Available online at www.sciencedirect.com
ScienceDirect
Chinese Journal of Aeronautics 22(2009) 70-74
Chinese Journal of Aeronautics
www.elsevier.com/locate/cja

Robust Wide Baseline Point Matching Based on Scale Invariant Feature Descriptor

Yue Sicong*, Wang Qing, Zhao Rongchun

School of Computer Science and Engineering, Northwestern Polytechnical University, Xi'an 710072, China

Received 21 March 2008; accepted 10 October 2008

Abstract

In order to obtain a large number of correct matches with high accuracy, this article proposes a robust wide baseline point matching method, which is based on Scott's proximity matrix and uses the scale invariant feature transform (SIFT). First, the distance between SIFT features is included in the equations of the proximity matrix to measure the similarity between two feature points; then the normalized cross correlation (NCC) used in Scott's method, which has been modified with adaptive scale and orientation, is used to put more weight on the correct pairs in the matrix since the SIFT feature is only invariant to linear changes of light. The proposed method removes all the proximity information about the distance between feature points' locations in the Scott's method, which causes mismatch in wide baseline matching. Experimental results show that the proposed algorithm is invariant to changes of scale, rotation, light, and thereby provides a new effective way for wide baseline matching.

Keywords: computer vision; image analysis; image match; scale invariant feature descriptor

1. Introduction

A key problem of computer vision[1] is to establish correspondence between two images because many complex tasks use the point matching as an initial procedure.

Many ways[2-8] have been proposed for the point matching. The classical ones assume a short baseline and are usually based on the correlation method. It is well known that the correlation method is unstable because it suffers from the changes of rotation, zoom, and affine transformation. As such, an elegant approach emerges from the spectral graph theory. G. Scott and H. Longuet-Higgins[9] were among the first researchers who used this theory.

According to Scott's matching method, the proximity matrix based on the distance between the feature points' locations was used. However, this measure led to bad performance. M. Pilu[10] improved Scott's method by using both the proximity information and the similarity information. The proximity measure was the same as in Scott's method, whereas the similarity was computed as the normalized cross correlation (NCC) between the points. Experiments show that the performance of Pilu's modification declined with the growing of baseline.

This article proposes a robust matching algorithm to deal with the correspondence problem of widebaseline. First, the distance between the scale invariant feature transform (SIFT) features[11] is used to build up the correspondence matrix. Then, NCC is modified with adaptive scale and orientation (denoted by ASO-NCC). This procedure results in a significant improvement in correspondence matrix. Finally, experiments on the different types of wide baseline are carried out to judge the performance of the proposed algorithm.

The ensuing part of the article is organized as follows. Section 2 describes the SIFT feature. Section 3 summarizes and explains the proposed robust wide baseline matching algorithm. Sections 4-5 display experimental results and conclusions, respectively.

2. Scale Invariant Feature Descriptor

In order to perform efficient matching of images with wide baseline, the stable feature descriptor is required invariant to the changes of rotation, scale, and illumination. This article first defines an image as a set of SIFT features, each representing a vector of local

*Corresponding author. Tel.: +86-13572137676.
E-mail address: yuesicong@mail.nwpu.edu.cn
Foundation items: National High-tech Research and Development Program (2007AA01Z314); National Natural Science Foundation of China (60873085)

doi: 10.1016/S1000-9361(08)60070-9

image measurements which is invariant to the changes of translation, scale, and rotation. However, the SIFT feature is partially invariant to the changes in illumination and affine deformations. The local and multiscale natures of the SIFT feature make them insensitive to noise, clutter, and occlusion, so they are highly selective for matching to large amounts of features. A comparative study on the effectiveness of the various invariant feature descriptors was presented in Mikolajczyk's work[12], which shows the SIFT feature that creates better performances than other existing approaches.

The feature points are localized directly on a scale space structure by searching the local maxima of difference of Gaussian (DoG) both on space and scale. At each point, an orientation is selected at the peak of the histogram of local image gradient orientations. A feature vector is formed by measuring the local image gradients in a region relative to the location, scale, and orientation of the point. The gradients are further blurred to reduce sensitivity to noise and small local image deformations. In summary, the SIFT feature renders local image features stable across multiple views of a scene.

3. Robust Wide Baseline Matching

This section will discuss the proposed robust matching algorithm. To begin with, some brief descriptions about the work on point matching with spectral graph analysis should be provided. Spectral graph analysis aims at characterizing the global properties of a graph using the eigenvalues and the eigenvectors. Most of these contributions are based on the so-called proximity or affinity matrix. The matrix elements are weights that reflect the strength of a pair relation. Usually, the proximity matrix $\boldsymbol{G}$ is defined as[9]

$$G_{ij} = \mathrm{e}^{-r_{ij}^2/2\sigma^2} \tag{1}$$

where σ is a free parameter, and r_{ij} the distance between the point $p_i = I_1(x_i, y_i)$ and the point $q_j = I_2(x_j, y_j)$. The element G_{ij} decreases monotonically from 1 to 0 with the increase of r_{ij}. G. Scott and H. Longuet-Higgins[9] were among the first to use spectral methods for the image matching. They use the Euclidean distance between two points to measure the relation of a pair. In Scott's algorithm, corresponding points must be the closest. For this reason, M. Pilu[10] added the normal NCC to quantify the similarity. The elements of $\boldsymbol{G}$ that model proximity and similarity can then be transformed into

$$G_{ij} = \frac{C_{ij}+1}{2}\mathrm{e}^{-r_{ij}^2/2\sigma^2} \tag{2}$$

where C_{ij} is the NCC between two points which ranges from −1 to 1. Stipulating the similarity constraint can eliminate rogue features, which should not be similar to anything, but the performance of Pilu's method declines with the baseline growing.

The reason of the worsening performance is to blame for the existing methods adopting improper feature and measure in the proximity matrix. On one hand, the distance between feature points' locations is unreliable for wide baseline because feature points may be detected or not from different views. On the other hand, the windows of fixed size and fixed orientation are used for the normal NCC without taking the changes of scale and orientation into account. Such an NCC can only tackle the situation with translation between two images. Evidently, the performance of the existing methods will aggravate when the baseline grows.

Considering the aforesaid problems, it is clear that more robust feature and measure must be adopted. First, the distance between SIFT features is used to reflect the strength of a pair relation, which means all the proximity information about the distance between points' locations are discarded. Then, ASO-NCC is used to put more weight on the correct pairs in the matrix, which can also eliminate the effects of illumination changes because the SIFT feature is only invariant to linear brightness changes.

Both the size and the orientation of the correlation windows are determined according to the characteristic scale and orientation of each feature point. Let $p_i = I_1(x_i, y_i)$ be a feature point with scale s_1 and orientation θ_1 in image I_1 and $q_j = I_2(x_j, y_j)$ be a feature point with scale s_2 and orientation θ_2 in the other image I_2. Without losing generality, $s_1 > s_2$ can be assumed. $\boldsymbol{A}_i$ and $\boldsymbol{B}_j$ are two correlation windows of size $(2w+1)\times(2w+1)$ centered on each point with $w = \lambda s_2$, where λ is a constant ($\lambda = 5$ in this article). Let $\theta = \theta_2 - \theta_1$ be the rotation angle and $\lambda = s_1/s_2$ be the scale change factor. Two correlation windows $\boldsymbol{A}_i$ and $\boldsymbol{B}_j$ can be represented by

$$\left.\begin{aligned} A_i^{uv} &= I_1(x_i + ru\cos\theta + rv\sin\theta, \\ &\qquad y_i + rv\cos\theta - ru\sin\theta) \\ B_j^{uv} &= I_2(x_j + u, y_j + v) \end{aligned}\right\} \tag{3}$$

where $u, v \in [-w, w]$. A^{uv} is calculated using bilinear interpolation. The ASO-NCC is then defined as

$$C_{ij} = \frac{\sum_{u=-w}^{w}\sum_{v=-w}^{w}(A_i^{uv} - \mathrm{mean}(\boldsymbol{A}_i))(B_j^{uv} - \mathrm{mean}(\boldsymbol{B}_j))}{(2w+1)(2w+1)\mathrm{stdv}(\boldsymbol{A}_i)\mathrm{stdv}(\boldsymbol{B}_j)} \tag{4}$$

where $\mathrm{mean}(\boldsymbol{A})$ is the average of subimage $\boldsymbol{A}$ and $\mathrm{stdv}(\boldsymbol{A})$ the standard deviation of subimage $\boldsymbol{A}$.

The next step of the proposed algorithm is to carry out the singular value decomposition of $\boldsymbol{G}$:

$$G = UDV^{\mathrm{T}} \tag{5}$$

Then compute a new pairing matrix $\boldsymbol{P}$ by converting the diagonal matrix $\boldsymbol{D}$ into a diagonal matrix $\boldsymbol{E}$, where each D_{ii} is replaced by 1.

$$P = UEV^{\mathrm{T}} \tag{6}$$

$\boldsymbol{P}$ has the same shape as $\boldsymbol{G}$, and $\boldsymbol{P}$ has the attractive property of amplifying good pairs and attenuating bad ones. Two features $\boldsymbol{A}_i$ and $\boldsymbol{B}_j$ are paired up if P_{ij} is the greatest element in both row i and column j. No threshold is required for choosing good pairs.

By adopting the SIFT feature and proposed ASO-NCC, the correspondence matrix $\boldsymbol{G}$ becomes more robust to severe scene variance than Scott's matrix. Finally, a reasonably good solution can be achieved through the singular value decomposition followed by a simple manipulation of the eigenvalues.

4. Experimental Results

To verify the proposed algorithm, experiments have been carried out on several types of image sequences with it to compare with Pilu's algorithm[10] (improved Scott's algorithm) and the SIFT distance matcher in Ref.[12]. The test image sequences include different geometric and photometric transformations, such as large rotation and scale variance, viewpoint changes, and illumination contrast. The variance grows in accordance with increase of the index of frames. Fig.1 shows some examples of test images. All the image sequences are from INRIA (http://lear.inrialpes.fr/people/Miko-lajczyk/Database/index.html).

In all cases, the first frame (index "0") of a sequence is fixed as the reference image and matched to the other frames. The method used for determining the correct matches depends on what geometric information on the camera geometry is available. When a set of data consists of fixed camera or sequences of planar scenes for which the homography between the different views are available, a pair of corresponding points $(\boldsymbol{p}, \boldsymbol{p}')$ could be considered a correct match if

$$\|\boldsymbol{p}' - \boldsymbol{H}\boldsymbol{p}\| < 3 \tag{7}$$

where $\boldsymbol{H}$ is the homography between the two images. For the sequence with camera's translation or viewpoint changes, epipolar geometry constraint is used to select the correct matches. In this case, because the scene is not planar, the fundamental matrix $\boldsymbol{F}$ is computed, and a pair of corresponding points $(\boldsymbol{p}, \boldsymbol{p}')$ is a correct match if

$$d(\boldsymbol{p}', \boldsymbol{F}\boldsymbol{p}) + d(\boldsymbol{p}, \boldsymbol{F}^{\mathrm{T}}\boldsymbol{p}') < 3 \tag{8}$$

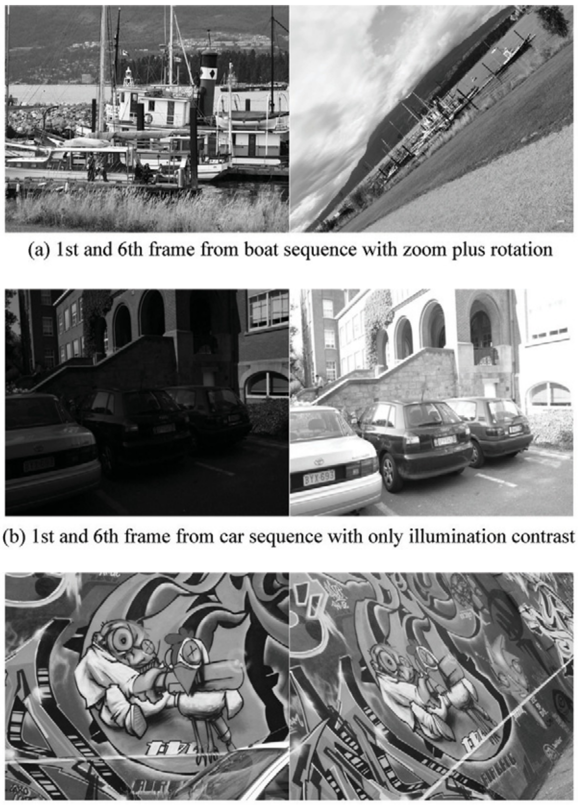
where $d(\boldsymbol{p}', \boldsymbol{F}\boldsymbol{p})$ is the distance from the point $\boldsymbol{p}'$ to the epipolar line corresponding to the point.

Let PM denote the Pilu's matching method, SDM the SIFT distance matcher, and RWM the proposed robust wide baseline matching method.

The results in Fig.2 are pertinent to a boat sequence of 6 frames with changes of different degrees in rotation and scale. Fig.3 shows the results of a car sequence of 6 frames with different light contrast. Table 1 displays the results of graffiti sequence with viewpoint changes. Since the results of the last 2 frames are close to 0 for three methods, Table 1 reveals only the matching results of 2nd, 3rd, and 4th frames. It is found that the accuracy of RWM and SDM are quite close, both much higher than that of PM for large rotation, scale, and viewpoint changes. For large illumination contrast, the accuracy of RWM and PM are roughly identical but that of SDM shows evident drops.

Apart from accuracy, the number of correct matches is also of great importance. This is because many matching methods with satisfied accuracy cannot provide adequate correct matches required by some applications in the case of wide baseline. In the experiment, there are four types of wide baseline changes in the test image sequences. In all cases, the proposed algorithm RWM yields the largest number of correct match. Especially, for the last few frames with variance of large degree, RWM can offer several times correct matches more than PM and SDM. In the largest scene variance pair of boat sequence or graffiti sequence, SDM and PM only afford 4-6 correct matches, whereas RWM 25-50 correct pairs.

For boat and graffiti sequences, SDM works better



(a) 1st and 6th frame from boat sequence with zoom plus rotation

(b) 1st and 6th frame from car sequence with only illumination contrast

(c) 1st and 4th frame from graffiti sequence with viewpoint change

Fig.1 Some examples of test image sequences.

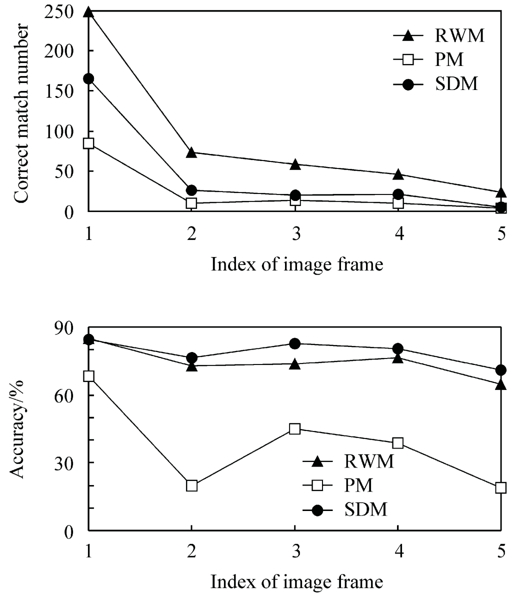


Fig.2 Curves of correct match number vs accuracy for three algorithms: comparison of results for boat sequence with scale and rotation changes.

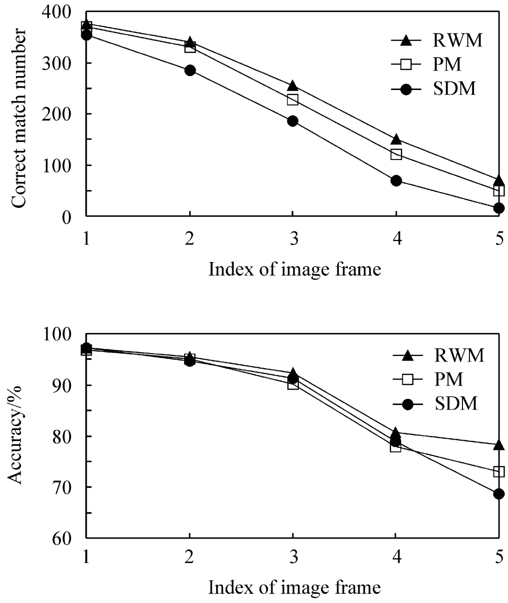


Fig.3 Curves of correct match number vs accuracy for three algorithms: comparison of results for car sequence with only illumination contrast.

Table 1 Correct matches and accuracy for three algorithms on a graffiti sequence

Method	Index of frame=1			Index of frame=2			Index of frame=3		
	RWM	PM	SDM	RWM	PM	SDM	RWM	PM	SDM
Correct match number	679	406	575	302	64	187	45	5	5
Accuracy/%	92	88	95	76	37	65	43	8	45

than PM, which is ascribed to the higher robustness of SIFT feature to scale, rotation, and viewpoint changes than that of NCC. PM does better than SDM in car sequence with only illumination changes, which shows that NCC is more robust to general light changes than SIFT features. Fig.4 and Fig.5 show some examples of matching maps.

(a) RWM

(b) PM

(c) SDM

Fig.4 Example of matching results for three methods (the pair of images is the 1st and 6th frame from boat sequence with large scale and rotation changes).

(a) RWM

(b) PM

(c) SDM

Fig.5 Example of matching results for three algorithms (the pair of images is the 1st and 4th frame from graffiti sequence with large viewpoint changes).

5. Conclusions

This article presents a new robust automatic matching approach for wide baseline based on spectral graph theory. By replacing the unreliable distance between points' locations by the distance between the SIFT features and adopting ASO-NCC, the similarity constraint of the optimal matches in the correspondence matrix is enhanced to satisfy changes of wide baseline, which has the advantage of reducing the risk of mismatches over existing methods. Experimental results show that the proposed algorithm can yield larger number of correct matches and higher accuracy than Pilu's matching method (improved Scott's method) and the SIFT distance matcher do for changes of scale, rotation, illumination, and viewpoint. Moreover, the proposed algorithm avoids choosing the threshold parameter for determining the optimal matches. It provides an easy and effective way to wide baseline matching for many applications in computer vision.

References

[1] Marr D. Early processing of visual information. Philosophical Transactions of the Royal Society of London (Biological Sciences) 1976; 275(942): 483-519.

[2] Zitov B, Flusser J. Image registration methods: a survey. Image and Vision Computing 2003; 21(11): 977-1000.

[3] Lu X Y, Manduchi R. Wide baseline feature matching using the cross epipolar ordering constraint. IEEE Proceedings of CVPR. 2004; 16-23.

[4] Pritchett P, Zisserman A. Wide baseline stereo matching. IEEE Proceedings of the ICCV. 1998; 754-760.

[5] Zhang Z, Deriche R, Faugeras O, et al. A robust technique for matching two uncalibrated images through the recovery of the unknown epipolar geometry. Artificial Intelligence 1995; 78(1-2): 87-119.

[6] Ferrari V, Tuytelaars T, Gool L V. Wide-baseline multiple-view correspondences. IEEE Proceedings of CVPR. 2003; 718-725.

[7] Carcassoni M, Hancock E. Spectral correspondence for point pattern matching. Pattern Recognition 2003; 36(1): 193-204.

[8] Sclaroff S, Pentland A. Modal matching for correspondence and recognition. IEEE Transactions on PAMI 1995; 17(6): 545-561.

[9] Scott G, Longuet-Higgins H. An algorithm for associating the features of two patterns. Proceedings of the Royal Society of London, Series B, Biological Sciences 1991; 244: 21-26.

[10] Pilu M. A direct method for stereo correspondence based on singular value decomposition. IEEE Proceedings of CVPR. 1997; 261-266.

[11] Lowe D. Distinctive image features from scale invariant key-points. International Journal of Computer Vision 2004; 60(2): 91-110.

[12] Mikolajczyk K, Schmid C. A performance evaluation of local descriptors. IEEE Transactions on PAMI 2005; 27(19): 1615-1630.

Biography:

Yue Sicong Born in 1979, he received B.S. and M.S. from Northwestern Polytechnical University in 2002 and 2005, respectively. Now, he is a Ph.D. candidate in Northwestern Polytechnical University. His main research interests lie in image processing, computer vision, and pattern recognition.
E-mail: yuesicong@mail.nwpu.edu.cn