A stereo matching algorithm using multi-peak candidate matches and geometric constraints

Yepeng Guan (管业鹏)[1,2]

[1]*School of Communication and Information Engineering, Shanghai University, Shanghai 200072*

[2]*Key Laboratory of Advanced Displays and System Application, Ministry of Education, Shanghai University, Shanghai 200072*

Received September 10, 2007

Gray cross correlation matching technique is adopted to extract candidate matches with gray cross correlation coefficients less than some certain range of maximal correlation coefficient called multi-peak candidate matches. Multi-peak candidates are extracted corresponding to three closest feature points at first. The corresponding multi-peak candidate matches are used to construct the model polygon. Correspondence is determined based on the local geometric relations between the three feature points and the multi-peak candidates. The disparity test and the global consistency checkout are applied to eliminate the remaining ambiguous matches that are not removed by the local geometric relational test. Experimental results show that the proposed algorithm is feasible and accurate.

OCIS codes: 150.0150, 100.6890, 100.5010, 100.2960.

Stereovision is a common method for extracting depth information from intensity images. In this method, a pair of images are acquired using two cameras separated from each other. After determining the stereo correspondence, the distance among various points is computed using triangulation[1]. Stereo correspondence is the problem of finding points in two or more images of the same scene, usually assuming known camera geometries. However, a critical issue in stereovision is to find corresponding pixels, points, or other features in both stereo images taken from two cameras. Stereo image matching is the first and most difficult step in recovering three-dimensional (3D) information from a pair of stereo images. It is also a fundamental task for many applications such as robot navigation and industrial automation[2].

Various computational algorithms with certain constraints and matching strategies have been proposed to reduce possibility of false matches, but many problems still remain in stereo correspondence[3,4]. Conventional image matching techniques may be classified as area-based or feature-based[5]. Feature-based matching is typically more reliable than area-based matching because the features are more stable to photometric variations and accurate than area-based schemes for the types of used features typically could be located to sub-pixel precision. In addition, feature-based matching is more efficient because it can produce global matching results and thus efficiently avoid searching blindly in a wide range[6−9].

Typical image features include the point, line, curve segment, and region, etc. Among them, points do not require a special pre-processing step, and many structured features are not prevalent in all images[10]. In this paper, a stereo algorithm based on feature points is proposed. This algorithm uses gray correlation matching technique to extract multi-peak feature points with correlation coefficients less than certain range of maximal correlation coefficient, which is called multi-peak candidate matches. A single point does offer any information about the pattern structure. However, if the point is considered in the context of the other points of the pattern, some useful information related with the pattern structure in a certain region could be drawn[11]. There exists already some published work on using geometric constraints for matching. Hu and Ahuja have proposed a matching algorithm involving geometry, rigidity, and disparity constraints[12]. Wong and Chung have proposed a method to refine correspondences by fitting an active contour model to the transferred feature points on the scene view[13]. Gupta and Mittal have proposed an affine invariant point matching using ordinal features to refine correspondences[14]. Gong and Yang have proposed an unambiguous stereo matching based on reliability measure[15]. Other researchers have recommended some geometric matching algorithms also.

The geometric stereo algorithm proposed in this paper differs from other existing geometric matching algorithms. We utilize the local geometric relations among the feature points to obtain valid matches from the multi-peak candidate matches. The disparity test and the global consistency checkout are applied to eliminate the remaining ambiguous matches that are not removed by the local geometric relational test.

As we know, unique correct correspondence cannot be obtained only by the use of gray cross correlation between the left and right images. Take a pair of actual parallel stereo images for example. Horizontal pixel coordinates of seven pairs of feature points in left and right images, respectively, are shown in Table 1.

Based on the directivity of stereo imaging, we adopt gray correlation matching technique to determine matches of two feature points in the left image, $x = 423$ and 586 respectively, across the right image. Suppose that the deviation towards the left is 300 pixels, namely the search scopes are $[123 - 423]$ and $[286 - 586]$ in the right image, respectively. The maps of gray correlation

Table 1. Horizontal Coordinates for Stereo Images (pixels)

Left Image	343	423	504	586	668	754	839
Right Image	205	268	331	396	462	529	598

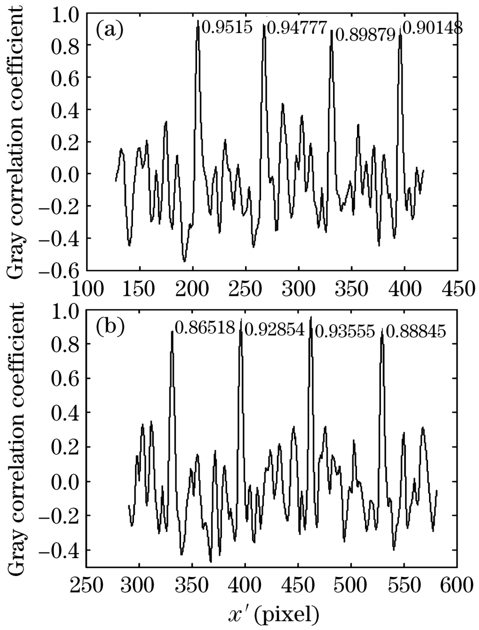

Fig. 1. Gray correlation between two feature points in the left image and in the right image. (a) $x = 423$; (b) $x = 586$.

are shown in Fig. 1. If the criterion of maximum gray correlation is adopted, it is easy to deduce that the corresponding coordinates would be $x' = 205$ and $x' = 462$ from Figs. 1(a) and (b), respectively, by maximum gray correlation coefficient. In fact, however, the actual coordinates should be $x' = 268$ and $x' = 396$.

Since gray cross correlation unilaterally describes gray similarity of feature points region between the left and right images, false matches will happen inevitably[16]. Gray cross correlation, however, can be adopted to determine gray correlation coefficients C_{ij} between the feature points in the left and right images. For confirming candidate matches in the next step, multi-peak feature points with normalized gray cross correlation coefficients satisfying $C_{ij} \geq k \cdot \max C$ (where $\max C = \max(C_{ij})$, k is a fraction less than 1 called the peak value ratio) are taken as the candidate matches, which are called multi-peak candidate matches.

Due to the fact that if the point is considered in the context of the other points of the pattern, some useful information related to the pattern structure in a certain region could be drawn[10]. In this paper, utilizing gray cross correlation, we extract three groups of multi-peak candidate matches corresponding with its closest three feature points respectively in the left image. Based on the local geometric relations between the three groups of multi-peak candidates and the three feature points, we determine correspondences of the three feature points from the multi-peak candidates. We use the corresponding multi-peak candidate matches to construct a model polygon. The cross ratio of point array is invariant after the projection transform[17,18], which is testified as follows.

Suppose that A, B, C, and D are four different points on line L, their $\mathbf{N}$ vectors are $\mathbf{m}_A$, $\mathbf{m}_B$, $\mathbf{m}_C$, and $\mathbf{m}_D$, the $\mathbf{N}$ vectors of their corresponding points A', B', C', and D' after projection transform F are $\mathbf{m}_{A'}$, $\mathbf{m}_{B'}$, $\mathbf{m}_{C'}$, and $\mathbf{m}_{D'}$,

$$\mathbf{m}_{A'} = \gamma_A F^{\mathrm{T}}\mathbf{m}_A, \quad \mathbf{m}_{B'} = \gamma_B F^{\mathrm{T}}\mathbf{m}_B,$$
$$\mathbf{m}_{C'} = \gamma_C F^{\mathrm{T}}\mathbf{m}_C, \quad \mathbf{m}_{D'} = \gamma_D F^{\mathrm{T}}\mathbf{m}_D, \tag{1}$$

where γ_A, γ_B, γ_C, and γ_D are constants which convert the vector into unit vector.

Suppose that the $\mathbf{N}$ vector of the line L is $\mathbf{n}$, and after projection F, it is transformed as

$$\mathbf{n}' = \gamma F^{-1}\mathbf{n}, \tag{2}$$

where γ is a constant.

Let $\boldsymbol{\nu}$ to be an arbitrary vector $(\boldsymbol{\nu}, \mathbf{n}) \neq 0$,

$$\boldsymbol{\nu}' = F^{\mathrm{T}}\boldsymbol{\nu}, \tag{3}$$

$$(\boldsymbol{\nu}', \mathbf{n}') = (F^{\mathrm{T}}\boldsymbol{\nu}, \gamma F^{-1}\mathbf{n}) = \gamma(\boldsymbol{\nu}, \mathbf{n}) \neq 0. \tag{4}$$

The cross ratio of A, B, C, and D is

$$R(A,B,C,D) = \frac{|\mathbf{m}_A, \mathbf{m}_C, \boldsymbol{\nu}|}{|\mathbf{m}_B, \mathbf{m}_C, \boldsymbol{\nu}|} : \frac{|\mathbf{m}_A, \mathbf{m}_D, \boldsymbol{\nu}|}{|\mathbf{m}_B, \mathbf{m}_D, \boldsymbol{\nu}|}. \tag{5}$$

The cross ratio of A', B', C', and D' is

$$R(A',B',C',D') = \frac{|\mathbf{m}_{A'}, \mathbf{m}_{C'}, \boldsymbol{\nu}|}{|\mathbf{m}_{B'}, \mathbf{m}_{C'}, \boldsymbol{\nu}|} : \frac{|\mathbf{m}_{A'}, \mathbf{m}_{D'}, \boldsymbol{\nu}|}{|\mathbf{m}_{B'}, \mathbf{m}_{D'}, \boldsymbol{\nu}|}, \tag{6}$$

$$\begin{aligned}
|\mathbf{m}_{A'}, \mathbf{m}_{C'}, \boldsymbol{\nu}| &= \gamma_A\gamma_C \left|F^{\mathrm{T}}\mathbf{m}_A, F^{\mathrm{T}}\mathbf{m}_C, F^{\mathrm{T}}\boldsymbol{\nu}\right| \\
&= \gamma_A\gamma_C \det F^{\mathrm{T}} \left|\mathbf{m}_A, \mathbf{m}_C, \boldsymbol{\nu}\right|, \\
|\mathbf{m}_{B'}, \mathbf{m}_{C'}, \boldsymbol{\nu}| &= \gamma_B\gamma_C \left|F^{\mathrm{T}}\mathbf{m}_B, F^{\mathrm{T}}\mathbf{m}_C, F^{\mathrm{T}}\boldsymbol{\nu}\right| \\
&= \gamma_B\gamma_C \det F^{\mathrm{T}} \left|\mathbf{m}_B, \mathbf{m}_C, \boldsymbol{\nu}\right|, \\
|\mathbf{m}_{A'}, \mathbf{m}_{D'}, \boldsymbol{\nu}| &= \gamma_A\gamma_D \left|F^{\mathrm{T}}\mathbf{m}_A, F^{\mathrm{T}}\mathbf{m}_D, F^{\mathrm{T}}\boldsymbol{\nu}\right| \\
&= \gamma_A\gamma_D \det F^{\mathrm{T}} \left|\mathbf{m}_A, \mathbf{m}_D, \boldsymbol{\nu}\right|, \\
|\mathbf{m}_{B'}, \mathbf{m}_{D'}, \boldsymbol{\nu}| &= \gamma_B\gamma_D \left|F^{\mathrm{T}}\mathbf{m}_B, F^{\mathrm{T}}\mathbf{m}_D, F^{\mathrm{T}}\boldsymbol{\nu}\right| \\
&= \gamma_B\gamma_D \det F^{\mathrm{T}} \left|\mathbf{m}_B, \mathbf{m}_D, \boldsymbol{\nu}\right|.
\end{aligned} \tag{7}$$

Since γ_A, γ_B, γ_C, γ_D and $\det F^{\mathrm{T}}$ are constants, Eq. (5) can be obtained by combining Eqs. (6) and (7), which means that $R(A,B,C,D) = R(A',B',C',D')$.

We compute the similarity between the triangles, which is examined by computing the proportions of the lengths of the corresponding sides of the triangles[11] (see Fig. 2). Suppose that the feature point p in the left image has multi-peak candidates (r_1, r_2) in the right image, and its closest two neighbor points q and t have corresponding multi-peak candidate matches (w_1, w_2), (s_1, s_2), respectively. L_i $(i = 1,2,3)$ is the length among p, q, and t. LC_i $(i = 1,2,3)$ are the lengths among (r_1, r_2), (q_1, q_2), and (s_1, s_2). Computing the lengths L_i and LC_i, based on the local geometric relations between the three feature points and their corresponding multi-peak candidates, we determine the correspondences of the three feature points from the multi-peak candidate matches by local geometric similarity.

Suppose that the matrix FL consists of lengths L_i $(i = 1,2,3)$,

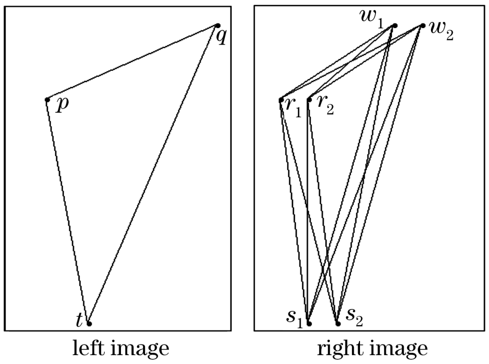

Fig. 2. Computation of similarity.

$$FL = [\ L_1 \quad L_2 \quad L_3\ ], \tag{8}$$

where

$$L_1 = \overline{pq}, \quad L_2 = \overline{qt}, \quad L_3 = \overline{tp}. \tag{9}$$

FL is a constant, which is completely determined by the three known points.

Suppose that the matrix ML consists of lengths LC_i $(i = 1, 2, 3)$,

$$ML = [\ LC_1 \quad LC_2 \quad LC_3\ ], \tag{10}$$

where

$$LC_1 = \overline{r_i w_j}, \quad LC_2 = \overline{w_i s_j}, \quad LC_3 = \overline{s_i r_j}, \quad (i = 1, 2; j = 1, 2). \tag{11}$$

If more than three points are used in the local geometric relational test, it may improve the quality of matched points, or eliminate some of the good matches and need more computation[11,12].

In the above discussion, the nearest neighbors are defined in terms of image plane distances. Perspective distortion could be a serious problem when the neighboring points in the image plane have large differences. To reduce this effect, the closest neighboring points could be required to have nearly the same disparities. Besides, the above local geometric relation tests are based on two-dimensional (2D) similarity and may hence result in errors; in other words, we still cannot guarantee that all valid matches are correct by local geometric relational test alone. The disparity test and global consistency checkout below are applied to eliminate the remaining ambiguous matches that are not removed by the local geometric relational test. The basic idea is to check if any point of matched pair in the right image can be re-matched by another point across the left image on the epipolar line. In Fig. 3, $P_{\rm L}$ and $P_{\rm R}$ represent points along two epipolar lines in the left and right image planes, respectively. The direction of two arrows on the top is consistent and it is one-to-one mapping, namely the point $P_{\rm L1}$ corresponding to the point $P_{\rm R1}$ is a pair of correct match. The arrow from the point $P_{\rm L3}$ in the left image is matched to the point $P_{\rm R2}$ in the right image plane, and $P_{\rm R2}$ is matched to the point P_{L5}. This is a mismatch. Consider a set of match points $p_{\rm l}(x, y)$ in the left image with the corresponding point $p_{\rm r}(x', y')$ in the right image. The disparity between these points is the displacement vector between the two points and displacements in x and y directions are given by ${\rm d}x_i = x_i - x_i'$ and ${\rm d}y_i = y_i - y_i'$, respectively. The disparity vectors for currently matched pairs $p_{\rm l}(x, y)$ and $p_{\rm r}(x', y')$ from left to right are $d_i = ({\rm d}x_i, {\rm d}y_i) = (x_i - x_i', y_i - y_i')$, and those from right to left are $d_j = ({\rm d}x_j, {\rm d}y_j) = (x_j - x_j', y_j - y_j')$. Assuming that the vertical and horizontal disparities should be the same for any correctly matched feature points, we simply repeat this operation for all candidate matches. If some points are re-matched by other points, we need to update the peak value ratio parameter k iteratively. The resulting disparities obtained from the new parameter k are checked as mentioned above repeatedly until the feature points in the left (or right) image are uniquely matched.

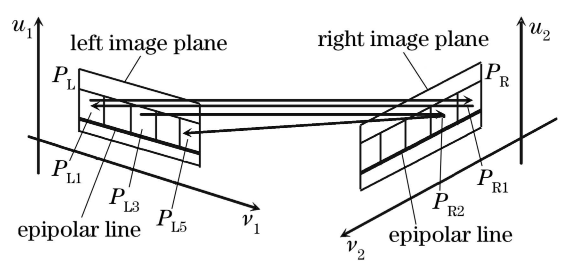

Fig. 3. Correspondences in the left and right image planes.

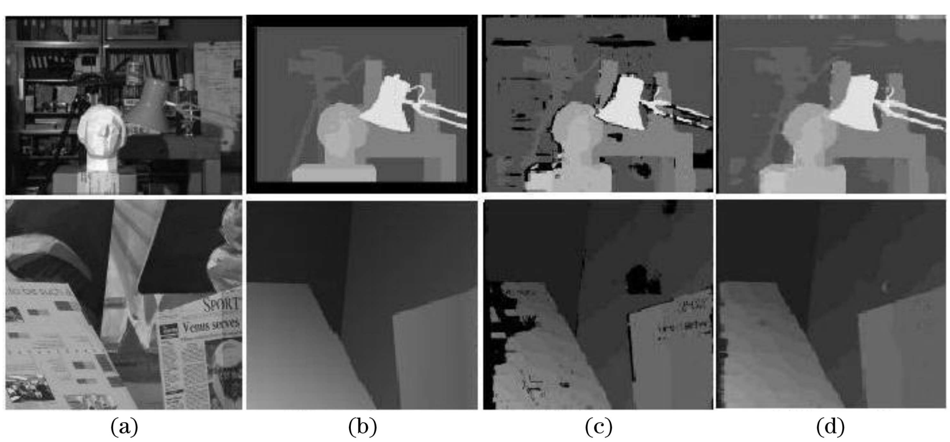

Fig. 4. Results for the Tsukuba and Venus images. (a) Left images; (b) ground truth; (c) before checkout; (d) after checkout.

Table 2. Performance of the Proposed Approach Shown with Error Statistics

Algorithm	Tsukuba			Venus		
	Nonocc	All	Disc	Nonocc	All	Disc
Ding[6]	5.23	7.14	12.81	4.27	5.64	10.71
Wong[13]	1.89	2.01	9.35	1.34	1.51	2.65
Gong[15]	1.53	1.82	7.62	1.17	1.25	2.36
Proposed	1.57	1.86	7.54	1.12	1.21	2.25

Nonocc: all pixels in Nonoccluded regions; All: all pixels in regions without texture; Disc: all pixels near discontinuities.

To verify the effectiveness of the proposed approaches, we performed experiments with some color images taken from the standard test images with ground truth from http://vision.middlebury.edu/stereo, which are often used for performance comparison of various methods. The results of stereo method for the tested images are given in Fig. 4. As shown in Fig. 4, the proposed method yields satisfying results for the tested images.

In the following parts, we apply a quantitative technique to evaluate the performance of the proposed algorithm for the above two images according to the method proposed in Ref. [19]. The performance of the proposed method for the tested images is compared with other methods in Table 2. Table 2 summarizes error statistics for each test case gathered within three areas: all pixels in nonoccluded regions, all pixels in regions without texture, and all pixels near discontinuities. It is clearly seen that the proposed approach performs better for the tested images.

Gray cross correlation matching technique can be adopted to extract multi-peak candidate matches with gray cross correlation coefficients less than some certain range of maximal correlation coefficient. Utilizing the corresponding multi-peak candidate matches to construct model polygon, we can efficiently avoid searching blindly in a wide range. Based on the local geometric relations between three feature points and their corresponding multi-peak candidates, we can determine correspondences of the three feature points from the multi-peak candidate matches. Combining disparity test and global consistency checkout, we can get valid correct matches.

This work was supported by the Leading Academic Discipline Project of Shanghai Educational Committee of China (J50104) and the Shanghai Leading Academic Disciplines of China (T0102). Y. Guan's e-mail address is ypguan@shu.edu.cn.

References

1. J. Jeon, K. Kim, C. Kim, and Y.-S. Ho, in *Proceedings of 2001 IEEE Pacific Rim Conference on Communications, Computers and Signal Processing* **1,** 263 (2001).
2. Y. Lu, K. Kubik, and M. Bennamoun, in *Proceedings of IEEE on Speech and Image Technologies for Computing and Telecommunications* **1,** 315 (1997).
3. A. M. Felicisimo and A. Cuartero, IEEE Trans. Geosci. Remote Sensing **44,** 2534 (2006).
4. G.-J. Wen and X.-Z. Zhou, Journal of Software (in Chinese) **16,** 708 (2005).
5. G. Li and Y. He, in *Proceedings of 2002 IEEE International Symposium on Circuits and Systems* **2,** 277 (2002).
6. H. Ding and M.-Y. Fu, Trans. Beijing Inst. Technol. (in Chinese) **26,** 987 (2006).
7. Z. Gu and X. Su, Opto-Electron. Eng. (in Chinese) **34,** (1) 95 (2007).
8. K.-J. Yoon and I. S. Kweon, in *Proceedings of IEEE 11th International Conference on Computer Vision* 4409002 (2007).
9. H. Li and Q. Zhang, Acta Opt. Sin. (in Chinese) **27,** 907 (2007).
10. M. S. Lew, T. S. Huang, and W. Kam, IEEE Trans. Pattern Anal. Machine Intell. **16,** 869 (1994).
11. M. Tico, C. Rusu, and P. Kuosmanen, in *Proceedings of International Conference on Image Processing* **2,** 462 (1999).
12. X. Hu and N. Ahuja, IEEE Trans. Pattern Anal. Machine Intell. **16,** 1041 (1994).
13. H.-S. Wong and R. Chung, in *Proceedings of 6th International Conference on Signal Processing* **2,** 929 (2002).
14. R. Gupta and A. Mittal, in *Proceedings of IEEE 11th International Conference on Computer Vision* 4409088 (2007).
15. M. Gong and Y.-H. Yang, IEEE Trans. Pattern Anal. Machine Intell. **27,** 998 (2005).
16. Y. Guan, L. Tong, W. Gu, J. Liu, and X. Ye, Chin. J. Electron. (in Chinese) **13,** 501 (2004).
17. S. Ma and Z.Zhang, *Computer Vision — Fundamentals of Computational Theory and Algorithms* (in Chinese) (Science press, Beijing, 2003) p.37.
18. Z. Zhang, G. Chen, M. Gao, and C. Song, J. Eng. Graphics (in Chinese) **28,** (3) 102 (2007).
19. D. Scharstein and R. Szeliski, Int. J. Computer Vision **47,** 7 (2002).