Robust view selection in multiview stereo

Zheng Gu (顾 征)[1,2], Xianyu Su (苏显渝)[1*], and Junbo Yang (杨俊波)[1]

[1] *Department of Optoelectronics, Sichuan University, Chengdu 610064*

[2] *China Academy of Space Technology, Beijing 100094*

[*] *E-mail: xysu@email.scu.edu.cn*

Received March 6, 2008

A robust view selection algorithm for multiview stereo matching is presented. Different from the existent view selection algorithms which pick only the believable views for matching, this method assigns an adaptive weight to each target view based on the dissimilarity between it and the reference view. So it can utilize the information of all views more sufficiently. The algorithm has been evaluated with different real images to demonstrate its robustness.

OCIS codes: 100.0100, 100.5010, 150.0150, 150.6910.

doi: 10.3788/COL20090703.0198.

Multiview stereo is an important part of computer vision[1,2]. Differing with binocular stereo, there are three or more images which can be used for matching in multiview vision[3−7]. So how to use the information of different target images appropriately is very important for handling occlusion and improving matching precision. Kang *et al.* proposed a temporal selection method[8] to handle occlusion. Rather than summing the matching errors over all the frames, they picked only the views where the pixels are visible. On the basis of temporal selection, Sun *et al.* proposed an adaptive convolution kernel[9] to reach the balance between keeping more information and reducing ambiguities. Their methods look more efficient in occlusion handling than temporal selection. By studying their methods, we can find that for each pixel's matching, they first divide all target views into two parts: believable views and unbelievable views. The views which have lowest dissimilarities with the reference view are picked as believable views and the other views are rejected as unbelievable views. So for each pixel, only a part of frames which are regarded as believable views can be used for its matching. If the pixel is in the occlusion region, this operation is appropriate. But if the pixel is visible in all views, using the methods in Refs. [8] and [9] will lose some useful information because some useable views are rejected as unbelievable views due to their little bigger dissimilarities. So both algorithms can handle occlusion in multiview stereo, but they are weak in some regions which are visible in all the views. Therefore, we propose a new view selection method. Instead of picking some believable views, the method assigns an adaptive weight for each view based on the dissimilarity between it and the reference view.

In a multiview stereo, the observation is a collection of images $\{I_k,\ k = 0, \cdots, K\}$. It is assumed that I_r is the reference view, $F(s, d_s, I_r, I_k)$ is the matching cost function of pixel s with disparity d_s between I_r and I_k. In our implementation, we use Birchfield and Tomasi's pixel dissimilarity measure (PDM)[10]. So we can define two kinds of weights for a certain target view I_k:

$$w_{ak} = \frac{\exp(-\alpha^{-1}F(s, d_s, I_r, I_k))}{\sum\limits_{i=0}^{r-1}\exp(-\alpha^{-1}F(s, d_s, I_r, I_i)) + \sum\limits_{i=r+1}^{K}\exp(-\alpha^{-1}F(s, d_s, I_r, I_i))}, \tag{1}$$

where α is parameter determined empirically, and

$$w_{sk} = \frac{\exp(-\alpha^{-1}F_s(s, d_s, I_r, I_k))}{\sum\limits_{i=0}^{r-1}\exp(-\alpha^{-1}F_s(s, d_s, I_r, I_i)) + \sum\limits_{i=r+1}^{K}\exp(-\alpha^{-1}F_s(s, d_s, I_r, I_i))}, \tag{2}$$

where $F_s(s, d_s, I_r, I_k) = \frac{\sum_{d_s} F(s,d_s,I_r,I_k)}{N_d}$, N_d is the number of disparity d in predefined search range. Like Ref. [9], to reach the balance between keeping more information and reducing ambiguities, we also define an adaptive weight as

$$w_k = \begin{cases} w_{ak} & \min\{F^-, F^+\} \geq t\max\{F^-, F^+\} \\ w_{sk} & \min\{F^-, F^+\} < t\max\{F^-, F^+\} \end{cases}, \tag{3}$$

where $F^- = \sum\limits_{k=0}^{r-1} F_s(s, d_s, I_r, I_k)$ and $F^+ = \sum\limits_{k=r+1}^{K} F_s(s, d_s, I_r, I_k)$, t $(0 < t < 1)$ is a winner threshold. If there is an obvious winner between F^- and F^+, w_{sk} is selected to reduce the ambiguity. Otherwise, w_{ak} is picked to keep more information.

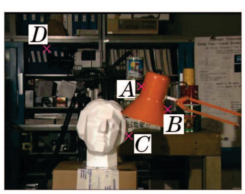
So the final matching cost $P(s, d_s)$ of the pixel s with disparity d_s can be expressed as

$$P(s,d_s)=\sum_{k=0}^{r-1} w_k F(s,d_s,I_r,I_k) + \sum_{k=r+1}^{K} w_k F(s,d_s,I_r,I_k). \quad (4)$$

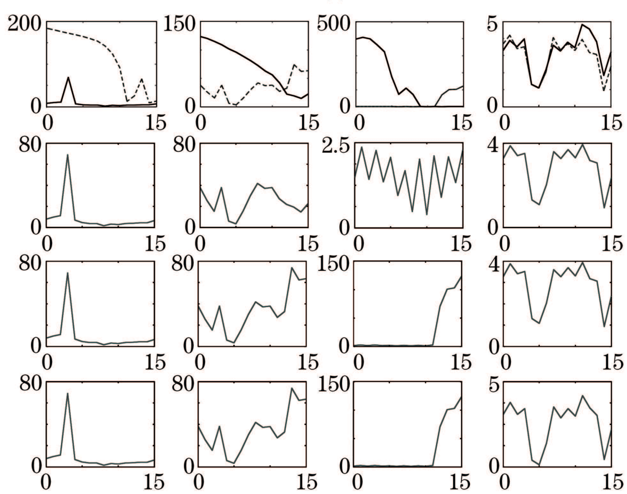
From the above description, we can see that if the pixel under matching is in the occlusion region, our method will assign a negligible weight to the view where the pixel is invisible. So only the views where the pixel is visible will play a crucial role in stereo matching. Furthermore, if the pixel under matching is visible in all views, the proposed method will assign an almost equal weight to each view. Then all views can be utilized to produce a reliable matching result.

We use the image sequences of Tsukuba (2nd, 3rd, and 4th frames), Venus (1st, 3rd, and 5th frames), Moebius (1st, 2nd, and 3rd frames), and Reindeer (1st, 2nd, and 3rd frames) as our test data. They can be downloaded from Middlebury website[11]. When comparing our method with other two methods, all modules of stereo matching algorithm are uniform except the view selection module (adaptive support-weight[12] for window selection, PDM for dissimilarity measure, and winner-takes-all (WTA) for disparity selection). The parameters set across both sequences are $\alpha = 5$ and $t = 0.5$.

Figure 1(a) is the reference image of the Tsukuba sequence. Points A, B, C, and D are four typical points in stereo matching. Points A, B, and C are semiocclusion points and point D is visible in all views. Each column of Fig. 1(b) corresponds to A, B, C, and D, respectively. The first row shows the distribution of matching cost, where the horizontal and vertical coordinates are disparity and matching cost $F(s,d_s,I_r,I_k)$, the dotted line is the dissimilarity $F(s,d_s,I_r,I_{r-1})$ between the left view and the reference view and the solid line is the dissimilarity $F(s,d_s,I_r,I_{r+1})$ between the right view and the reference view. The second, third, and fourth rows of Fig. 1(b) show the distribution of final matching costs which are computed by temporal selection[8], Sun's method[9], and our method, respectively, the horizontal and vertical coordinates are disparity and final matching cost $P(s,d_s)$.



(a)



(b)

Fig. 1. Comparison of three methods with Tsukuba sequence. (a) Reference image; (b) comparison of three methods.

Table 1. Disparities of Points A, B, C, and D

	A	B	C	D
True Disparity	8	5	8	5
Temporal Selection	8	5	10	14
Sun's Method	8	5	8	14
Our Method	8	5	8	5

Table 1 shows the true disparities of the points A, B, C, and D and the disparities worked out by the WTA strategy based on the final matching costs in Fig. 1(b). From Table 1 and Fig. 1(b), we can find that for points A and B, all three methods can acquire right disparities. For the point C, Sun's method and our method can obtain a right disparity by using an adaptive strategy as Eq. (3) while the temporal selection method fails. For the point D which is visible in all views, temporal selection and Sun's method are invalid to acquire a right disparity because of their insufficient usage of all target views' information. Our method, however, also produces a right result because it can use the information of all views more sufficiently.

Figure 2 shows the reference image and matching results of the Venus sequence. When comparing the results are obtained by temporal selection, Sun's method, and our method, we find that fewer than 10% pixels in the disparity maps change. However, many regions in Figs. 2(b) and (c) (some of them are marked by squares) where inaccurate results are produced because of the insufficient usage of all views' information are meliorated by our method, as shown in Fig. 2(d).

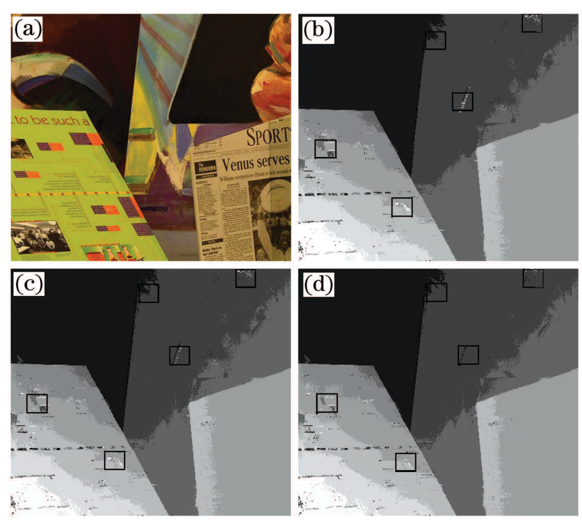


Fig. 2. Reference image and matching results of Venus sequence. (a) Reference image; (b) result by temporal selection; (c) result by Sun's method; (d) our result.

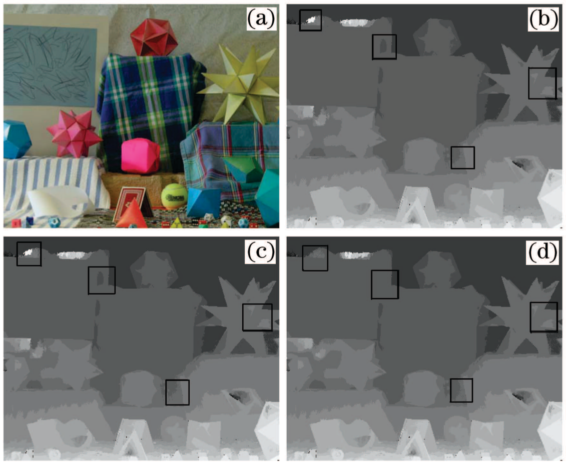


Fig. 3. Reference image and matching results of Moebius sequence. (a) Reference image; (b) result by temporal selection; (c) result by Sun's method; (d) our result.

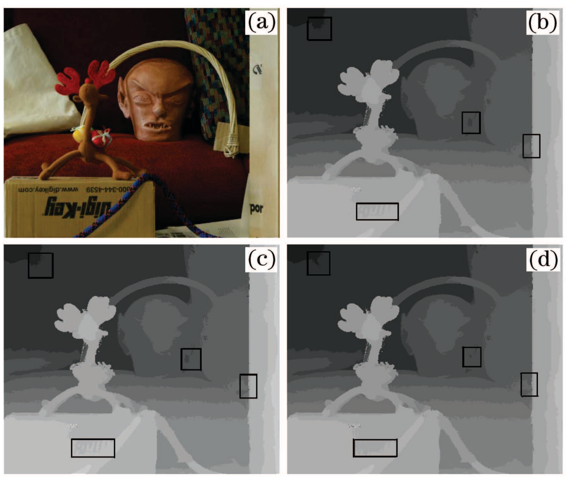


Fig. 4. Reference image and matching results of Reindeer sequence. (a) Reference image; (b) result by temporal selection; (c) result by Sun's method; (d) our result.

The matching results of Meobius and Reindeer sequences are shown in Figs. 3 and 4, respectively. Similar to the result of the Venus sequence, some inaccurate pixels in temporal selection and Sun's method are revised with our method. Some of them are also marked by squares.

In conclusion, a new view selection method which assigns an adaptive weight for each view according to the dissimilarity between it and the reference view has been proposed. Experimental results demonstrate that the method can utilize the information of all views more sufficiently. It is robust for view selection.

This work was supported by the National Natural Science Foundation of China under Grant No. 60527001.

References

1. H. Li and Q. Zhang, Acta Opt. Sin. (in Chinese) **27,** 907 (2007).
2. Y. Guan, Chin. Opt. Lett. **6,** 465 (2008).
3. S. M. Seitz, B. Curless, J. Diebel, D. Scharstein, and R. Szeliski, in *Proceedings of the 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition* **1,** 519 (2006).
4. T. Ueshiba, in *Proceedings of 18th International Conference on Pattern Recognition* **1,** 1076 (2006).
5. G. Vogiatzis, P. H. S. Torr, and R. Cipolla. in *Proceedings of IEEE Computer Society Conference on Pattern Recognition 2005* **2,** 391 (2005).
6. E. S. Larsen, P. Mordohai, M. Pollefeys, and H. Fuchs, in *Proceedings of the Third International Symposium on 3D Data Processing, Visualization, and Transmission* 342 (2006).
7. M.-A. Drouin, M. Trudeau, and S. Roy, in *Proceedings of the Fifth International Conference on 3D Digital Imaging and Modeling* 540 (2005).
8. S. B. Kang, R. Szeliski, and J. Chai, in *Proceedings of 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition* **1,** 103 (2001).
9. J. Sun, N.-N. Zheng, and H.-Y. Shum, IEEE Trans. Pattern Anal. Mach. Intell. **25,** 787 (2003).
10. S. Birchfield and C. Tomasi, IEEE Trans. Pattern Anal. Mach. Intell. **20,** 401 (1998).
11. D. Scharstein and R. Szeliski, Int. J. Comput. Vis. **47,** 7 (2002).
12. K.-J. Yoon and I. S. Kweon, IEEE Trans. Pattern Anal. Mach. Intell. **28,** 650 (2006).