The Design of Image Depth Information Extraction Algorithm Based on Joint Bilateral Filtering

Lin Yong
Department of Communication, Chongqing College of Electronic Engineering, Chongqing 401331, China
Lin Yong: 65437478@qq.com

Abstract: With the vigorous development of science and technology, specialists and scholars pay their attention to the 2D-to-3D media conversion more and more. As the next generational display, 3D-TV has lots of problems to solve, which must depend on depth estimation. Here it proposes a novel line tracing method and depth refinement filter as core of depth estimation framework. First, edge detection is performed with the downsampled input image. The line tracing algorithm traces strong edge positions to generate an initial staircase depth map. The initial depth map is further improved by a recursive depth refinement filter. It finally presents visual results from depth estimation and stereo image generation. The experimental results show that the algorithm is effective and feasible which leads a good result.

1. Introduction
With the progress of human society and the development of science and technology, people have higher and higher requirements for information processing and information exchange. Traditional machine vision can no longer meet the requirements of three-dimensional object recognition[1,2]. 2D to 3D technology has become an important direction in the development of 3D technology. Depth estimation is one of the key technologies of 3D conversion from 2D to 3D. It can automatically and effectively estimate depth information from one or more images of a scene. The conversion of 2D to 3D has achieved fruitful results overseas. Typical technology providers include ILM, Pass more and other companies. In China, however, there is still no mature technology. This paper is expected to provide a more practical method for depth estimation, improve the image quality of 3D view and users’ stereo vision comfort, and further promote the industrialization of this technology in digital television chip applications.

2. Introduction
In order to use depth information in depth estimation framework[3], a linear tracking method of Y. Chang is referred. First, before processing the edge, a grayscale image is processed. Then, linear tracing starts to work, which traces from the leftmost boundary of the edge map to the rightmost boundary. The result of linear tracing produces a line trajectory graph, which starts from the parallel line graph of the initial state, and finally forms the whole line trajectory graph. The initial line trace is composed of the number of parallel lines between the common areas of the line and the line. The number of lines is an important parameter affecting the whole result. The area depth between the bottom trajectories is almost zero, while the area between the top trajectories is the largest. Therefore, a ladder shaped line trace map is obtained.
After deep improved filtering, a stereo image is transformed from the depth map to [4]. The horizontal transfer process is based on the depth value of each pixel in the depth map.

3. algorithm description

3.1 Grayscale acquisition and down sampling

Research shows that parallax and 2D video can generate binocular vision. In image processing, YUV space is usually used to represent a frame of 256 color grayscale, that is depth map. Y and UV denote brightness and chromaticity respectively. Each of the 3 components is represented by 8bit, which is represented by 10 in the middle of 0-255. In a gray-scale image with U=V=128, the brightness value Y of each pixel and 2D image form a one-to-one correspondence, and the relative distance between the human eye and the 2D pixel of the corresponding point is expressed by 0 to 255. It is usually agreed that white (255) means the nearest distance, while black (0) means the farthest distance from the human eye. Therefore, the depth information of objects in the 2D diagram is more intuitive. If you want to see a 3D stereo image, you can use a 3D display to process ordinary 2D video and add the corresponding depth map[5].

In this article, we use the following formula to convert the color RGB value to the depth value [5].

\[
\text{Gray} = (R \times 38 + G \times 75 + B \times 1) \geq 7
\]  

(1)

Image down-sampling is an operation to reduce computational complexity. In image super-resolution reconstruction, image resampling is often involved. In all the deduction formulas, image is vectorized, and then the vector is multiplied by a down-sampling matrix D. In this paper, we use matlab to generate the down sampling matrix D of any scale.

For an image with M*N size, the resolution image with (M/s)* (N/s) size can be obtained by S-TIMES down-sampling. So s should be the common number of M and N. If a matrix image is considered, the original image in the s*s window is transformed into a pixel whose value is the average of all the pixels in the window:

\[
p_k = \frac{\sum_{i \in \text{win}(k)} I_i}{s^2}
\]  

(2)

After the image is vectorized into a vector of 1*(MN), the downsampling process should also have a corresponding matrix whose size is (MN/s^2)*(MN).

Take a 4*4 image as an example: the number represents the pixel location [6].

| 0 | 4 | 8 | 12 |
|---|---|---|----|
| 1 | 5 | 9 | 13 |
| 2 | 6 | 10| 14 |
| 3 | 7 | 11| 15 |

The vectorization size of the image is 16*1, and the element is:

[0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15]

The new image after 2 times downsampling has four pixels, which are the mean of 0 145 position, 23 677 position, 8 9 1213 position and 10 11 1415 position respectively.

3.2 Edge detection Sobel operator and its implementation

Compared with other common edge detection operators, Sobel operator has the advantages of fast detection speed, smoothing effect on noise, and some ability to suppress noise. However, because it detects some false edges, it makes the edges rough and reduces the accuracy of detection and positioning. In order to solve these problems, neighborhood or weighted averaging and first-order differential processing are carried out sequentially, and then edge detection is carried out to get edge [7]. Figure 2 shows the matrix window used by the operator (both horizontal and vertical).
1) detect vertical edges in horizontal gradient direction.
2) vertical gradient direction, horizontal edge detection

Figure 1. Sobel Operator horizontal and vertical template

The operator matrix window (3) and (4) represent the horizontal and vertical convolution operations respectively, and the gradient values are obtained by using the following formulas:

\[
\Delta x = \left[ f(x-1,y-1) + 2f(x,y) + f(x+1,y+1) \right] - \left[ f(x,y) + 2f(x,y-1) + f(x+1,y+1) \right]
\]

\[
\Delta y = \left[ f(x-1,y-1) + 2f(x,y) + f(x+1,y+1) \right] - \left[ f(x-1,y) + 2f(x,y-1) + f(x+1,y) \right]
\]

The algorithm is to make the image two valued, and set the threshold TH to achieve it. The idea of edge detection is that the gradient value of non-edge points is less than the threshold value, on the contrary, the edge points. The main steps of the algorithm are as follows:

1) Move the two direction templates from one pixel to another along the image, and then overlap one pixel position with the center of the pixel.
2) multiplying the coefficients in the template with the corresponding pixel values:
3) add up all the multiplied values.
4) using the values of two convolutions, calculate the gradient value of the place, that is, the new gray value.
5) Select the appropriate threshold TH, if the gray value of the new pixel is greater than or equal to TH, then the pixel is the image edge point.

3.3 Linear tracking

The line tracking algorithm improves an energy function. This energy function is modeled by three constraints. The first one is the constraint condition for tracking the Strong boundary. The second is a smoothing constraint for punning sudden changes in the vertical direction. The third is an elastic constraint on the penalty for significant changes in the vertical direction of the line, so that the vertical position can be avoided from being too far from the initial vertical position.

The following equations describe the three constraints [8]:

Constraint condition 1: edge tracking condition,

\[
E_e(x,y) = \exp\left(-\text{edge}(x,y)/a\right),
\]

Constraint condition 2: smoothing constraints:

\[
E_s(x,y) = d_s(x,y)/b,
\]

Constraint condition 3: elastic constraint conditions:

\[
E_e(x,y) = d_e(x,y)/c.
\]

The control parameters a,b,c, depend on the characteristics of the input image and are tentatively, \(E_e(x,y)\) defined as edge tracking constraints, \(E_s(x,y)\) as the control parameters of edge trajectory constraints, \(E_e(x,y)\) as the control parameters of smooth constraints, \(E_s(x,y)\) as the elastic constraints, \(\text{edge}(x,y)\) as the control parameters of elastic constraints, representing the boundary values of points on the edge graph. Represents the vertical distance from the current pixel point to the substitute pixel point. Represents the vertical distance from the starting position of the left boundary on the original line diagram to the point substitute pixel.
3.4 Deep assignment
Line trajectory tracking will track the obvious edges of the edge image from left to right, and get the horizontal non-intersecting line trajectory map[8-10]. Next, we will get the gradient depth map from the bottom to the top according to the line trace. Specific steps are as follows: the distance D of the horizontal trajectory, the position of the initial trajectory and the depth corresponding to each horizontal trajectory are determined by the initial line number n of the line tracing graph and the reference image line number Hi.

The depth assignment of linear graphs must strictly obey the rule of increasing from bottom to top. Each trajectory corresponds to a fixed depth value, which is scanned from the bottom to the top of the column in the assignment. At the beginning, the value is 255. At the time of scanning to line 1, the value is 255-1*d, D is the interval of depth values, until the next trajectory line n is scanned, the value is 255-n*d, and so on, until a row of scanning and assignment is completed and the next column is carried out.

3.5 Joint bilateral filtering
The filter is a mathematical model, through which the image data can be transformed into energy. If the energy is low, the filter can be eliminated. Noise is a low-energy part. If the ideal filter is used, the ringing phenomenon will appear in the image. Using the joint bilateral filter, the system function is smooth, avoiding the ringing phenomenon.

The two nearest key frame is very important for joint bilateral filtering. The key frame is the basis of depth image generation in a certain direction. The nearest backward key frame is the non-key frame. The depth $d_{bw}$ is generated by the inverse bilateral filtering depth recursion algorithm. Next, we use linear interpolation to combine the two methods to get a new depth.

$$t_{ff} = (1 - \frac{t}{T})d_{ff} + \frac{t}{T}d_{bw}$$

The $t$ and $T$ in the formula represent the time distance between the current frame and the forward key frame and the time distance between the two key frames, respectively. However, it can not really reflect the confidence of depth generated from each direction, because the depth weights generated from the front and back end are based on the time distance. The final result is that both the unobscured and the obscured areas may be blurred, even if the averaging reduces part [11].

A simple example of a deep improvement filter will be given by the above equation. Where x and Y represent the coordinates of the image, Z represents the initial depth value calculated by the second part of linear tracking. The first equation describes the task of feature extraction module. The weight of the filter is determined by a series of pixel pairs, and the same quantity is determined by the size of the filter. Here, Y represents pixel brightness while Sigma is a filter parameter. The second equation describes the task of cyclic depth filtering module. Among them, K denotes normalization factor and the ETA represents a set of adjacent pixel positions that describe the filter's length range. This filtering operation runs once or repeatedly until the final depth map is obtained. It relies on the stereograph descriptor. The sample, which is obtained through the use of single-scale loop deep filtering, is shown.

4. Experimental results
In this paper, we use downsampling, Sobel operator for edge detection, linear tracking, depth assignment and joint bilateral filtering to achieve a depth estimation process. The programming tool is MATLAB R2010a. The experiment is based on Win7 Professional platform. The experiment uses ordinary 2D images with resolution of 642*642.

The depth estimation quality of this algorithm is closely related to the parameter setting of Table 2. Generally, the number of line trajectories is 50. When the number of line trajectories is less than 10, discontinuous perceptual depth will be generated due to large depth changes. The control parameters a are generally the average of the pixel values of the edge trajectory, B and C are 1/4 of the image height, the weight a is 0.4, and both beta and are 0.3. In order to obtain depth information, the algorithm only uses monocular video sequence. Although the accuracy of this method is not as good as that of
binocular system, it can be seen from the results that the depth map is clearer after joint bilateral filtering. Although depth blur still exists, it has been greatly reduced. Experiments show that the algorithm is simple and easy to use and can get more accurate depth maps, which provides a good basis for further improvement of depth maps.

Table 2. Parameter values used in depth estimation

| a   | b   | c   | α  | β  | Υ  | n  |
|-----|-----|-----|----|----|----|----|
| 0.107 | 125 | 125 | 0.4 | 0.3 | 0.3 | 50 |

5. Conclusion

Image depth estimation is an important basis of computer graphics based on image feature analysis and extraction. It is widely used in the fields of 3D display, animation, film and television, computer design, stereo photography and so on. This paper explores some of the current mainstream image depth estimation algorithms, and also studies the related technologies involved in depth estimation. Aiming at depth estimation in 3D TV system, a depth estimation algorithm based on relative height cues is proposed. The algorithm uses horizontal line trace to track edge image and template near bottom to achieve depth estimation. The experimental results show that the proposed algorithm can effectively extract the depth information of the image, and the computational complexity is small. It is a method of extracting the depth information of the image that is worth popularizing and applying.

Acknowledgment

Science and technology research project of Chongqing Education Commission(No. KJQN201803104)
Science and technology research project of Chongqing College of Electronic Engineering (No.XJZK201806)

References

[1] P. Harman, J. Flack, S. Fox and M. Dowley, “Rapid 2D to 3D Conversion[M],” Proceedings of SPIE, (2002)
[2] Y. Chang, C. Fang, L. Ding, S. Chen, and L. Chen, Map Generation for 2D-to-3D Conversion by Short Motion Assisted Color Segmentation[J], International Conference on Multimedia and Expo, 2007, 12(6): 1958-1961
[3] J. Ko, M. Kim and C. Kim, “2D-To-3D Stereoscopic Conversion: Depth-Map Estimation in a 2D Single-View Image[J],” Proceedings of SPIE, Vol. 6696, (2007)
[4] S. Battiato, A. Carpa, S. Curti and M. La Cascia, “3D Stereoscopic Image Pairs by Depth-Map Generation[J],” Proceedings of 3DPVT, (2004)
[5] Cheng, C.-C., C.-T. Li, L.-G. Chen. A novel 2Dd-to-3D conversion system using edge information[J]. IEEE Transactions on Consumer Electronics, 2010. 56(3): 1739-1745.
[6] W. Tam, A. Yee, J. Ferreira, S. Tariq and F. Speranza, “Stereoscopic Image Rendering Based on Depth Maps Created From Blur and Edge Information[J],” Proceedings of SPIE, (2005)
[7] Tam, W.J., L. Zhang. 3D-TV content generation: 2D-TO-3D conversion[C]. in 2006 IEEE International Conference on Multimedia and Expo, ICME 2006, July 9, 2006 - July 12, 2006. 2006: 1869-1872.
[8] S. Valencia and R. Dagnino, “Synthesizing Stereo 3D Views from Focus Cues in Monoscopic 2D Images[M],” Proceedings of SPIE, (2003)
[9] LIU Ran, XIE Hui, TAI Guoqin, TAN Yingchun. An Approach to Eliminate Folds Based on view judgment for DIBR. Tongji Daxue Xuebao/Journal of Tongji University, 2013. 41(1):142-147
[10] A. Redert, R.-P. Berretty, C. Varekamp, O. Willemsen, J. Swillens, and H. Diessen, Philips 3D Solutions: From Content Creation to Visualization, The 3rd Int. Symposium on 3D Data Processing, Visualization, and Transmission[J], 2006, 14(6): 429-431
[11] J. F. Canny, “A Computational Approach To Edge Detection[J],” IEEE Trans. Pattern Analysis and Machine Intelligence, 8, 679-714, (1986)