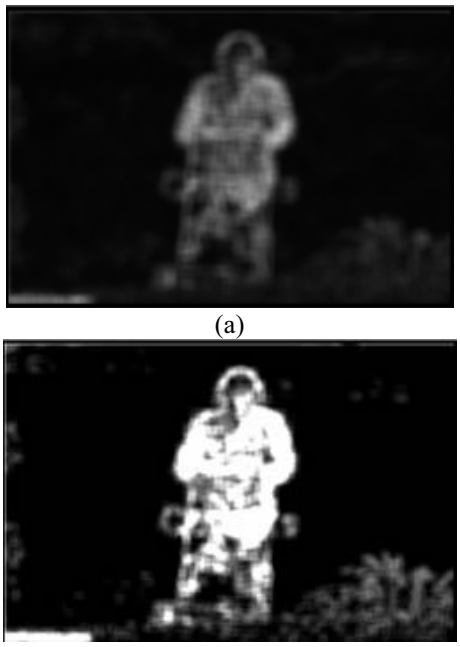
(b)

Figure 4. The N-level scaling with a threshold: (a) m=3, T=100, N=3 (b) m=3, T=100, N=5

4. Results

The same image used by Valencia and Rodríguez-Dagnino is processed in this preliminary experiment. The mowing man is the focused object with a little of

grass in the bottom and left of the foreground. Though some edge details are revealed using the wavelet analysis after Y component extraction, there are a lot of noises before the first smoothing in Figure 2(b). In the stage of N-level scaling, the focused objects are much more obvious. Afterwards, a further smoothing and double contrast are performed. Our result shows that the details of the depth map are clearly revealed in Figure 5(b), where the facial luminance difference and the grass in the front are obvious.



(a) (b)

Figure 5. Second smoothing and double contrast results: (a) N=3, m =2 and (b) N=5, m =2(optimal)

5. Conclusions

A fast depth estimation method is proposed here, by which a 2D image of limited depth of field can be displayed in a 3D format. Compared with the previous studies[3][4], although our preliminary results show that the depth map can be derived through fewer steps based on wavelet analysis, the degree of details and smoothness of the depth map estimation still need to be improved in our future study.

*M.-J. Hsiao is currently an instructor at Kang-Ning Junior College of Medical Care and Management.

References

[1] A. Redert, R.-P. Berretty, C. Varekamp, O. Willemsen, J. Swillens, and H. Driessen, "Philips 3D Solutions: From Content Creation to Visualization," *The 3rd Int. Symposium on 3D Data Processing, Visualization, and Transmission,* pp.429-431, June 2006.

[2] C. Fehn, R.D.L. Barre, and S. Pastoor, "Interactive 3-DTV – Concepts and Key Technologies," *Proc. IEEE,* vol. 94, no. 3, March 2006.

[3] 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.

[4] G. Gou, N. Zhang, L. Hou and W. Gao, " 2D to 3D Conversion Based on Edge Defocus and Segmentation", *Proc. JCASSP,* pp.2181-2184, 2008.

[5] A. D. Bimbo, Visual Information Retrieval, San Francisco: Morgan Kaufmann, 1999.

[6] Daubechies, I., "The Wavelet Transform, Time-Frequency Localization and Signal Analysis," *IEEE Trans. on Information Theory,* vol. 36, pp.961-1005, 1990.

[7] Mallat, S. G., "A theory for multiresolution signal decomposition: the wavelet representation," IEEE Trans. on Pattern Analysis and Machine Intelligence, vol. 11, no. 7, pp.674-693, 1989.