Fast sub-pixel accuracy stereo image matching based on disparity plane

Nan Jiang, Yufu Qu, Yueping Li
School of Instrument Science and Opto-Electronics Engineering, Beihang University, Beijing 100191, China

ABSTRACT

To avoid "stair-casing effect" in the disparity map when dealing with slant plane, curved surface and weak texture region, an improved fast dense stereo matching algorithm based on disparity plane estimation is proposed. First, a set of support points are extracted from the edge of the original matching images and the description images. Second, Delaunay triangulation disparity planes are calculated using all the support points. Third, the sub-pixel disparity map is computed from the best support points and parameters of the Delaunay triangulation disparity planes. Finally, experimental results show that the "stair-casing effect" caused by slant plane, curved surface and weak texture region is eliminated by using the presented method. In addition, the proposed method spends less than 600ms on a one-megapixel image averagely.

Keywords: Fast dense stereo matching, 3D reconstruction, sub-pixel, disparity plane, stair-casing

1. INTRODUCTION

Three-dimensional reconstruction of objects is an important issue in the fields such as computer graphics, computer measurement and robotics. There is a variety of ways to obtain the three-dimensional information, among these ways, reconstruction based on dense stereo matching has become a very important method because of its low cost, simple control and high precision. Four steps are generally included in dense stereo matching algorithms: matching cost computation, cost aggregation, disparity computation and disparity refinement. According to whether cost aggregation is involved or not, dense stereo matching algorithms can be divided into two types of methods, local methods and global methods. Local methods compute the matching costs by the image intensities or color values inside fixed or varying support windows. Global methods calculate the disparity of each pixel by global energy functions which involve all other pixels. Though global methods can obtain highly accurate matching results, they require a long time computation and a huge computing space which cannot reach the requirement of real-time three-dimensional reconstruction. Local methods, however, have a fast computing speed, but generate "stair-casing effect" and have lower accuracy when dealing with slant plane, curved surface and weak texture region.

In order to avoid "stair-casing effect" in the disparity map, a few methods are introduced to obtain sub-pixel disparity maps. At first, interpolation and fitting were taken by researchers for disparity refinement. Tian Qi [1] used parabolic fit method to get sub-pixel disparity, but this method took probability of disparity nearby the integer pixels far more than the probability around 0.5 pixels. Then Nehab et al. [2] improved Tian Qi's [1] method by means of two-dimensional fitting. These methods could easily be achieved but had a high computational complexity, and the algorithm accuracy of these methods were affected by the effect of interpolation. Instead of interpolation and fitting, some methods based on

2015 International Conference on Optical Instruments and Technology: Optoelectronic Imaging and Processing Technology, edited by Guangming Shi, Xuelong Li, Bormin Huang, Proc. of SPIE Vol. 9622, 96221C · © 2015 SPIE · CCC code: 0277-786X/15/$18 · doi: 10.1117/12.2193512

correlation algorithm made a good effect on sub-pixel stereo matching. Takita et al. [3] proposed a high-accuracy sub-pixel method based on phase-only correlation. By improving Takita's [3] method, Takuma Shibahara et al [4] used image Gaussian pyramid to calculate cost aggregation and reduce the search range from two-dimensional to one-dimensional which reduced the computational complexity. To get higher accuracy, a method using color correspondences between different scales of images was proposed by Yang et al. [5]. The method obtained a high accuracy sub-pixel matching result by applying a bilateral filter to the cost aggregation. Though this method raised the accuracy of computing disparity, it computed different parameters for each pair of images and cost a lot of computation space and time. In recent years, some method using plain fitting and image segmentation managed to solve this problem [6]. A method proposed by Bleyer [7] divided image into several regions by using graphic-cut and then built models in different regions, which got a very good result in sub-pixel image matching, but this method also needed a lot of computation space and time.

Geiger et al. [8] proposed a method combine local and global method, which used a set of support points to build triangle planes and computed disparity based on Bayesian. The method has high accuracy and is fast, however, likes other local methods, it generates "stair-casing effect" and has lower accuracy when dealing with slant plane, curved surface and weak texture region. In order to eliminate "stair-casing effect", a fast sub-pixel accuracy stereo matching method is presented in this paper based on disparity plane which is improved from the method of Geiger [8]. A prior edge disparity is built in our method by use of Canny descriptor to preserve disparity discontinuities and computes sub-pixel disparity from disparity of the support points and the triangulation disparity plane.

2. METHODOLOGY

The framework of our method is provided in Figure 1. First, a set of support points are extracted from the original matching images. Among them some points are obtained from the description images which are calculated from the original matching images by Sobel descriptor, the others are obtained from the edge of the original matching images by Canny edge detector. Second, all the support points are used to create the triangle disparity planes by Delaunay triangulation. In each disparity plane, each point can obtain a disparity by the parameters of the Delaunay triangulation disparity plane. Finally, the sub-pixel disparity map is obtained by a proportion computed from disparity of the Delaunay triangulation disparity plane and the distance between the point and the triangle plane on the image. In our approach, we assume that the left and right images have been rectified.

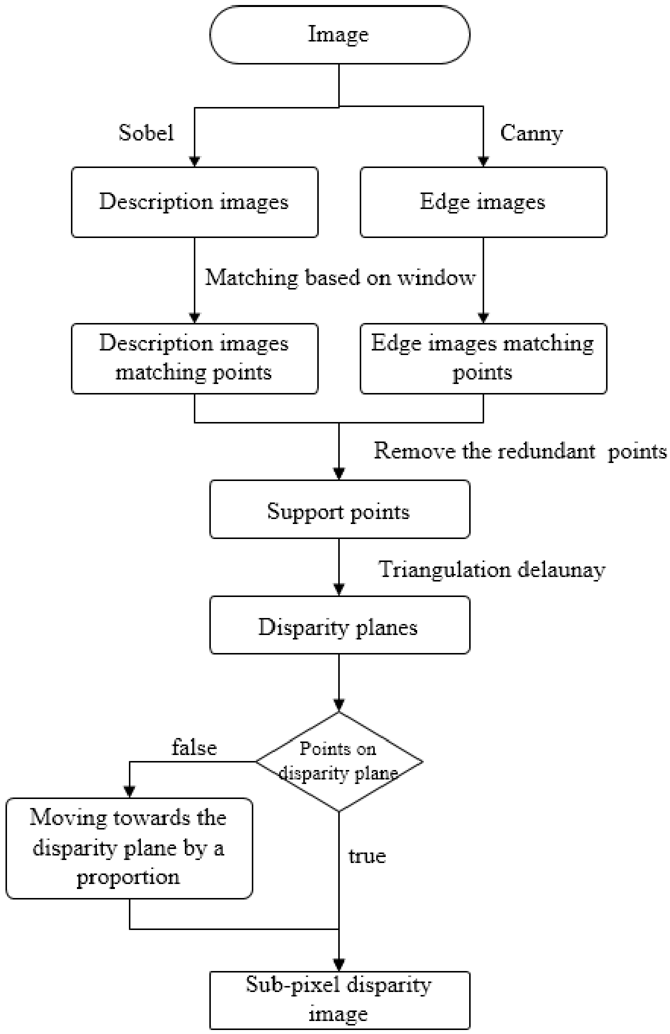

Figure 1. Flow chart of our approach.

2.1 Support points extracting

As it is shown in figure 2, a set of support points, which can be highly matched from left and right images, are obtained by description images and edge images. The number of support points has a great influence on the accuracy of matching. To get high accuracy support points, the horizontal and vertical Sobel filter is used to generate the description images. The description image I_{des} is a 16-channel image, each point of it is made up of each 8 points nearby in horizontal Sobel filter images I_{HSobel} and vertical Sobel filter images I_{VSobel} .

$$I_{des}\left(u,v\right)=\left\{I_{HSobel}\left(u-1,v-1\right),\cdots,I_{HSobel}\left(u+1,v+1\right),I_{VSobel}\left(u-1,v-1\right),\cdots,I_{VSobel}\left(u+1,v+1\right)\right\} \quad (1)$$

Part of support points can be obtained by matching the description images from left-to-right and right-to-left to make sure the points have exact correspondences between the images. The corresponding matching point in the right description is given by calculating the minimum of D_{des} .

$$D_{des} = \sqrt{\sum_{i=1}^{15} \left[I_{desl}\left(u,v\right) - I_{desr}\left(u+i,v\right) \right]^2} \tag{2}$$

The support points obtained by the above method have isolated points which have dissimilar disparity with all the support points around. These isolated points have to be removed based on the assumption that the disparity field is continuous.

To preserve disparity discontinuities, edge points are used as support points which are obtained by detecting the original images by Canny edge detector. The histogram of image is used to get the best parameters for Canny edge detector. To compute the disparity of the whole image, the points at the image corners are added to the support points.

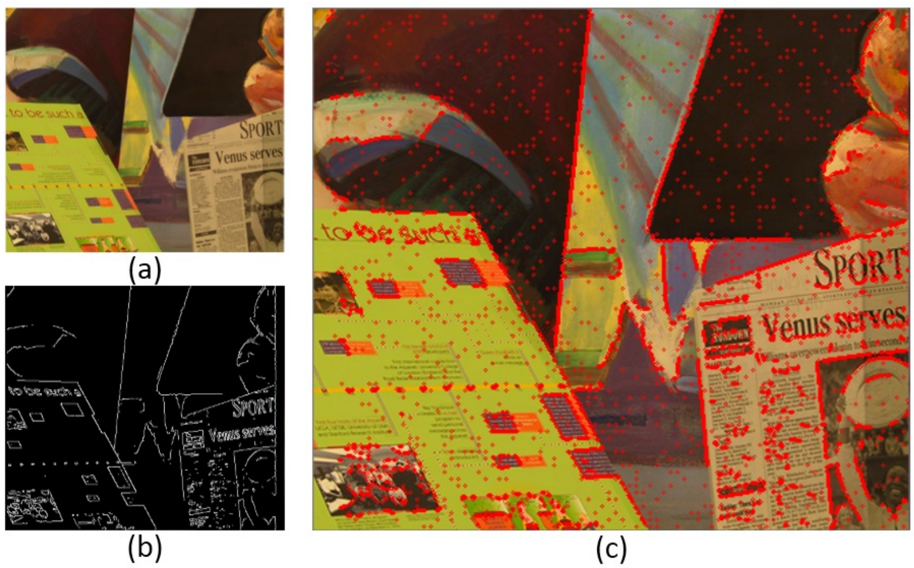

Figure 2. Support points result on Middlebury datasets. (a)Original left image. (b) Edge image using Canny edge detector. (c) All the support points shown on the original left image.

2.2 Sub-pixel Disparity estimation with disparity plane

To get the disparities of the points remained, the support points are used to create the disparity planes based on Delaunay triangulation to build prior information. As figure 3 shows, the image is divided into a number of triangular area, each area corresponds to a disparity plane which can be get with each three non-linear points which is defined as α_d :

$$\alpha_d = au_i + bv_i + c \tag{3}$$

For each point P_o inside the triangular area tri_o on the image, the disparity of it is confined to the corresponding plane where $d\left(u_{P_o}, v_{P_o}\right) \in \left(\min \alpha_d\left(u_{tri_o}, v_{tri_o}\right), \max \alpha_d\left(u_{tri_o}, v_{tri_o}\right)\right)$.

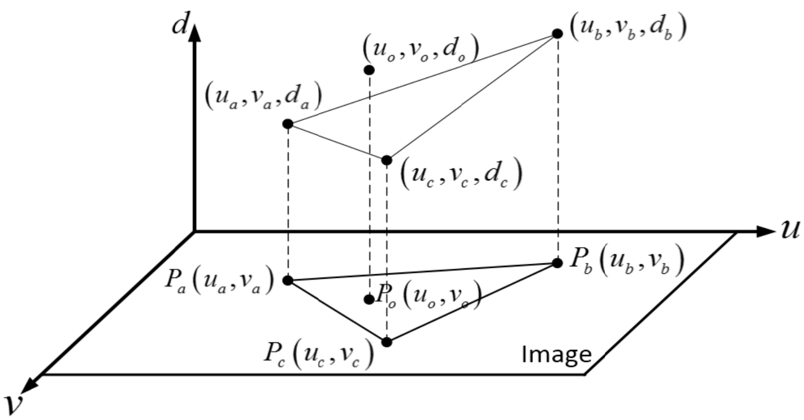

Figure 3. Sketch showing the correspondence between the image and the disparity plane. P_a, P_b, P_c are three support points, P_o is an unmatched point inside the disparity plane.

After the establishment of the disparity planes, it can be seen that the search range for unmatched points has been limited into a triangular area which make computing dense matching at a short time scales. Different from traditional dense local matching, our method use the maximum a-posteriori estimation (MAP) to compute the disparity instead of using the minimum matching cost with winner-takes-all (WTA) decision criteria. To compute the disparity of target point $p_l(u,v)$ in the left image, inspired by Geiger's method[8], a set of support points which around the target point $p_l(u,v)$ in a 17*17 pixel support window are used to build a prior disparity space $\mu(u,v,d)=\{\mu_1(u_1,v_1,d_1),\cdots,\mu_n(u_n,v_n,d_n)\}$. The disparity of target point $p_l(u,v)$ is computed using the maximum a-posteriori estimation as:

$$d = \arg\max P(d \mid p_l, p_r, \mu) \tag{4}$$

The image likelihood is defined by description image I_{des}, which obeys a Gaussian distribution as:

$$P(p_r \mid p_l, d) \propto \lambda_d + \exp\left[-\frac{\| I_{des}(u,v) \|_1}{2\sigma^2}\right] \tag{5}$$

λ_d is a constant. The conditional probability of the disparity d is described as:

$$P(d \mid p_l, p_r, \mu) \propto P(d \mid p_l, \mu) P(p_r \mid p_l, d) \tag{6}$$

To simplify the calculation procedure, some support points are removed where the disparity is dissimilar with the value of disparity plane. Based on Markov random fields, the posterior probability distribution can be described as $P(d \mid I) = \exp(-E(d))$ .By minimizing $E(d)$ [9], an initial disparity map can be obtained.

2.3 Sub-pixel Disparity Refinement

To eliminate the "stair-casing effect" caused by fronto-parallel matching windows, the disparity plane is used to refine the disparity of each point inside. As figure 4 shows, in the disparity coordinate system $O-uvd$, for point $P_o(u_o,v_o,d_o)$ with integer disparity value, the distance d_1 between $P_o(u_o,v_o,d_o)$ and the disparity plane α_d can be obtained with

Eq.(3) that:

$$d_1 = \frac{\left|au_o + bv_o - d_o + c\right|}{\sqrt{a^2 + b^2 + 1}} \tag{7}$$

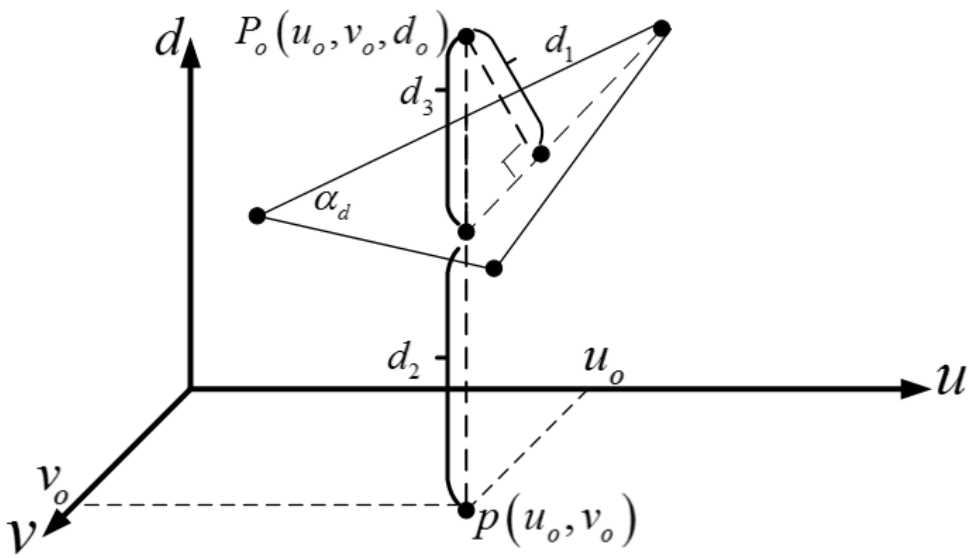

Figure 4 The disparity coordinate system of disparity refinement.

Point $P_2(u_o, v_o, d_2)$ is computed by projecting the point $P_o(u_o, v_o, d_o)$ to the disparity plane α_d ,the value of d_2 is as follows:

$$d_2 = au_0 + bv_o + c \tag{8}$$

The distance from $P_o(u_o, v_o, d_o)$ to $P_2(u_o, v_o, d_2)$ can be represented as $d_3 = d_o - d_2$. Our method adopts Gaussian probability model to depict the correlation between point $P_o(u_o, v_o, d_o)$ and the disparity plane α_d :

$$p(d_o, \alpha_d) = \begin{cases} \gamma + \exp\left(-\frac{\left(\left(d_o - \alpha_d(P_o) \right) - d_1 \right)^2}{2\sigma^2} \right) & if \left| \left(d_o - \alpha_d(P_o) \right) - d_1 \right| < 3\sigma \\ 0 & otherwise \end{cases} \tag{9}$$

with γ an additional probability .Our method gain accuracy by excluding the disparity farther than 3σ from d_1 .Then the sub-pixel disparity map can be computed by moving the point towards the disparity plane with:

$$d_o^* = d_3 \times p\left(d_o, \alpha_d\right) + d_2 \quad (10)$$

At last, our method uses a left and right consistency check and a median filter to eliminate spurious mismatches.

3. EXPERIMENT

In this section, a few of experiments are performed to evaluate the quality of the proposed approach. Since our goal is to obtain the accurate sub-pixel disparity map, the method is applied on real and benchmark datasets in the experiments to verify the accuracy. First, to show the ability of eliminating the "stair-casing effect", a stereo image and a dense disparity image are generated with real scenes compared to ELAS. Second, to evaluate our method on more scenes, pairs of stereo images which contain large texture-less region are used to generate the three-dimensional surface model. Then, Middlebury Stereo Dataset is used to generate disparity images to compare the accuracy and running time with state-of-the-art methods. Besides, a set of different resolution images are used to show the efficient processing capacity of our method. In all the experiments, the main parameters of our method are set to $\gamma = 0.05$, $\sigma = 1$ and $\lambda_d = 0.1$ in the experiment.

The improvement is showed with a set of high resolution reality scene images of 1024*1024. The comparison results are shown in Figure 5. The point cloud image shows obviously that our method have a good effect on eliminating the "stair-casing effect".

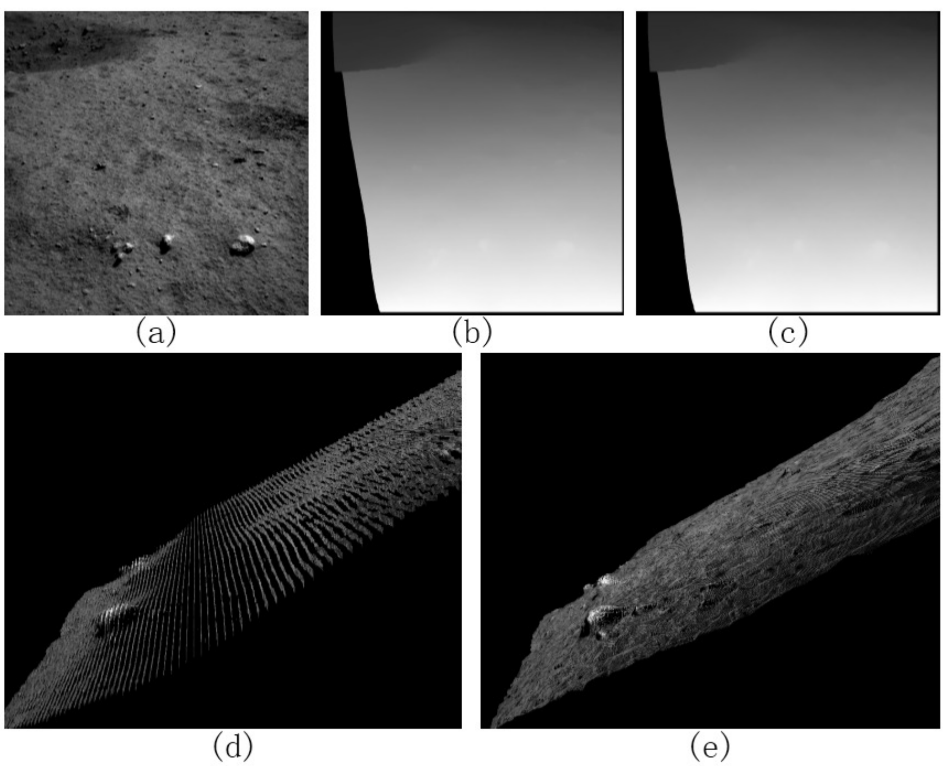

Figure 5. Contrast result of reality scene image between ELAS and our method. (a) Reality scene images. (b) Depth image of ELAS. (c) Depth image of our method. (d) Dense point cloud of ELAS. (e) Dense point cloud of our method.

To show the processing power, experiments which are made on reality scene in different situation are shown in Figure 6. Figure 6(a ,b) contain large texture-less region , Figure 6(c ,d) is a pair of images with a wide baseline. As it is shown in Figure 6(e ,f), our method's performance remains constant in such situation. We also compare the accuracy and running time of our approach with AdaptWeight and FastBilateral in stereo images from the Middlebury data sets: Venus, Teddy and Cone. The results are shown in Figure 7 and Table 1 shows the accuracy and the running time comparison. The result shows our method preform competitively to these classic matching method but with such short running time.

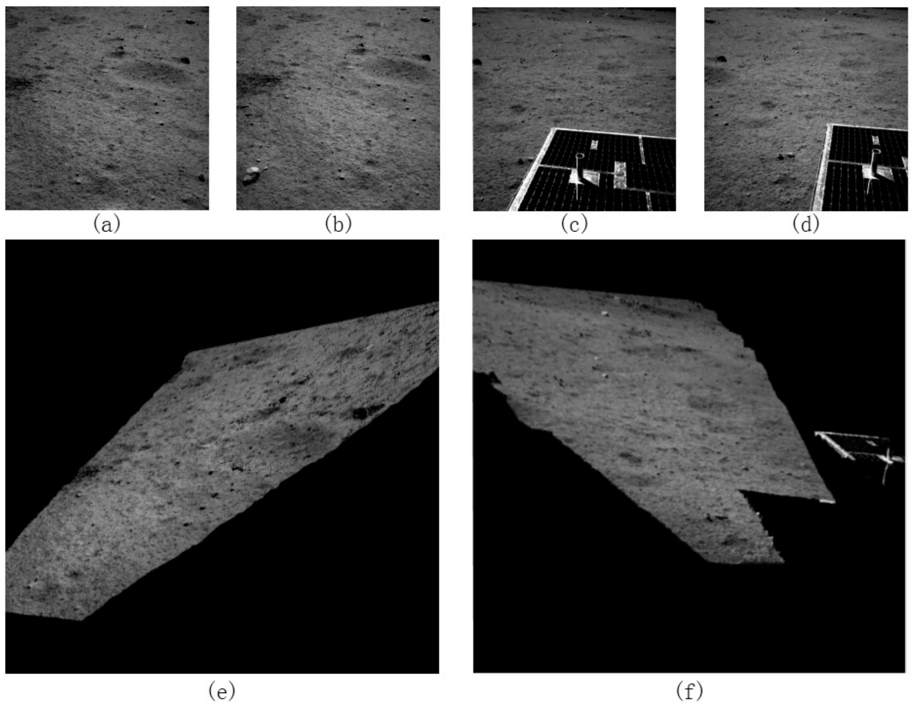

Figure 6. Performance of our method in reality scene.

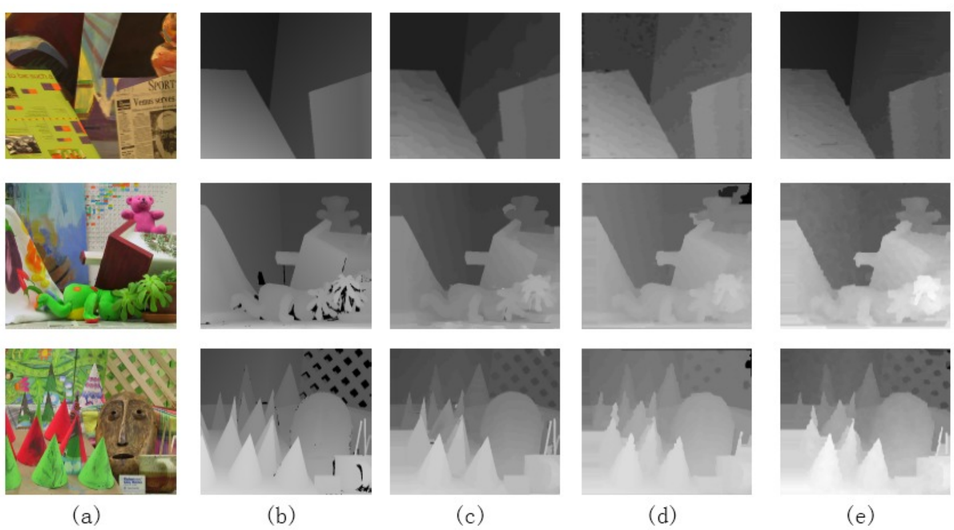

Figure 7 Results on Middle bury datasets. (a) Middlebury test images. (b) Ground truth disparities. (c) Using AdaptWeight. (d) Using ELAS. (e) Using our method.

Table 1. Comparison of the algorithms for error threshold 1.

	Venus	Teddy	Cone	Average time
AdaptWeight	1.19	13.3	9.79	5 min 35 s
FastBilateral	0.92	15.3	9.31	32 s
ELAS	1.42	13.6	13.3	<1s
Our method	1.26	12.8	7.17	<0.6s

We also evaluate the running time for different resolutions image from 0.26 Megapixel (512*512) to 4 Megapixels (2048*2048) on the reality images, we find out that the running time increase from 0.5s to 2.7s, which is very fast.

4. CONCLUSION AND FUTURE WORK

In this paper, a fast sub-pixel stereo matching algorithm is presented to eliminating the “stair-casing effect” when dealing with slant plane, curved surface and weak texture region. To achieve the effect of fast matching, a disparity plane is used to limit the search range inside a triangle area. And, by establishing the Gaussian probability model, the accurate sub-pixel disparity can be calculated for each point inside the disparity plane. As it is shown in the experiments, the result on eliminating the “stair-casing effect” is very obvious in the three-dimensional reconstruction figure. And compared with some classic methods, our method has shown the high accuracy and it takes 0.6s on average to process one Megapixel image which is very fast. To achieve faster speed of 3D reconstruction, in the future, the GPU will be used to accelerate the method to meet the requirement of processing a pair of one Megapixel images in real time.

ACKNOWLEDGMENTS

This work is supported by the National Natural Science Foundation of China under Grant No. 51105027.

REFERENCES

[1] Tian Q., and Huhns M. N., “Algorithms for subpixel registration,” Computer Vision, Graphics, and Image Processing 35(2), 220-233 (1986).

[2] Nehab D., Rusinkiewiez S., and Davis J., "Improved sub-pixel stereo correspondences through symmetric refinement," Computer Vision, 2005. ICCV 2005. Tenth IEEE International Conference, 1, 557-563 (2005).

[3] Takita K., Sasaki Y., Higuchi T. et al., “High-accuracy subpixel image registration based on phase-only correlation,” IEICE transactions on fundamentals of electronics, communications and computer sciences 86(8), 1925-1934 (2003).

[4] Shibahara T., Aoki T., Nakajima H. et al., "A sub-pixel stereo correspondence technique based on 1D phase-only correlation," Image Processing, 2007. ICIP 2007. IEEE International Conference, 5, 221-224 (2007).

[5] Yang Q., Yang R., Davis J. et al., "Spatial-depth super resolution for range images," Computer Vision and Pattern Recognition, 2007. CVPR'07. IEEE Conference, 1-8 (2007).

[6] Yang Q., "A non-local cost aggregation method for stereo matching," Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference, 1402-1409 (2012).

[7] Bleyer M., Rother C., and Kohli P., "Surface stereo with soft segmentation," Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference, 1570-1577 (2010).

[8] Geiger A., Roser M., and Urtasun R.,"Efficient large-scale stereo matching," Springer, (2011).

[9] Donate A., Wang Y., Liu X. et al., "Efficient and accurate subpixel path based stereo matching," Pattern Recognition, 2008. ICPR 2008. 19th International Conference, 1-4 (2008).