Discarding wide baseline mismatches with global and local transformation consistency

H.B. Zhou, D.Z. Zhang, C. Chen and J.W. Tian

A novel method called global and local transformation consistency constraints, which combines the scale, orientation and spatial layout information of ‘scale invariant feature transform’ (SIFT) features, is proposed for discarding mismatches from given putative point correspondences. Experiments show that the proposed method can efficiently extract high-precision matches from low-precision putative SIFT matches for wide baseline image pairs, and outperforms or performs close to state-of-the-art approaches.

Introduction: Local features are powerful tools for finding correspondences between wide baseline views of the same scenes. Some feature-based algorithms first establish putative correspondences, and then estimate the best global geometry relationship (such as homography) interpreting them. However, many well-known robust estimators (such as RANSAC [1]) perform poorly when the ratio of inliers is lower than 50% [2], while discarding mismatches before estimating this relationship yields important improvements, especially in the case where incorrect matches strongly outnumber the correct ones. Previous works (see e.g. [3, 4]) for discarding mismatches mainly employ the geometrical and topological relationship among putative matches, but ignore the scale and orientation information of the potential feature pairs, which can express a similarity transformation.

This Letter focuses on rejecting mismatches via evaluating the quality of each potential correspondence, which is measured by both global and local transformation consistency. To address the mismatches discarding problem, we divide the algorithm into two steps. First, using global constraint, we will be retaining a part of the matches of which the scale log-ratio and orientation difference are approximate to global scaling and rotation factor, respectively. Then, using local constraint, we will reject more incorrect matches from the first step, with a stricter constraint by requiring that neighbouring feature pairs have the similar transformation. Experiments show that the approach presented in this Letter improves the currently achieved wide baseline matching precision, with 10% fewer errors on most of the six well-known wide baseline image pairs, which were offered by Tuytellaars and Van Gool [5].

Global scaling and rotation consistency: Mikolajczyk and Schmid compared the performance of local features in [6], and the result showed that Lowe’s SIFT [2] performs best. A SIFT feature comprises four components: location p, scale σ, orientation θ, and feature vector f. This quaternion $Q(p,\sigma,\theta,f)$ contains a wealth of information. However, a popular approach in most SIFT matching algorithms is discarding the information about scale, orientation and spatial layout of features, and finding corresponds using appearance only. As the first step of this Letter, the idea is to use the scale and orientation information of SIFT features to reinforce the global scaling and rotation consistency, and to weed out a part of the false positive matches.

As mentioned above, if two features are a correct match, their scale log-ratio and orientation difference should be limited in scope, and approximate to global scaling and rotation factor, respectively. Let Q_1 $(p_1, \sigma_1, \theta_1, f_1)$ and Q_2 $(p_2, \sigma_2, \theta_2, f_2)$ denote a putative SIFT matched pair between image I_1 and I_2, where the dominant orientation (θ) should be modified into $[0, \pi]$ according to contrast insensitive criterion. Scale log-ratio $(\Delta\sigma)$ and orientation difference $(\Delta\theta)$ are given by:

$$\Delta\sigma = \log_2(\sigma_1/\sigma_2) \tag{1}$$

$$\Delta\theta = \theta_1 - \theta_2 \tag{2}$$

Meanwhile, the correct matches should satisfy these following constraints:

$$|\Delta\sigma - \overline{\Delta\sigma}| < \tau_\sigma \tag{3}$$

$$|\Delta\theta - \overline{\Delta\theta}| < \tau_\theta \tag{4}$$

First, the scale log-ratio histogram and orientation difference histogram of all putative matches are formed, and the values of scaling factor $(\overline{\Delta\sigma})$ and rotation factor $(\overline{\Delta\theta})$ are assigned by the values corresponding to the highest peaks of their histograms, respectively. τ_σ and τ_θ are two thresholds to reject outliers. The correct ratio improves after this step, but it is still too low for applying as a global geometry relationship estimator. Hence, the following local similarity transformation consistency constraint has been developed.

Local similarity transformation consistency: A simple global scaling and rotation consistency test discards lots of the false matches, and then a stricter local transformation consistency approach is used to increase more evidence for surviving matches. The idea of this step is inspired by spatial consistency voting proposed by Sivic and Zisserman [7], however Sivic and Zisserman’s method is applying to visual retrieval, and its constraint is too loose for wide baseline matching. In this Letter, the score of each match is co-determined by the qualities of matches of spatially neighbouring features. Let m: $(p, p', \Delta\sigma, \Delta\theta, N_{pp'})$ be a match of two SIFT features, where p and p' denote the corresponding location, $\Delta\sigma$, $\Delta\theta$ represent scale log-ratio and orientation difference, respectively, and $N_{pp'}$ indicates the set of K nearest neighbouring matches of p and p', where the nearby features are determined by p if $\Delta\sigma \le 0$, otherwise by p'. By means of global scaling and rotation filtering, the scale log-ratio $(\Delta\sigma)$ and orientation difference $(\Delta\theta)$ left could represent a local similarity transformation. This stage computes how well the neighbouring feature pairs satisfy this transformation by $Score(m)$ which is given by:

$$Score(m) = \frac{1}{N}\sum\nolimits_{(q,q')\in N_{pp'}} (\lambda d_\sigma(m) + (1-\lambda)d_\theta(m)) \tag{5}$$

where:

$$d_\sigma(m) = \frac{|\|p-q\|_2 - 2^{\Delta\sigma}\|p'-q'\|_2|}{\|p-q\|_2 + 2^{\Delta\sigma}\|p'-q'\|_2} \tag{6}$$

$$d_\theta(m) = \left|\left|\arccos\left(\frac{<p-q, p'-q'>}{\|p-q\|_2\ \|p'-q'\|_2}\right)\right| - |\Delta\theta|\right| \tag{7}$$

where d_σ penalises the changes in length, d_θ penalises the changes in direction, and N stands for the number of nearby feature pairs which are close to both p and p', hence, $1/N$ penalises the matches that satisfy no spatial consistency, and $N \le K$. The constant λ weighs the two terms.

Intuitively, *Score(m)* evaluates the local similarity transformation consistency between match (p, p') by computing how well the segment $\overline{pq}$ is similar with the segment $\overline{p'q'}$ in terms of length, direction, and spatial layout, which is illustrated in Fig. 1. In this case, a threshold τ_{score} is set. If $Score(m)$ is less than τ_{score}, the match left is reserved, otherwise it is removed.

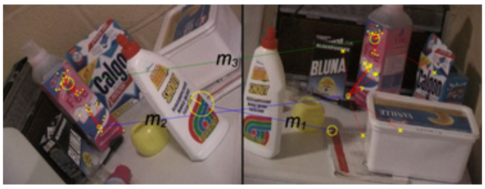

Fig. 1 *Image pair with three detected SIFT matches*

A putative SIFT feature point is noted by a circle and a segment, which represent scale and orientation, respectively. In left image, the points connected to the SIFT points are $K(=15$ in this example) nearest neighbouring points. In right image, the points connected to the SIFT points are corresponding to the nearest neighbouring points in the left image, where the points marked with ‘x’ are not the nearest neighbouring points of the right SIFT points, and others are. In this wash example, we figure out $\overline{\Delta\sigma}$ is 0 and $\overline{\Delta\theta}$ is -1.2863. We can see that m_3 is a correct match, and m_1 and m_2 are false. For m_1, $\Delta\sigma_1 = 1.3333$, $\Delta\theta_1 = 0.6685$, so m_1 is a false match according to formulas (3) and (4). For m_2, $\Delta\sigma_2 = 0$, $\Delta\theta_2 = -0.9065$, $N_2 = 6$, $Score(m_2) = 1.3162$, this match should be discarded in terms of LSTC constraint. For m_3, $\Delta\sigma_3 = 0.3333$, $\Delta\theta_3 = -0.9741$, $N_3 = 14$. $Score(m_3) = 0.0972$, it is a correct match

Experiment results and discussion: The proposed method was evaluated on six famous wide baseline image pairs (see Fig. 2), and compared with topological filtering (TF) [3] and topological clustering (TC) [4]. The parameters of the proposed method were fixed as $\tau_\sigma = 1$, $\tau_\theta = 0.5$, $\tau_{score} = 1.1$, $K = 15$ and $\lambda = 0.65$ through the experiments. For each image pair, the original putative matches were obtained by SIFT matching. A SIFT feature vector $f_1 \in F_1$ and a SIFT feature vector $f_2 \in F_2$ are taken as candidates for a match if and only if f_1 is the most similar measurement to f_2 and vice versa, i.e.

$$\begin{aligned}\forall f_1' \in F_1/f_1 : \|f_1 - f_2\|_2 < \|f_1' - f_2\|_2,\\ \forall f_2' \in F_2/f_2 : \|f_1 - f_2\|_2 < \|f_1 - f_2'\|_2\end{aligned} \quad (8)$$

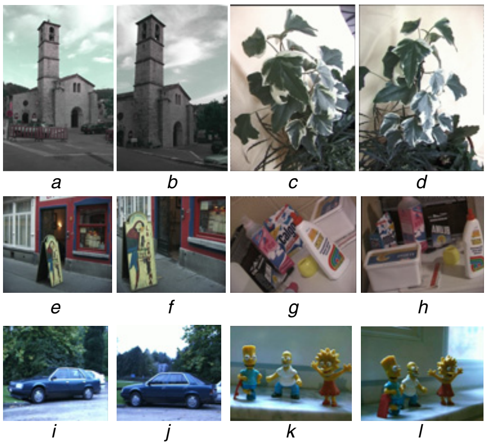

Fig. 2 *Image pair for test*

a, b Church
c, d Tree
e, f Mex
g, h Wash
i, j Auto
k, l Simpsons

The correct matches are identified manually, which could be used to compute the correct ratio of the original putative matches and the resulting matches. The experiment results are summarised in Table 1, the SIFT matching method establishes putative matches, and the other three methods discard mismatches. From the Table, we can see that the precision of our method is much higher than the original putative matches, and the number of resulting matches is considerably more than what is required in global estimating. Furthermore, our approach improves the currently achieved wide baseline matching performance, compared with state-of-the-art approaches, such as TF and TC.

Table 1: Precision and match number comparison of several popular mismatch discarding methods

Image pairs \ Method		SIFT [2]	TF [3]	TC [4]	Our method
Church	Number of matches	281	131	84	105
	Precision (%)	38.79	75.69	83.33	92.38
Tree	Number of matches	450	176	129	147
	Precision (%)	38.22	81.43	90.69	97.96
Mex	Number of matches	315	148	96	126
	Precision (%)	42.22	76.79	95.83	91.27
Wash	Number of matches	394	163	87	106
	Precision (%)	28.42	59.41	90.80	90.57
Auto	Number of matches	732	181	78	27
	Precision (%)	8.87	37.33	66.67	96.30
Simpsons	Number of matches	180	32	9	23
	Precision (%)	18.89	78.13	100	82.60

Conclusion and future work: The method proposed in this Letter combines the scale, orientation and spatial layout information of the SIFT features, and is divided into two steps: global scaling and rotation consistency (GSRC) constraint and local similarity transformation consistency (LSTC) constraint. The contrast of experimental results demonstrates that the proposed method can efficiently extract reliable SIFT matches from low-precision putative matches for wide baseline image pairs and outperforms or performs close to state-of-the-art approaches. Future work will focus on employing these high-precision matches to address advanced computer vision tasks, such as image retrieval, image registration and objection recognition.

21 October 2010
doi: 10.1049/el.2010.2967
One or more of the Figures in this Letter are available in colour online.

H.B. Zhou, D.Z. Zhang, C. Chen and J.W. Tian (*Institute for Pattern Recognition and Artificial Intelligence, Huazhong University of Science and Technology, Wuhan 430074, People's Republic of China*)

E-mail: zhouhuabing@gmail.com

References

1 Torr, P.H.S., and Murray, D.W.: 'The development and comparison of robust methods for estimating the fundamental matrix', *Int. J. Comput. Vis.*, 1997, **24**, (3), pp. 271–300

2 Lowe, D.: 'Distinctive image features from scale-invariant keypoints', *Int. J. Comput. Vis.*, 2004, **2**, (60), pp. 91–110

3 Ferrari, V., Tuytelaars, T., and Van Gool, L.: 'Wide-baseline multipleview correspondences'. Proc. IEEE CVPR, Madison, WI, USA, 2003, Vol. 1, pp. 718–725

4 Wang, Y.T., Zhang, D.Z., and Tian, J.W.: 'Discarding wide baseline mismatches via topological clustering', *Electron. Lett.*, 2008, **44**, (11), pp. 670–671

5 Tuytelaars, T., and Van Gool, L.: 'Matching widely separated views based on affine invariant regions', *Int. J. Comput. Vis.*, 2004, **1**, (54), pp. 61–85

6 Mikolajczyk, K., and Schmid, C.: 'A performance evaluation of local descriptors', *IEEE Trans. Pattern Anal. Mach. Intell.*, 2005, **27**, (10), pp. 1615–1629

7 Sivic, J., and Zisserman, A.: 'Efficient visual search of videos cast as text retrieval', *IEEE Trans. Pattern Anal. Mach. Intell.*, 2009, **31**, (4), pp. 591–606