Depth Map Reconstruction Using Wavelet Analysis

Yu-Hong Lin¹ Chwen-Tzeng Su² Tienwei Tsai³ Te-Wei Chiang⁴

¹Dept. of Multimedia Design, Chihlee Institute of Technology
²Dept. of Industrial Engr. and Management, National Yunlin Univ. of Science and Tech.
³Dept. of Information Management, Chihlee Institute of Technology
⁴Dept. of Accounting Information Systems, Chihlee Institute of Technology
suct@yuntech.edu.tw

Abstract

A method for depth estimation in a 2D image using wavelet analysis has been implemented with less computation that converts two-dimensional images of limited depth of field (DOF) into depth estimation for 3D image construction. Each image is first transformed to YUV domain and the Y component of the image is further transformed to the wavelet domain for major object detection. According to our observation, the high-frequency wavelet subbands of the image represent the focused foreground. After the preliminary stage, smoothing, N-level scaling, and autocorrelation techniques are utilized to find the positions of the focused objects. The experimental result shows that the proposed approach is effective and efficient for depth map estimation.

1. Introduction

In 1950s, 3D movies were first introduced using anaglyph solution in the analogue age. However, due to image quality and distribution problems, the popularity of 3D movies did not last more than 5 years. Until 2000, James Cameron’s Avatar initiated the wave of 3D movies in this century. Digital technology solved most of the problems in the past regarding 3D movie production and distribution. With the efforts by the vendors, there are already several 3D display products for consumer markets in recent years. Users need to wear bulky glasses to access the stereoscopic visual effects, which restricts the usability for some audience. On the other hand, the number of 3D titles is much less than the traditional DVD format. Users also consider the price to acquire such new technologies. Under these limitations, the 3D display market was not as promising as the vendors predicted [1]. To get rid of glasses, a technology called Depth Image-Based Rendering (DIBR) [2] has been applied to autostereoscopic 3-D displays. In the displays, 2-D images and their associated per-pixel depth maps are compiled to form 3-D data representations. The depth maps, also called the Z-buffers, are used to describe the depth information of each point in gray levels. The stereoscopic images are then constructed through signal processors and optical devices, such as multiview or light-field systems. Philips’s multiview system, WOWvx technology, used a 2D image and its depth map to composite stereoscopic images on a display with lenticular lenses. The WOWvx technology also provided computer graphics solutions of 2D-plus-depth format content creation by 3D software. To immediately utilize the current 2D titles, BlueBox service was provided to convert 2D videos to 3D contents in a semi-automatic way.

2. Previous Works

There are several approaches based on wavelet analysis. Edge defocus estimation based on Lipschitz exponents was proposed in [3]. 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. Images were then processed as series of 1-D row signals, with the resulting horizontal stripes in the depth map. To overcome this issue, an incremental algorithm based on wavelet transform and edge focus analysis in two-dimensions was proposed in [4], 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. Second, image features were utilized to estimate the depth information. Scenes were classified into personal images, outdoor images, and close-up images. Finally, the depth map is assigned according to each category using log curves. [5]
3. The Proposed Approach

In this paper, a more simple approach is proposed to obtain the depth map of an image. In our approach, each image is first transformed to the YUV color space; then the Y component of the image is further transformed to the wavelet domain. Secondly, the high-frequency area of an image can be obtained from analyzing the high-frequency wavelet subbands of the image. After that, combinations of N-level separation and smoothing techniques are applied to find the positions of the focused objects in the image.

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[6]. 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[7][8]. Larger values mean closer to the camera. The results of the components and their sum are shown in Figure 2.

3.3. 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

\[
S_{i,j} = \frac{\sum_{x=(i-m), y=(j-m)}^{x=(i+m), y=(j+m)} Y_{x,y}}{n} \quad \text{and} \quad n = (2m+1)^2,
\]

where \( S_{i,j} \) is the point in the smoothed signal and \( x_{i,y} \) 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. Autocorrelation

Autocorrelation is used to find linear dependence of the focused object in the image. Each pixel of the wavelet subband image corresponds to a wavelet coefficient. The larger the value of the corresponding wavelet coefficient, the larger the energy within the pixel. The depth map is shifted linearly in X-Y directions and then added up. As a pixel and its neighboring pixels are high energy, this process will enhance the structure of the focused object and further remove some noise.

4. Results

The same image used by Valencia and Rodríguez-Dagnino is processed in this 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 the stage of N-level scaling, the focused objects are much more obvious. Afterwards, a further smoothing and autocorrelation are performed. Our result shows that the details of the depth map are
clearly revealed in Figure 3 where the facial luminance difference and the grass in the front are obvious.

Figure 1. (a) The Y component of the image (b) The wavelet analysis components

Figure 2. Results by Guo et al [4], (a) without edge defocus (b) refined with edge defocus

Figure 3. Our experiment result

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. Our preliminary results, compared with the results(Figure 2) by Guo et al [4], shows more details in the front. With less computation, the depth map can be derived based on wavelet analysis with combination of smoothing, N-level separation and autocorrelation. For practical use, the degree of details and refines of the depth map still need to be improved in our future study.

6. References
