International Journal of Multimedia and Ubiquitous Engineering
Vol. 10, No. 2 (2015), pp. 395-404
http://dx.doi.org/10.14257/ijmue.2015.10.2.36

Stereo Matching Based on Least Square

Yu Haihao, Fang Yurun, Kong Leilei and Wang Xin

Heilongjiang Institute of Technology
Harbin, China

Abstract

Least Square method is widely adopted in stereo matching owing to it high precision, but the fact that transformation parameters is obtained by solving linear equations leads to the instability of its solutions and the process of matching oscillates and decreases convergence speed. To overcome this disadvantage, improve convergence speed and keep high precision, this paper provides gradient method to resolve stereo matching. The experiments show that the algorithm is valid and practical.

Keywords: *least square, Stereo matching, Gradient descent method*

1. Introduction

Stereo matching is one of important research areas in computer vision, which obtains the disparities by finding corresponding points in two or more images taken from different viewpoints. Then, depths are calculated by their disparities, which are fed into the process for three dimensional reconstruction and scene synthesis.

Stereo matching is the most critical and most difficult step in binocular stereo vision. The difference from ordinary image registration is that stereo images are taken from two cameras at the same time or one camera at different time. The different gray values of corresponding points and occlusions produced by different viewpoints and shooting at different time make matching process more difficult. So far, a lot of matching algorithms have been proposed. They are classified into four citatory: area based stereo matching [1, 2], feature based stereo matching [3-10], phase based stereo matching [12-14] and energy based stereo matching [15, 16]. In the feature-based stereo matching, the correspondences are established for the features extracted from images, such as edges [3, 4], line segments [5-7] and curves [8-10] and so on. Its advantages are high accurate disparities and low time and space complexity. However, its disadvantage is that sparse disparity maps only can be obtained and interpolation for dense disparity maps introduces errors again. In area based stereo matching, disparities are obtained by local regions around the pixel under consideration, which is very suitable for texture regions, but produces poor results at discontinuities, especially for slope. In phase base stereo matching, the displacement in space domain is translated into phase in frequency domain according to Fourier Transform and disparities are computed by phase difference or phase correlation. In this method, the disparity range should be less than the half wavelength of the image, so multi-resolution strategy is used for solving the large disparity range. Phases chosen as matching elements can reflect the structural information of the signal very well and have good noise suppression and produce subpixel disparities. However, frequency transfer confusion influences matching accuracy. The energy-based matching algorithm obtains disparities of corresponding points by minimizing the global energy function, which consists of the data term and the smooth term. The performance of the algorithms of the type is superior to local algorithms.

ISSN: 1975-0080 IJMUE

Stereo matching algorithm based on least square obtains disparities by minimizing the sum of the squared difference of gray levels of pixels between local regions around the pixels under consideration. The algorithm of this type assumes that disparities of corresponding points satisfies linear transformation, and then obtains transformation parameters by solving the over determined linear equations. This process will lead to the instability and fluctuation of the solution. In order to overcome this drawback, this paper employs the gradient descent method to solve the equations along the negative gradient direction to achieve transformation parameters, where the minimum of the equations is obtained.

2. Traditional Stereo Matching Algorithm based on Least Square

In the disparity estimation based on least square, the local window is chosen in the left and right image, and then assumed that pixels in local windows satisfy linear transformation, and then the least square is used to compute the transformation parameters, lastly disparities are calculated by the transformation parameters. The process is shown in Figure 1.

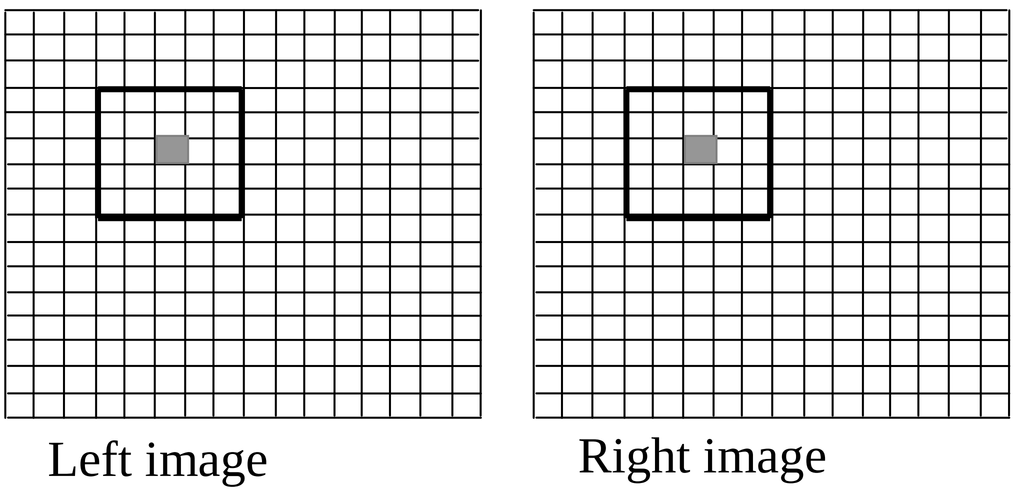


Figure 1. Stereo Matching Process

Let (x_1, y_1) and (x_2, y_2) the ordinates of the image of the same scene element, in the left and right images respectively. According to the gray constancy assumption, their gray values remain unchanged:

$$l(x_1, y_1) = r(x_2, y_2) \tag{1}$$

In the imaging process, Illumination change and sensor noise renders their gray values inconsistent, so they are related by

$$l(x_1, y_1) = r(x_2, y_2) + e \tag{2}$$

where e denotes the imaging noise. In order to obtain the disparity of the corresponding pixels, Equation (3) need to be minimized in local windows of the left and right images

$$\min = (l(x_1, y_1) - r(x_2, y_2))^2 \tag{3}$$

The pixels in local windows around the pixels (x_1, y_1) and (x_2, y_2) satisfy the linear transformation:

$$\begin{aligned} x_2 &= k_0 + k_1 x_1 + k_2 y_1 \\ y_2 &= k_3 + k_4 x_1 + k_5 y_1 \end{aligned} \tag{4}$$

Plunging Equation (4) into Equation (5):

$$\min = (l(x_1, y_1) - r(k_0 + k_1 x_1 + k_2 y_1, k_3 + k_4 x_1 + k_5 y_1))^2 \tag{5}$$

The $r(x_2, y_2)$ in Equation (5) is expanded at the (x_1, y_1) :

$$\begin{aligned} r(x_2, y_2) = \frac{\partial r}{\partial x} dk_0 + x\frac{\partial r}{\partial x} dk_1 + y\frac{\partial r}{\partial x} dk_2 + \frac{\partial r}{\partial y} dk_3 + \\ x\frac{\partial r}{\partial y} dk_4 + y\frac{\partial r}{\partial y} dk_5 + r(x_1, y_1) \end{aligned} \tag{6}$$

The difference of the gray values of corresponding pixels is：

$$\begin{aligned} \min = r - l = \frac{\partial r}{\partial x} dk_0 + x\frac{\partial r}{\partial x} dk_1 + y\frac{\partial r}{\partial x} dk_2 \\ + \frac{\partial r}{\partial y} dk_3 + x\frac{\partial r}{\partial y} dk_4 + y\frac{\partial r}{\partial y} dk_5 \\ + r(x_1, y_1) - l(x_1, y_1) \end{aligned} \tag{7}$$

The traditional least square is based on the minimizing of the sum of the squared differences of gray values in local windows. In order to obtain the minimum of Equation (7), the derivatives with respect to linear transformation parameters are taken:

$$\frac{\partial \sum \min^2}{\partial dk_i} = 0 (i = 0, 1 \ldots 5) \tag{8}$$

The linear transformation parameters are obtained according to iteratively solving linear equations (8), with which disparities is computed.

Stereo matching based on least square is an iterative process. Firstly, assume that very pixel has an initial disparity, and then linear transformation parameters are computed according to least square estimation, and then corresponding points are calculated according to transformation parameters, and then linear transformation parameters are computed according to least square estimation again until the residual error between local windows is less than the threshold. Lastly, the coordinates of corresponding points are obtained according to equation (4) and the disparity of corresponding points is given by

$$d = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2} \tag{9}$$

3. Least Square based on Gradient Descent Method

3.1. Gradient Descent Method

Gradient descent method is a first-order optimization algorithm, which determines the search direction by using the negative gradient direction, then along this direction the steps proportion to the negative direction are taken to find a local minimum of a function.

If real-value function $F(\mathbf{x})$ is defined and is differentiable in a neighborhood of a point $\mathbf{a}$, then $F(\mathbf{x})$ decreases fastest in the direction of the negative gradient of F .It follows that, if

$$b = a - \gamma \nabla F(a)$$

for $\gamma > 0$ a small enough number, then $F(a) \geq F(b)$ .According to the above analysis, starting with a guess $\mathbf{x}_0$ to search for a local minimum of F , and considering the sequence $\mathbf{x}_0, \mathbf{x}_1, \mathbf{x}_2 \ldots$ such that

$$x_{n+1} = x_n - \gamma_n \nabla F(x_n), n \geq 0 .$$

We have

$$F(x_0) \geq F(x_1) \geq F(x_2) \geq \ldots$$

The value of the step size γ is allowed to be changed at every iteration. With certain assumption, convergence to a local minimum can be guaranteed.

3.2. Least Square based on Gradient Descent for Stereo Matching

Stereo matching process is a process in which corresponding points are found in the left and right images. According to the assumption that the world is composed of Lambertian planes, that is, the gray values of corresponding points remains constant, the value of F for corresponding points should be equal to zero , but in practice the value of F is not equal to zero because of image noises and lighting situation. The least square stereo matching based on gradient descent computes transformation parameters by finding local minimum of the objective function F along the negative gradient direction. The objective function is given by

$$F = (r(x_1, y_1) - l(x_2, y_2))^2 \tag{10}$$

$$\nabla F = [\frac{\partial F}{\partial k_1} \quad \frac{\partial F}{\partial k_2} \quad \frac{\partial F}{\partial k_3} \quad \frac{\partial F}{\partial k_4} \quad \frac{\partial F}{\partial k_5}]^T \tag{11}$$

$$\vec{k} = \vec{k}_0 - \lambda \nabla F \tag{12}$$

Where $\vec{k}$ denotes transformation parameters vector, $\vec{k}_0$ is a initial value such as $\left[0 \ 1 \ 0 \ 0 \ 1 \ 0\right]^T$. After several iterations, $\vec{k}$ is found, on which the value of the objective function achieves the minimum. Then, the coordinate of corresponding points is computed by the computed linear transformation parameter. Furthermore, disparities are obtained.

In the actual iterative process, in order to make the matching process more stable and more robustness against noise, the objective function is applied in local windows around the pixels under consideration. The corresponding objective function and gradient direction are given by

$$F = \sum_{-w/2<i<w/2} \sum_{-w/2<j<w/2} r(x_1 + i, y_1 + j) - l(r(x_2 + i, y_2 + j))^2$$

$$\nabla F = \sum_{-w/2<i<w/2} \sum_{-w/2<j<w/2} [\frac{\partial F}{\partial k_1} \quad \frac{\partial F}{\partial k_2} \quad \frac{\partial F}{\partial k_3} \quad \frac{\partial F}{\partial k_4} \quad \frac{\partial F}{\partial k_5}]^T \tag{13}$$

The specific steps of least square stereo matching method based on gradient descent are as follows:

(1) For any pixel $\left(x_1, y_1\right)$ in the reference image and the initial corresponding point $\left(x_2, y_2\right)$ in the matching image, According to linear transformation parameters k_1, k_2, k_3, k_4, k_5 ,their relations is established by

$$x_2 = k_0 + k_1 x_1 + k_2 y_1$$
$$y_2 = k_3 + k_4 x_1 + k_5 y_1$$

In general, $k_0 = 0$, $k_1 = 1$, $k_2 = 0$, $k_3 = 0$, $k_4 = 1$, $k_5 = 0$ is chosen as the initial value of the transformation parameters.

(2) Resample the window in the matching image by using the interpolation algorithm according to the coordinates of corresponding point $\left(x_2, y_2\right)$.

(3) Solving the linear transformation parameters according to the gradient descent method.

(4)Compute residual error F ,if F is less than a threshold, then stop and go to (5),else go to(1).

(5) Compute x_2 and y_2 according to the linear transformation parameters.

(6) Compute a disparity according to Equation (9).

4. Post Processing

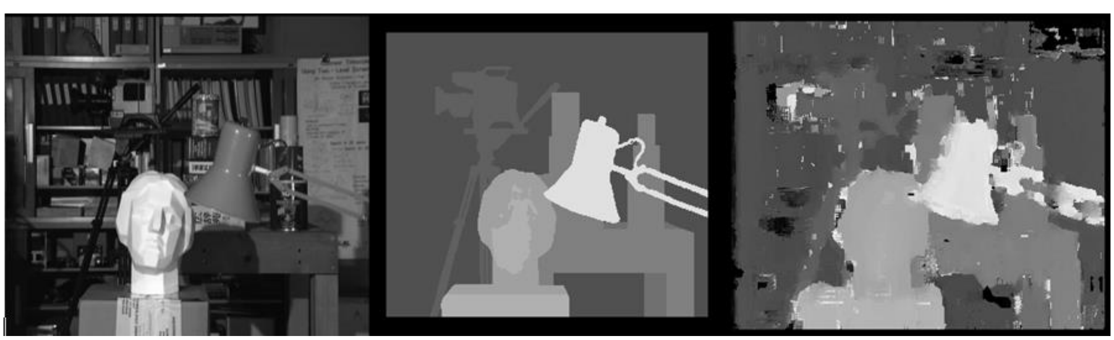
Occlusion in stereo images is a region, which only is present in an image and is not in another image. In the stereo matching process, a lot of mismatches are produced due to occlusions. In our approach, occlusions are detected by checking the left-right consistency. First, the left image is chosen the reference image, and the left disparity map is computed. Then the right image is chosen the reference image, and the right disparity map is computed. For any pixel $\left(x_1, y_1\right)$ in the left image and its disparity, the points that do not satisfy

$$d_l\left(x_l, y_l\right) = -d_r\left(x_r + d_l\left(x_l, y_l\right), y_r\right)$$

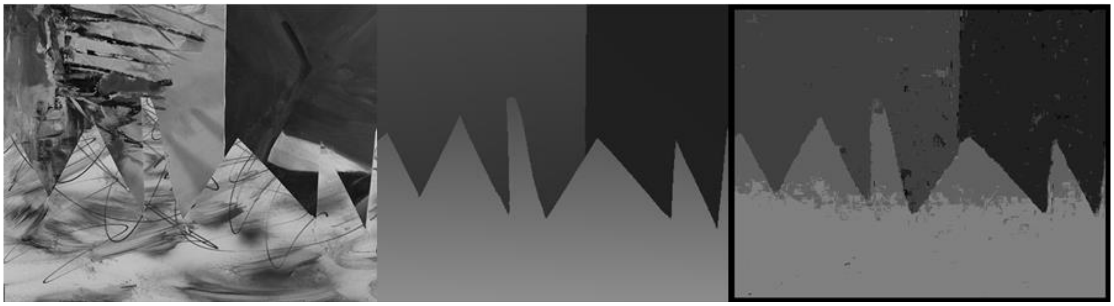
are considered as occlusion. Finally, disparities of occlusions are filled by disparities of pixel neighboring it.

5. Experimental Results and Analysis

In order to verify the effectiveness of the proposed algorithm, stereo images provide by Middlebury are adopted. This proposed algorithm is tested using Tsukuba, Sawtooth, Map and Venus on The Middlebury stereo benchmark dataset and is evaluated by measuring the percent of bad matching pixels (where the absolute disparity error is larger than 1 pixel) for three subsets of an image: nonocc (the pixels in the nonoccluded region), all (all the pixels), and disc (the visible pixels near the occluded regions). In this experiment, the window size 19×19 and the step size 0.1 are used. The experimental results are showed in Figure 2, which shows that our experimental result is nearly close to true disparity maps.



(a)Tsukuba, ground truth and computed disparity map



(b) Sawtooth, ground truth and computed disparity map

(c)Map, ground truth and computed disparity map

(d) Venus, ground truth and computed disparity map

Figure 2. Resullt of Tsukuba, Sawtooth, Map and Venus

The disparity accuracy is related with the window size and the step size. The principle of the window selection is that in the textureless area the large window should be selected to include much enough gray variation; in texture area the small size should be selected to avoid depth discontinuities. An experimental result is shown in Table 1.

Table 1. Impact of Window Size on Accuracy

Window size	Tsukuba	Sawtooth	Venus	Map
9×9	19.54	21.15	17.41	18.24
17×17	17.39	18.31	15.41	14.37
21×21	12.27	14.12	11.49	11.62
25×25	15.25	17.37	12.27	16.45

For analysis of the performance of our proposed stereo matching algorithm, we compares our algorithm with other stereo matching algorithm listed in reference, which is showed in Table 2. We found that the matching accuracy of the proposed method at the top of all compared stereo matching methods.

Table 2. Performance Evaluation of Disparity Accuracy

Algorithm	Tsukuba			Sawtooth			Map			Venus		
	nocc	all	disc	nocc	all	disc	nocc	all	disc	nocc	all	disc
Our method	2.09	2.44	7.12	0.14	0.38	1.92	5.92	11.3	15.5	2.70	7.93	7.48
Ref.[1]	1.51	1.85	7.61	0.20	0.39	2.42	6.16	11.8	16.0	2.71	8.24	7.66
Ref.[16]	1.45	1.83	7.71	0.14	0.26	1.90	6.88	13.2	16.1	2.94	8.89	8.32
Ref.[7]	2.11	2.38	7.45	0.17	0.39	2.17	6.53	11.9	16.3	2.91	8.16	7.97
Ref.[9]	1.88	2.35	7.64	0.38	0.82	3.02	5.99	11.3	13.3	2.84	8.33	8.09
Ref.[14]	1.85	2.51	7.45	0.35	0.88	3.01	6.28	12.1	14.3	2.80	8.91	7.79
Ref.[6]	1.38	1.96	7.14	0.44	1.13	4.87	6.80	11.9	17.3	3.60	8.57	9.36
Ref.[2]	1.10	1.67	5.92	0.53	0.89	5.71	6.69	12.0	15.9	2.60	8.44	6.71

Our matching algorithm improve the speed in comparison with a stereo matching algorithm based on traditional least squared (or MATLS, for short). We use stereo images Tsukuba, Sawtooth, Map, Venus and select the window size 17×17 to test matching speed of our matching algorithm and MATLS. Experimental results are shown in Table 3.

Table 3. Comparison of Matching Speed (Unit: Second)

Algorithm	Tsukuba	Sawtooth	Map	Venus
Our method	0.188	0.406	0.422	0.125
MATLS	0.485	0.862	0.875	0.328

The results of some new representative data on the Middlebury website are presented in Figure 3. They are: Dolls, Cloth2, Baby3 and Wood1. These stereo pairs are captured by high-end cameras in a controlled lab environment. The obtained disparity maps are quite close to the ground truth. Data Dolls presents indoor scene with many stacking objects. The proposed method can calculate accurate disparity values for most parts of the scene, and the disparity of small toys on the floor is correctly recovered. For data Cloth2, a smooth disparity map is generated for continuous cloth with repeated texture. In Baby3, objects with curved surfaces are presented, e.g. ball and cylinder. Disparity of these objects is both accurate and smooth. The disparity of the background (map) is also obtained with few mismatches. InWood1, the texture information is much weaker, nevertheless, our method can still generate accurate and smooth disparity map that is close to the ground truth. In fact, high-definition stereo sequences are much easier to be obtained, and accurate disparity maps can be generated by the proposed method.

(a) (b) (c)

Figure 3. Results of Representative Data on the Middlebury Website. From the Top to Bottom are: Dolls, Cloth2, Baby3 and Wood1. (a) Left Image; (b) Results of the Proposed Method; (c) Ground truth

6. Conclusion

In order to overcome high computational complexity and slow convergence in the traditional least square stereo matching algorithm, we propose a least square stereo matching algorithm based on the gradient method. Experimental results shows that this proposed algorithm is a fast and effective algorithm and can obtain satisfactory results.

Acknowledgements

This paper is supported by National Natural Science Foundation of China(61370170), National Social Science Fund Project - Youth Project of China(14CTQ032) and Heilongjiang Province Educational Committee Science Foundation(12541683).

References

[1] H. Han, X. Han and F. Yang, International Journal for Light and Electron Optics, vol. 16, no. 9, **(2014).**
[2] X. Z. Zhou, G. J. Wen, G. Jian and R. S. Wang, Chinese Journal of Computers, vol. 3, no. 1, **(2006).**
[3] T. Hu, B. Qi, T. Wu, X. Xu and H. He, Computer Vision and Image Understanding, vol. 7, no. 2, **(2012).**
[4] S. B. Pollard, J. E. W. Mayhew and J. P. Frisby, Perception, vol. 14, no. 1, **(1985).**
[5] Ayache and N. F. Lustman, "Fast and reliable passive trinocular stereovision", Proceedings of the 1th International Conference on Computer Vision, **(1987)** August 8-11, London, UK.
[6] J. H. Mcintish and K. M. Mutch, Computer Vision Graphics Image Process, vol. 43, no. 1, **(1988)**.
[7] G. Medioni and R. Netatia, Computer Vision Graphics Image Process, vol. 31, no. 1, **(1985).**
[8] A. T. Brint and M. Brady, Image and Vision Computing, vol. 8, no. 12, **(1990)**.
[9] N. M. Nasrabadi, IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 14, no. 5, **(1992).**
[10] L. Robert and O. Faugeras, "Curve-based stereo, Figural continuity and curvature", Proceedings of IEEE International Conference Computer Vision Pattern Recognition, **(1991)** June 3-6, Maui, Hawaii.
[11] J. Barron, D. Fleet, S. Beauchemin and T. Burkitt, "Performance of optical flow techniques", Proceedings of the IEEE on Computer Vision and Pattern Recognition, **(1992)** June 15-18, Champaign, IL.
[12] D.J. Fleet and A.D. Jepson, Journal of Computer Vision, vol. 5, no. 1, **(1990)**.
[13] D. J. Fleet and A. D. Jepson, IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 15, no. 12, **(1993)**.
[14] C. Kuglin and D. Hines, "The phase correlation image alignment method", Proceedings of International Conference on Cybernetics and Society, **(1975)** September 23-25, San Francisco, California.
[15] F. Heitz, P. Perez and P. Bouthemy, Image Understanding, vol. 1, no. 59, **(1994)**.
[16] L. Robert, R. Deriche, "Dense depth map reconstruction, A minimization and regularization approach which preserves discontinuities", Proceedings of the 4th European Conference on Computer Vision, **(1996)** April 15-18, Cambridge, UK.

Author

Yu Haihao, he was born in Huanan, China in 1974. He received his Bachelor of Engineering degree in Welding Technology and Equipment from Jia MuSi University in 1997 and the Master of Engineering degree in Technology of Computer Application from Northeast Forestry University in 2008. His major field of study is Artificial Intelligence.