Pattern Recognition Letters 33 (2012) 872–881

Contents lists available at SciVerse ScienceDirect

Pattern Recognition Letters

journal homepage: www.elsevier.com/locate/patrec

Good match exploration using triangle constraint

Xiaojie Guo, Xiaochun Cao *

School of Computer Science and Technology, Tianjin University, Tianjin 300072, China

ARTICLE INFO

Article history:
Available online 19 September 2011

Keywords:
Triangle constraint
Matching score
Number of correct matches
Similarity measurement

ABSTRACT

This paper presents a novel method for addressing the problem of finding more good feature pairs between images, which is one of the most fundamental tasks in computer vision and pattern recognition. We first select matched features by Bi-matching as seed points, then organize these seed points by adopting the Delaunay triangulation algorithm. Finally, triangle constraint is used to explore good matches. The experimental evaluation shows that our method is robust to most geometric and photometric transformations including rotation, scale change, blur, viewpoint change, JPEG compression and illumination change, and significantly improves both the number of correct matches and the matching score. And the application on estimating the fundamental matrix for a pair of images is also shown. Both the experiments and the application demonstrate the robust performance of our method.

1. Introduction

Finding correspondences between two images of the same scene or object is a very important issue for quite varied applications in computer vision and pattern recognition, such as symmetry detection (Lee and Liu, 2009), wide baseline matching (Schaffalitzky and Zisserman, 2002; Tuytelaars and Van Gool, 2004), building panoramas (Brown and Lowe, 2003), image classification (Bosch et al., 2008), image and video retrieval (Jegou et al., 2008), object recognition (Vedaldi and Soatto, 2006; Ferrari et al., 2006; Kannala and Brandt, 2007; Cho et al., 2010), action recognition (Scovanner et al., 2007), 3D reconstruction (Tola et al., 2008) and human detection (Dalal and Triggs, 2005; Wang et al., 2009).

To achieve the goal of finding correspondences, the existing methods generally include three main parts, *i.e.* feature detection, description and matching. First, feature detection is to capture discriminative and stable features, which has now reached some maturity in the literature. The existing feature detection schemes can be grouped into two categories: (1) point-based, like Harris corner detector (Harris and Stephens, 1988), Hessian–Laplace detector (Lowe, 2004; Mikolajczyk and Schmid, 2005) and Hessian-Matrix based detector (Bay et al., 2006), and (2) region-based, *e.g.* EBR (Tuytelaars and Van Gool, 1999), IBR (Tuytelaars and Van Gool, 2000), Harris-Affine region (Mikolajczyk and Schmid, 2001), Hessian-Affine region (Mikolajczyk and Schmid, 2002), MSER (Matas et al., 2004) and salient region (Kadir et al., 2004). As suggested by Nielsen and Lillholm (2001), different types of image features are appropriate to different tasks. Next, the neighborhood around every interest point for point-based schemes or every interest region for region-based is represented by a feature descriptor based on the concentrated information (for instance, gradient (Lowe, 2004) and Haar-Wavelet response (Bay et al., 2006)). The descriptor has to be distinctive meanwhile robust to geometric and photometric deformations and noises. Finally, the similarity measurement of features is employed to associate features, which is typically based on a distance between feature descriptors, such as the Mahalanobis or Euclidean distance.

Although the feature detection and description are well developed, the performance of finding correspondences between images and current similarity measurement approaches are still limited. The widely used similarity measurement (Lowe, 2004) adopts dot product between descriptors and then compares the ratio between the nearest and the next nearest neighbors against a predefined threshold to decide whether they are matched or not (we call this ratio test measurement the original matching method, OMM for short). Even though this strategy reduces influences from many geometric and photometric transformations, the drawback is that it sacrifices a great many of feature pairs that should be on the list of correct matches, which significantly affects the power of feature based applications such as reconstruction and image searching. One example is shown in Fig. 1(a). To solve the problem resulting from low matching score (the ratio between the number of correct matches and the number of total matches), a few improvements of similarity measurement have been developed. Leordeanu and Herbert (Leordeanu and Hebert, 2005) proposed a spectral technique for correspondence problem using pairwise constraints. The approach builds the adjacency matrix of a graph whose nodes represent the potential correspondences. The weights on the links represent pairwise agreements between potential correspondences. The authors then use the principal eigenvector of the

* Corresponding author. Tel.: +86 13820682739; fax: +86 22 27406538.
E-mail address: xcao@tju.edu.cn (X. Cao).

0167-8655/$ - see front matter © 2011 Elsevier B.V. All rights reserved.
doi:10.1016/j.patrec.2011.08.021

Fig. 1. Matching result comparison between the OMM and our method TCM with the same matching threshold. To better illustrate, we only show 1/6 of correct matches in each feature set (green '+'s) randomly. There are 39 hits (matching score 85.97%) in (a) and 216 hits (matching score 93.11%) in (b). (For interpretation of the references in colour in this figure legend, the reader is referred to the web version of this article.)

adjacency matrix and impose the mapping constraints to find correspondences between images. In (Brown and Lowe, 2003), RANSAC is employed to estimate the homography between images and filter out false matches according to the geometric constraint provided by the estimated homography. The method called Circular Earth Mover's Distance (CEMD) Rabin et al. (2008) effectively measures descriptors, which relies on an adaptation of Earth Mover's Distance. Jiang and Yu (2009) introduced a linear formulation that simultaneously finds feature point correspondences and global geometrical transformation in a constrained solution space, which can accurately match features. These measurements improve the matching accuracy, but with the cost of a reduced number of correct matches. Therefore, to boost the performance of finding correspondences, an effective technique for finding more good correspondences between two sets of features is imperative.

We aim to propose an effective similarity measurement approach, called Triangle Constraint Measurement, for exploring good (discriminative and accurate) matches. In comparison to the state-of-the-art, our method increases both the matching score (the ratio between the number of correct matches and the number of total matches) and the number of correct matches (see Fig. 1).

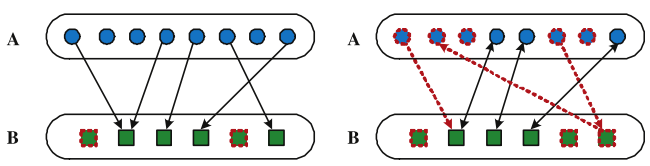
The rest of the paper is structured as follows. The flow of our proposed method is presented in Section 2. Parameter selection and performance evaluation are in Section 3. Section 4 shows the application on fundamental matrix estimation. Finally, the paper is concluded in Section 5.

2. Triangle Constraint Measurement

In this section, we detail our method, Triangle Constraint Measurement (TCM), ordered according to the stages *seed point selection*, *seed point organization* and *match exploration*. Before those stages, *feature extraction* should be performed. While distinctive and reliable features are expected to ensure the quality of the final result, the particular method used to extract features does not exert direct restraint on our proposed approach.

2.1. Seed point selection

We adopt the Bi-matching for selecting seed point pairs (stable matches) for the rest of the tasks, which is based on the observation that true positive matches between the feature set **A** from the reference image I_A and **B** from the target image I_B are more likely to be Bi-matched (from **A** to **B** and vice versa). In other words, if the relationship between one pair of features is unidirectional, then it is more probable for the feature pair to be either incorrect or unstable. Fig. 2 illustrates the Bi-matching procedure. We denote the process shown in Fig. 2(a) as $\mathbf{P}_{A\to B}$. As can be seen, $\mathbf{P}_{A\to B}$ has obtained matches from *A* to *B*, possibly containing a subset of unstable or even false positive matching pairs. According to the nature of the Bi-matching method, the matches are bidirectional. That is to say, the features that are unmatched by any of



(a) Matching result after $\mathbf{P}_{A\to B}$ (b) Matching result after $\mathbf{P}_{B\to A}$

Fig. 2. Illustration of Bi-matching.

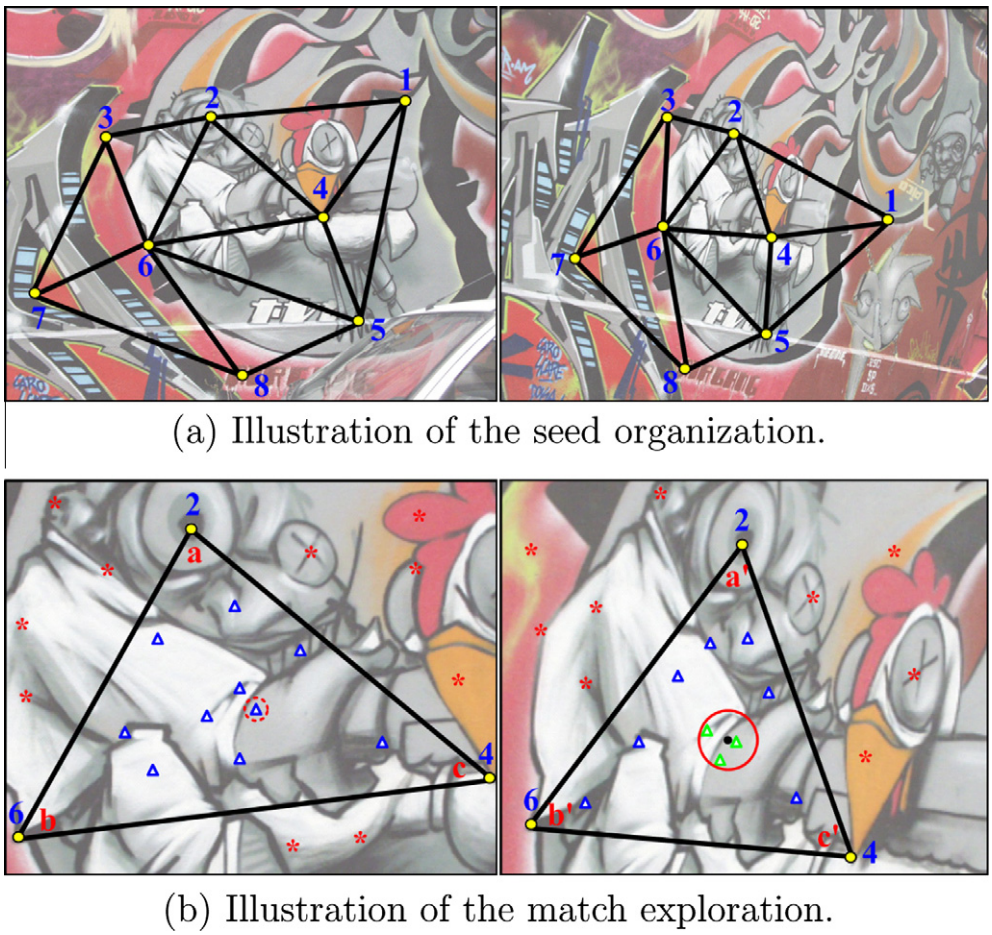


Fig. 3. Example of the Triangle Constraint Measurement. The triangle $\triangle_{246}$ in (a) is zoomed in at (b) to illustrate the match exploration process.

$\mathbf{P}_{A\to B}$ or $\mathbf{P}_{B\to A}$ (the second process as shown in Fig. 2(b)) will not be the elements of the set of seed points. Therefore, prior to executing $\mathbf{P}_{B\to A}$, eliminating the unmatched features from B in $\mathbf{P}_{A\to B}$ significantly reduces computational time compared to processing all the features from $\mathbf{B}$ in $\mathbf{P}_{B\to A}$, while introducing no negative effects to the result of seed point selection. Note that all the features in $\mathbf{A}$ participate in $\mathbf{P}_{B\to A}$. The solid arrows from A to B represent the matches from A to B and vice versa. The red[1] dashed arrows stand for the false or unstable matches. Similarly, the red dashed circles and squares are features filtered out by either $\mathbf{P}_{A\to B}$ or $\mathbf{P}_{B\to A}$. If there are few stable matches, the corresponding images are very likely to be irrelative or significantly distorted. The features finally matched by the Bi-matching method are considered seed points, *e.g.* the yellow points shown in Fig. 3, which are more reliable to execute the remaining processing steps.

2.2. Seed point organization

After selecting seed points, we need to organize the isolated seed points so that we can further explore more good feature pairs using geometric constraint. The geometric constraint here denotes an affine transformation which approximately relates a matched pair of local patches. The first reason to select an affine transformation rather than a perspective one is that the state of the art feature descriptor is only tolerant to affine transformation. The reason why we prefer affine rather than a more restricted similarity transformation is that it is closer to the real scenario. Since a 2D affine transformation has six degrees of freedom, three non-collinear points are sufficient to embody the geometric constraint. So triangle is preferred for such purpose. And, given the three vertices, the triangle is established without extra ordering confirmation of the vertices. To quickly organize the seed points obtained from the seed point selection stage, we adopt the Delaunay triangulation algorithm. The algorithm integrates the isolated points as triangles and, more importantly, maximizes the minimum angle of all the three angles of the triangles in the triangle mesh. In other words, it tends to avoid skinny triangles. The skinny triangle might cause degenerated expression of the affine transformation since the three vertices of a skinny triangle tend to be collinear. From another perspective, the triangulation divides the image plane into small pieces (triangles), which are more robust to real complex distortions than big patches.

In the implementation, we only apply the Delaunay triangulation algorithm to the seed points from I_A. As a result, the triangles are joint but not overlapped in I_A. The seed points from I_B are organized according to the same order with those from I_A according to the one-to-one relationship between the seed points from I_A and I_B. An example of the Delaunay triangulation is shown in Fig. 3(a). Note that the triangle mesh in I_B might not be exactly the same with that in I_A due to the false positive matches survived from the Bi-matching, *e.g.* the seed point pair 1 in Fig. 3(a). The match exploration described later is suitable to solve this problem.

2.3. Match exploration

So far, we have obtained the triangles in I_A with their correspondences in I_B. We take the case shown in Fig. 3(b) for example to

[1] For interpretation of color in Figs. 2 and 3, the reader is referred to the web version of this article.

explain how to explore matches using triangle constraint. For generalizing the case, we replace the vertices 2, 6 and 4 by a, b and c (a', b' and c') respectively for the left (right) triangle. The features are limited by the triangles $\triangle_{abc}$ and $\triangle_{a'b'c'}$, *i.e.* the features out of the triangles (the red '$*$'s in Fig. 3(b)) are not considered in current case. In other words, only the sets of features inside the triangles (the '$\triangle$'s in Fig. 3(b)) denoted as $\mathcal{P}_A$ for the reference triangle (left) and $\mathcal{P}_B$ for the target triangle (right) are involved.

2.3.1. Triangle constraint

For each feature $\mathbf{P}_i \in \mathcal{P}_A$ (*e.g.* the feature marked by dashed red circle) inside $\triangle_{abc}$, the relationship between $\mathbf{P}_i$ and the vertices of $\triangle_{abc}$ is

$$\mathbf{P}_i = \mathbf{a} + \beta(\mathbf{b} - \mathbf{a}) + \gamma(\mathbf{c} - \mathbf{a}), \tag{1}$$

where β and γ are the scale coefficients of the vector $(\mathbf{b} - \mathbf{a})$ and $(\mathbf{c} - \mathbf{a})$ respectively. Fortunately, the three vertices ($a = [x_a, y_a, 1]^T$, $b = [x_b, y_b, 1]^T$ and $c = [x_c, y_c, 1]^T$) of $\triangle_{abc}$ are known and the parameters $\mathbf{K}$ can be computed by

$$\mathbf{K} = \begin{bmatrix} \alpha \\ \beta \\ \gamma \end{bmatrix} = \begin{bmatrix} x_a & x_b & x_c \\ y_a & y_b & y_c \\ 1 & 1 & 1 \end{bmatrix}^{-1} \begin{bmatrix} x_i \\ y_i \\ 1 \end{bmatrix}, \tag{2}$$

where $\mathbf{P}_i = [x_i, y_i, 1]^T$ and $\alpha = 1 - \beta - \gamma$. We will shortly prove that if $\triangle_{abc}$ and $\triangle_{a'b'c'}$ are related by an affine transformation, then $\mathbf{P}'_i$ corresponding to $\mathbf{P}_i$ and the vertices of $\triangle_{a'b'c'}$ are also related by the same $\mathbf{K}$. Therefore, under translation, scale, rotation and affine transformations, the estimated point $\mathbf{P}_e$ (the black point in Fig. 3(b)) corresponding to $\mathbf{P}_i$ is computed via

$$\mathbf{P}_e = \begin{bmatrix} x_e \\ y_e \\ 1 \end{bmatrix} = \begin{bmatrix} x_{a'} & x_{b'} & x_{c'} \\ y_{a'} & y_{b'} & y_{c'} \\ 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} \alpha \\ \beta \\ \gamma \end{bmatrix}. \tag{3}$$

2.3.2. Applicability analysis

We assume that three vertices of $\triangle_{abc}$, *i.e.* $[x_a, y_a, 1]^T$, $[x_b, y_b, 1]^T$ and $[x_c, y_c, 1]^T$, from the reference image are projected into $[x'_a, y'_a, 1]^T$, $[x'_b, y'_b, 1]^T$ and $[x'_c, y'_c, 1]^T$ in the target image by a transformation matrix $\mathbf{H}$. Alike, the point $[x_i, y_i, 1]$ inside $\triangle_{abc}$ transforms through $\mathbf{H}[x_i, y_i, 1]^T$ as $[x'_i, y'_i, 1]^T$ accordingly. For isometries, similarity transformations and affine transformations, the corresponding parameters hence can be equally expressed like:

$$\begin{aligned} \mathbf{K}_\mathbf{H} &= \begin{bmatrix} \alpha' \\ \beta' \\ \gamma' \end{bmatrix} = \begin{bmatrix} x'_a & x'_b & x'_c \\ y'_a & y'_b & y'_c \\ 1 & 1 & 1 \end{bmatrix}^{-1} \begin{bmatrix} x'_i \\ y'_i \\ 1 \end{bmatrix} \\ &= \left(\mathbf{H} \begin{bmatrix} x_a & x_b & x_c \\ y_a & y_b & y_c \\ 1 & 1 & 1 \end{bmatrix} \right)^{-1} \left(\mathbf{H} \begin{bmatrix} x_i \\ y_i \\ 1 \end{bmatrix} \right). \end{aligned}$$

It is easy to verify that $\mathbf{K}_\mathbf{H} = \mathbf{K}$, *i.e.* triangle constraint is invariant to translation, rotation, scale and affine transformations, since the third dimension of the points' homogeneous coordinates keeps being 1 without normalization. When the perspective transformation happens, triangle constraint would not hold true due to the normalization difference for the homogeneous coordinates of points. But the triangulation splitting the image plane into small patches makes perspective transformation can be approximately treated as affine. In other words, match exploration is tolerant to perspective transformation.

To be more robust to noises and distortions, the area around the $\mathbf{P}_e$ within R pixels (in our experiments, $R = 3$) is defined as candidate area, as shown by the red solid circle in Fig. 3(b). The features $\mathbf{C}_j$ inside the candidate area are defined as candidate features (the green '$\triangle$'s in Fig. 3(b)). We denote $\mathcal{C} = \{\mathbf{C}_j\}$ as the set of candidate features. The similarity score between the $\mathbf{P}_i$ and the candidate feature $\mathbf{C}_j$ is measured by

$$\mathbf{s}_{i,j} = 1.5^{-(dist_{i,j}/R)^2} \times \mathbf{D}_i^T \mathbf{D}_{c_j}, \quad (j = 1, 2, \ldots, |\mathcal{C}|), \tag{4}$$

where $dist_{i,j}$ is the Euclidean distance between $\mathbf{P}_i$ and $\mathbf{C}_j$, $\mathbf{D}_i$ and $\mathbf{D}_{c_j}$ are the descriptors for $\mathbf{P}_i$ and $\mathbf{C}_j$ respectively, and $|\circ|$ stands for the cardinality of a set. The term $1.5^{-(dist_{i,j}/R)^2} \in (0, 1]$ reflects the spatial distance between the estimated location $\mathbf{P}_e$ and the location of $\mathbf{C}_j$. And the dot product $\mathbf{D}_i^T \mathbf{D}_{c_j}$ measures the similarity between the descriptors. Since the descriptors (*e.g.* SIFT used in our experiments) are normalized, the similarity can be measured by the cosine value of the angle θ between two descriptors, *i.e.* $\cos\theta = \mathbf{D}_i^T \mathbf{D}_{c_j}$. If the maximum score of all the features in $\mathcal{C}$ is greater than a predefined threshold τ, the corresponding feature pair is considered as temporary match, and there is at most one match for every feature $\mathbf{P}_i$. The value of τ is selected based on parameter selection (Section 3.2) for the remaining tasks including performance evaluation (Section 3.3) and application (Section 4).

After processing all the features from $\mathcal{P}_A$, there is a set $\mathcal{T}$ containing all the temporary matches between $\mathcal{P}_A$ and $\mathcal{P}_B$. The temporary matches are accepted as final matches if they satisfy

$$|\mathcal{T}| > \lambda \min\{|\mathcal{P}_A|, |\mathcal{P}_B|\}, \tag{5}$$

where λ is the size coefficient, the selection of which is described in Section 3.2. Otherwise, the triangle pair $\triangle_{abc}$ and $\triangle_{a'b'c'}$ and the temporary matches between them are discarded. This strategy is based on the observation that if the pair of the triangles is true correspondence, they will approximately satisfy the triangle constraint. On the other hand, if all the relative triangles of a vertex are discarded, the vertex is then removed from the seed points. This strategy is able to reduce wrong matches in the triangle constraint step and filter out the false positive matches survived from the Bi-matching step. Note that the removal of wrong matches would result in an uncomplete triangle mesh. Therefore, after removing a vertex (vertices), we repair the triangle mesh by detecting holes (broken parts) from the broken triangle mesh and then repairing them using the same triangulation algorithm meanwhile keeping the intact part. To avoid missing good matches that lie inside the newborn triangles, we also do the exploration in these triangles.

2.4. Complexity consideration

Suppose N and M are the numbers of features extracted from I_A and I_B respectively. To obtain n matches, OMM costs $\mathcal{O}(NM)$, where $n \ll \min(N; M)$ typically. As for Bi-matching procedure, $\mathbf{P}_{A\to B}$ spends actually the same with OMM. The complexity of $\mathbf{P}_{B\to A}$ is also $\mathcal{O}(NM)$ in theory, which is however hard to reach since the search space of features is significantly reduced by $\mathbf{P}_{A\to B}$ as described in Section 2.1. To triangulate n matches (Section 2.2), the Delaunay triangulation algorithm costs $\mathcal{O}(n \log n)$. There are $N - n$ and $M - n$ unmatched features for I_A and I_B after Bi-matching. The worst case of match exploration (Section 2.3) takes $\mathcal{O}((N - n)(M - n))$. This worst scenario, however, is unlikely to happen in practice, mainly because match exploration executes on corresponding triangle pairs separately, *i.e.* only the features inside the executing triangle pair are involved. Therefore, the worst computational complexity of our method is the same with the traditional matching method OMM, i.e. $\mathcal{O}(NM)$.

3. Experiments

The techniques proposed in this work are suitable to apply to all the local image features. In our experiments, we use the popular SIFT as the example. And three experiments including verification

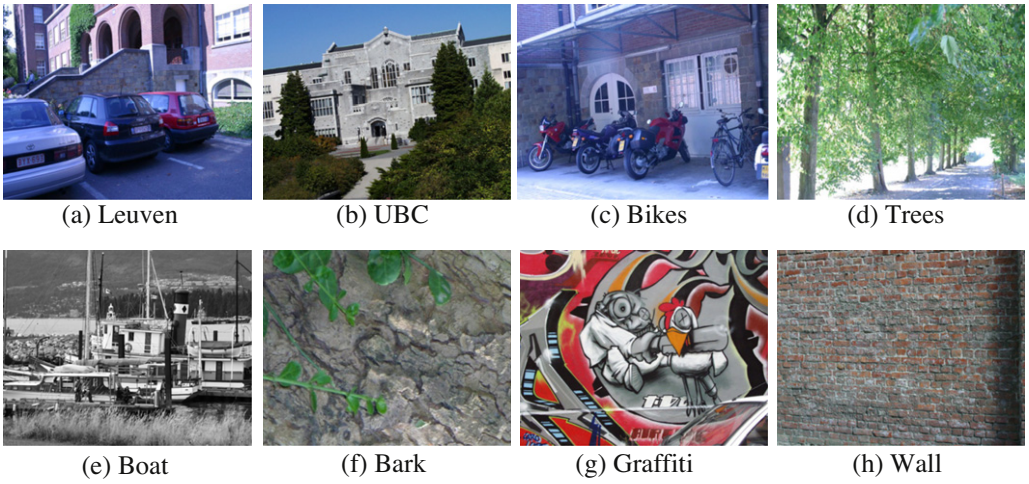


Fig. 4. Example images from eight categories of Oxford dataset.

of triangle constraint, parameter selection and performance evaluation, are presented in this section.

Dataset. The verification of triangle constraint is performed on simulated data described in Section 3.1. The evaluation of both the parameter effects and the performance of our method is carried on real images with different geometric and photometric transformations and for different scene types. Fig. 4 shows the first image (the original image) in every category of Oxford dataset[2] (Leuven, illumination change; UBC, JPEG compression; Bikes, blur, structured scene; Trees, blur, textured scene; Boat, scale change, structured scene; Bark, scale change, textured scene; Graffiti, viewpoint change, structure scene; Wall, viewpoint change, textured scene). Along with the images, the homographies are also provided.

Performance concern. The performance that we are concerned here includes two aspects, *i.e.* the number of correct matches (NCM) and the matching score (MS, the ratio between the number of correct matches and the number of all matches). Since the experiments are based on the same feature extraction scheme, the numbers of correspondences for involved different similarity measurements have no difference. Therefore, these two aspects are sufficient to compare the performance difference with respect to the traditional ratio test matching method OMM (Lowe, 2004).

3.1. Verification of triangle constraint

Triangle constraint is performed based on features extracted from images. The photometric invariance is guaranteed by the property of features, *e.g.* SIFT is robust to illumination change since descriptors are normalized. In other words, the photometric transformations do not change the triangulation result. Section 3.3 gives the results (Fig. 7(a) and Fig. 8(a)) that demonstrate the robustness to illumination change. The geometric transformations influence the shape of the triangle mesh. We employ simulation to verify that triangle constraint is tolerant to the geometric transformations including translation, scale change, rotation and affine transformations. First, on one image plane, we randomly generate 30 points as original seed points and another 1000 points as original candidate points. The generated 1030 points $P_i^1 (1 \leqslant i \leqslant 1030)$ are then transformed by different transformations including Euclidean, similarity and affine transformations to be P_i^2. All the image points P_i^2 are disturbed by adding zero-mean Gaussian noises with the standard derivation ranging from 0 to 8 pixels. Seed points are organized by using the Delaunay triangulation algorithm. The performance is measured via comparing the parameters $\mathbf{K} = [\alpha, \beta, \gamma]^T$ that relate original candidate points to their corresponding triangles and $\mathbf{K_H} = [\alpha', \beta', \gamma']^T$ that relate transformed candidate points to their corresponding triangles. The error is computed as follows:

$$err = \frac{|\alpha - \alpha'| + |\beta - \beta'| + |\gamma - \gamma'|}{3}. \tag{6}$$

The simulation involves three kinds of geometric transformations, *i.e.* Euclidean, similarity and affine transformations. Note that for similarity and affine transformations, scale change is involved. To reveal the influence of scale change, scaling up (Similarity1 and Affine1 in Fig. 5) and scaling down (Similarity2 and Affine2 in Fig. 5) are carried out. For each transformation under one noise level, we run the experiment results 100 times, and the mean and standard derivations of the errors in Eq. (6) are shown in Fig. 5. As can be seen from Fig. 5, triangle constraint under different geometric transformation performs robustly. When there are zero noises, the $\mathbf{K}$ and $\mathbf{K_H}$ are exactly same for all the three transformations. The errors increase almost linearly with respect to the level of noise. Note that the errors of Similarity2 (Affine2) are slightly higher than Similarity1 (Affine1), since triangles become smaller, which might influence the accuracy of estimated parameters.

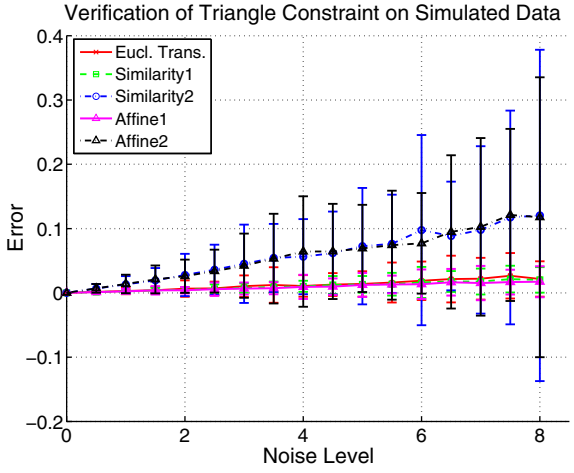


Fig. 5. Simulation results for verifying the tolerance of triangle constraint to geometric transformations. The difference between Similarity1 (Affine1) and Similarity2 (Affine2) is the value of scale factors. Similarity1 (Affine1) is the result of scaling up. Alternatively, Similarity2 (Affine2) is of scaling down.

[2] http://www.robots.ox.ac.uk/vgg/research/affine/.

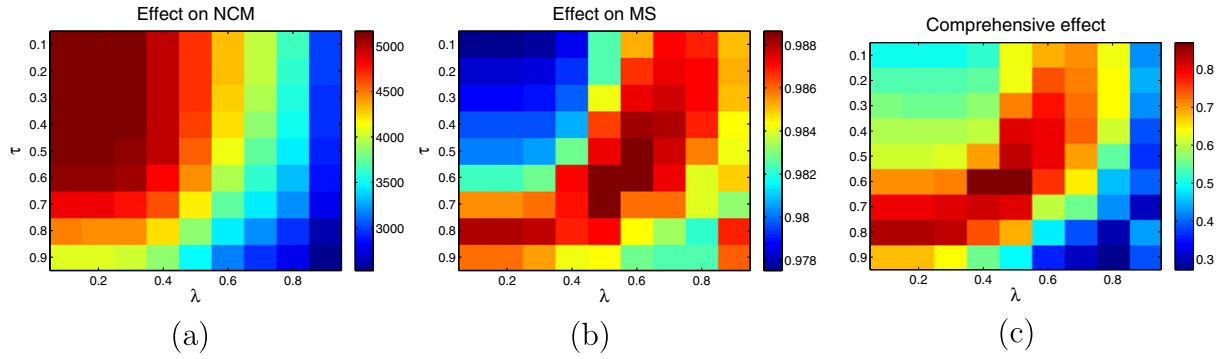


Fig. 6. NCM and MS differences by different parameter combinations. (a) Parameter effect on NCM. (b) Parameter effect on MS. (c) Comprehensive effect.

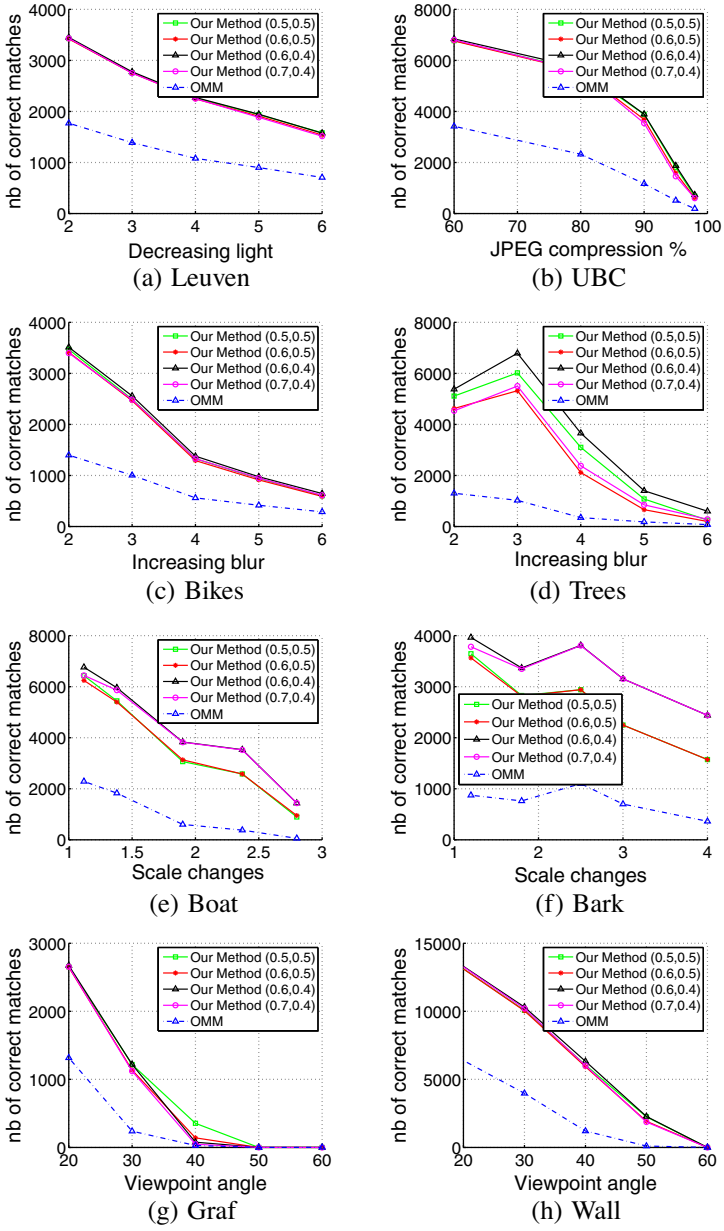


Fig. 7. Experiment results of the number of correct matches (NCM) for different categories from Oxford dataset.

3.2. Parameter selection

Good match exploration involves two parameters that introduced in Section 2.3, *i.e.* the similarity threshold, τ, for determining whether a feature pair is good enough to be selected as temporary match, and the size coefficient, λ, for testing if the temporary matches from corresponding triangles can be accepted as final matches. These two parameters affect the performance of the proposed method in terms of NCM and MS. Intuitively, the number of correct matches and the matching score vary as one falling and the other raising. To find optimal parameters empirically, we test different parameter combinations on Oxford dataset. Note that a match is defined correct if the distance between the ground truth location (projected by the provided ground truth homography for each pair of relative images) and the estimated location (computed by OMM or our method) is less than 6 pixels, incorrect otherwise.

3.2.1. Experimental setting

The first and the third images from every category of Oxford dataset are used for evaluating the performance difference between parameter combinations. The similarity threshold, τ, ranges from 0.0 to 1.0 in theory. Specially, $\tau = 1.0$ means that the match exploration is closed, *i.e.*, no matches will be explored. Conversely, the restriction of selecting good matches is totally released when $\tau = 0.0$. The size coefficient λ can also be set using a value in $[0.0, 1.0]$, where $\lambda = 0.0$ and $\lambda = 1.0$ are extreme cases. The case where $\lambda = 0.0$ turns all temporary matches into final matches no matter how many temporary matches are explored between the corresponding triangles. That is to say, the discriminating power is completely loss. On the other hand, $\lambda = 1.0$, directly rejects all of temporary matches. We skip over these extreme cases that violate the original intention of good match exploration, and adopt 9 options for both τ and λ, *i.e.* 0.1 to 0.9 at a fixed interval 0.1, to accomplish the task of parameter selection.

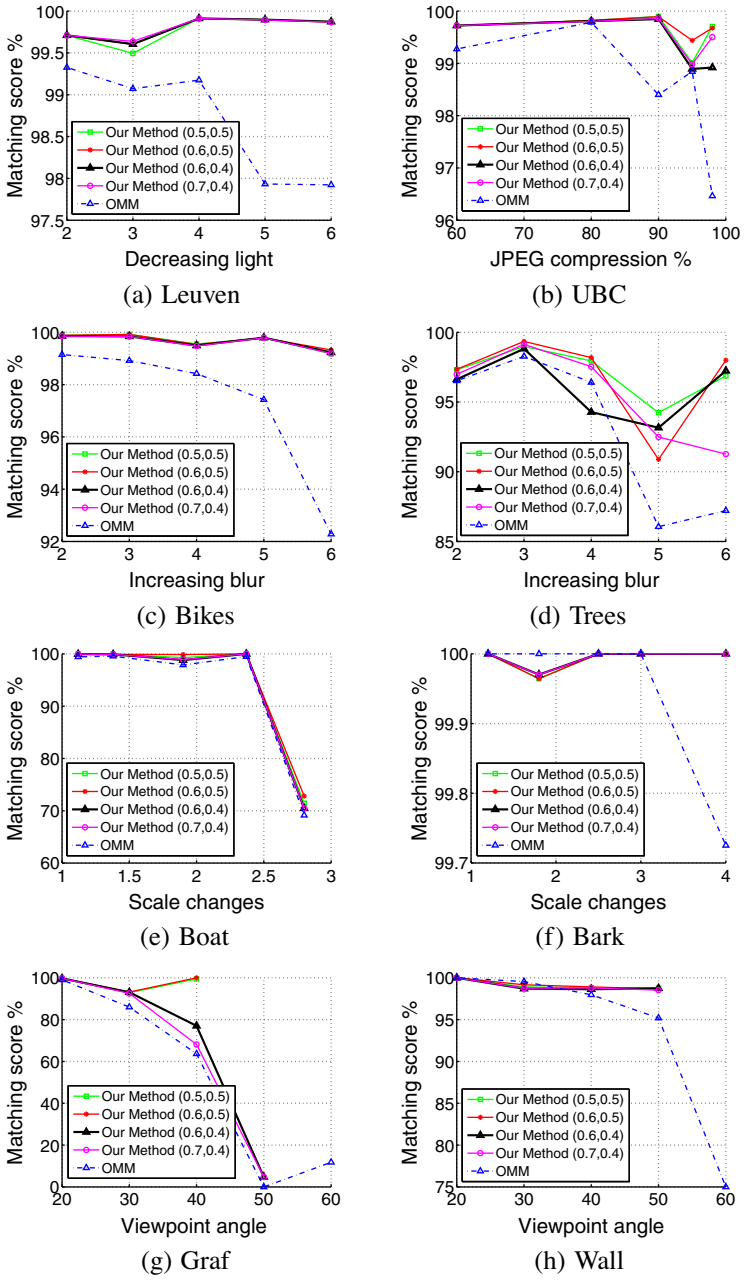


Fig. 8. Experiment results of the matching score (MS) for different categories from Oxford dataset.

The trade-off between the capability of exploring more matches (NCM) and the accuracy of matching (MS) lies on the parameter choice. Intuitively, higher τ and λ lead to less good matches to be explored but keep the matching score higher. On the other hand, the quantity of correct matches increases and the matching score decreases as τ and λ turn down. First, the influences of τ and λ on NCM and MS are separately evaluated as shown in Fig. 6(a) and (b). To generalize different categories, we then integrate the results of parameter influences on NCM by

$$Encm_{\tau,i,\lambda,j} = \frac{\sum_{c=1}^{8} NCM^{c}_{\tau,i,\lambda,j}}{8}. \quad (7)$$

As can be seen in Fig. 6(a), the quantity of *Encm* is low when τ and λ are all set to 0.9, which grows as τ and λ decrease and reaches the maxima when both τ and λ change to 0.1. Fig. 6(b) gives the integration of the parameter influence on MS, *Ems*, calculated in the same way with that of NCM, the trend of which however is not like the intuition described above. The reason is that high similarity threshold and size coefficient lead to few good matches explored by the procedure. So the improvement of MS will be slight. But if the restriction is too loose, bad matches are also accepted together with good ones as final matches, which will instead reduce the matching score. Therefore, the parameter combinations increasing MS are those exploring good matches much more than bad ones.

To analyze the comprehensive effect of two parameters, we merge the results of Fig. 6(a) and (b) as follows. For eliminating the influence brought by the difference of value range, the results are normalized into $[0,1]$ separately as (take *Encm* for example, and *Ems* in the same way)

$$CEncm_{\tau,i,\lambda,j} \leftarrow Encm_{\tau,i,\lambda,j} - \min\{Encm\}, \quad (8)$$

$$CEncm_{\tau,i,\lambda,j} \leftarrow \frac{CEncm_{\tau,i,\lambda,j}}{\max\{CEncm\}}. \quad (9)$$

To balance the importance of *CEncm* and *CEms*, the comprehensive effect CE is computed as follows

$$CE_{\tau,i,\lambda,j} \leftarrow \frac{CEncm_{\tau,i,\lambda,j} + CEms_{\tau,i,\lambda,j}}{2}. \quad (10)$$

Fig. 6(c) draws the visual result of comprehensive effect. As can be seen in the graph, the overall best performance is obtained by the parameter combination ($\tau = 0.6, \lambda = 0.4$). As the values of the parameters vary around, the performance decreases. Together with the best combination, *i.e.* ($\tau = 0.6, \lambda = 0.4$), we additionally select three parameter combinations ($\tau = 0.5, \lambda = 0.5$), ($\tau = 0.6, \lambda = 0.5$) and ($\tau = 0.7, \lambda = 0.4$) for further comparing the performance of our method against OMM.

3.3. Performance evaluation

3.3.1. Relative image pair matching

Every category of Oxford dataset contains six images capturing the same scene with geometric and/or photometric transformations. Figs. 7 and 8 show the results in terms of NCM and MS for different categories, respectively. As can be seen, all of the NCMs by our method using four different parameter combinations selected based on the parameter comprehensive effect are considerably more than those obtained by OMM for all eight categories as shown in Fig. 7. And the quantities of correct matches using our method from high to low are obtained by setting ($\tau = 0.6, \lambda = 0.4$), ($\tau = 0.7, \lambda = 0.4$), ($\tau = 0.5, \lambda = 0.5$) and ($\tau = 0.6, \lambda = 0.5$), which is consistent with the trend shown in Fig. 6(a). With respect to MS, TCM performs mostly better than OMM except for the third pairs of the image sequence of Trees with ($\tau = 0.6, \lambda = 0.4$) as shown in Fig. 8(d). The reason is that there are a huge amount of features (the largest number of correct matches reaches about 8000 by TCM) that increase the possibility of accidentally considering incorrect matches as correct. Note that we do not draw the matching scores for the image pairs with no matches by using our method (only very few unstable correct matches by OMM), *i.e.* the matching score is undefined ($\frac{0}{0}$). The margins between the MS curves of our method using different parameters are not as distinct as those of the NCM curves, since they are close to 100%. Again, the results displayed in Figs. 7 and 8 certify that there is a trade-off between NCM and MS, which can be adjusted according to the demand of task. From the results, we can see that TCM using the parameter combinations selected from Section 3.2 significantly outperforms OMM.

3.3.2. Irrelative image pair matching

From another perspective, a good measurement has to be robust to noisy images. Therefore, we use an additional pair of

Table 1
Performance comparison between our method and OMM for fundamental matrix estimation. The error is measured using Symmetric Epipolar Distance.

Image pair	MC I	MC II	MC III	UL	WC
Our method	2.2385	1.2107	1.9037	1.9783	2.2544
OMM	3.4593	1.9446	2.4936	3.8952	2.6654

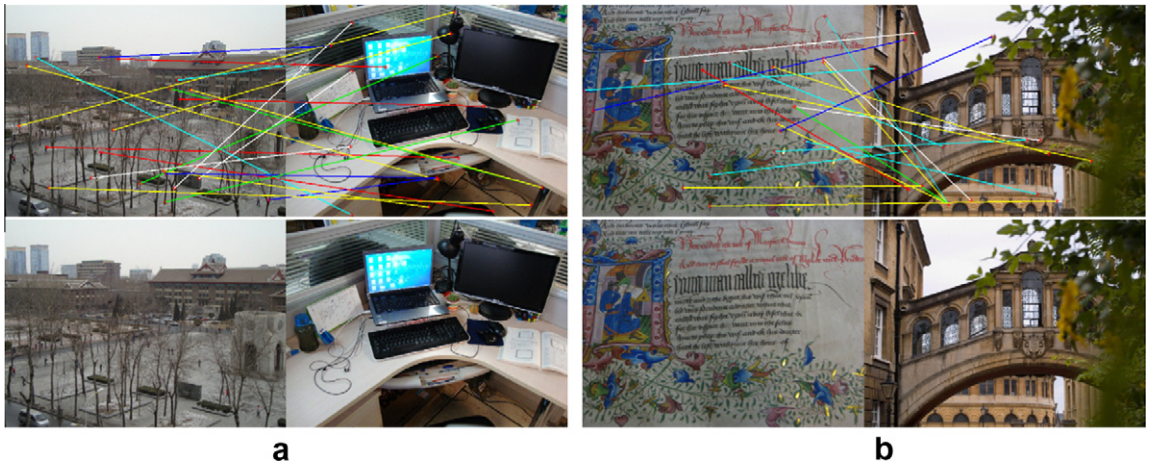


Fig. 9. Irrelative image pair. **Upper row:** OMM results and **Lower row:** our results. There are 22 hits by OMM and 0 hit by our method for image pair (a), and 20 hits by OMM and 0 hit by our method for image pair (b). False positive matches are marked by '∗' and linked by lines.

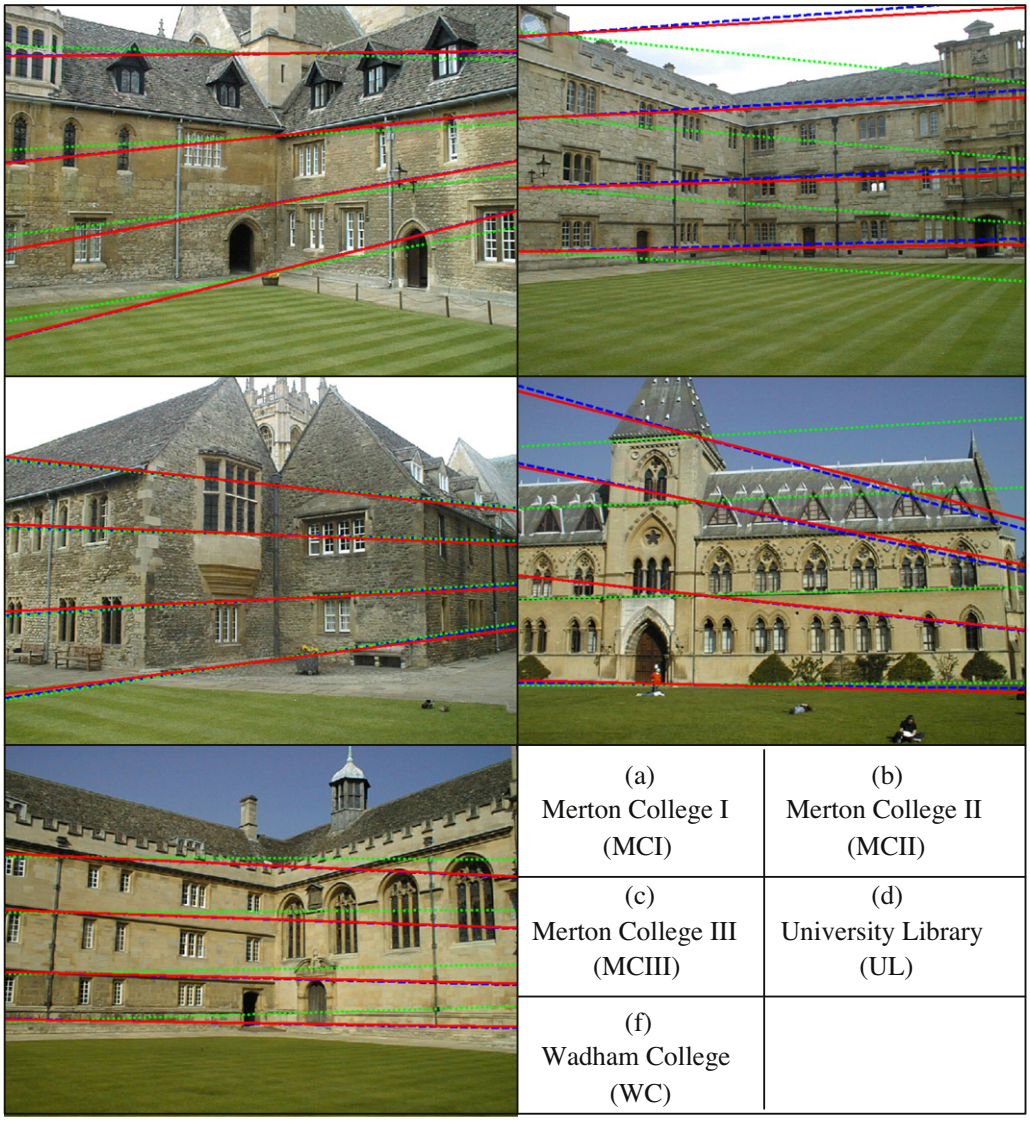


Fig. 10. Visual results of epipolar geometry estimation for each category from Oxford Colleges dataset. The ground truth, our result and the result by OMM are plotted in red solid line, blue dashed line and green dotted line, respectively. (For interpretation of the references in colour in this figure legend, the reader is referred to the web version of this article.)

irrelative images taken by ourselves to test and verify the validity of our method. As shown in Fig. 9, there are 22 hits by OMM marked in red '$*$' and linked by lines and 0 hit by our method. No doubt that the 22 matches are all false positive.

4. Application to fundamental matrix estimation

As one of the most fundamental tasks in computer vision and pattern recognition, many applications benefit from more good matches between images, such as tracking (Ta et al., 2009), depth-map estimation (Tola et al., 2008) and image retrieval (Jegou et al., 2008). In this section, we apply Triangle Constraint Measurement to the task of estimating the fundamental matrix, F, from point correspondences, which is of practical importance. The fundamental matrix relating two images is estimated from a number of correspondences between the images, which are assumed to be projections of the same 3D points. As is known, fundamental matrix can be described by a 3×3 singular matrix, which connects two perspective images of an object or scene.

F estimation has been the focus of many researchers, and a few techniques have been developed (Luong and Faugeras, 1996; Chesi et al., 2002; Hartley and Zisserman, 2004). We adopt a popular and simple scheme, *i.e.* fitting F using RANSAC,[3] to estimate the F based on the matches between real images from Oxford colleges dataset[4] using our method and OMM, respectively, and compare experimentally the errors between the estimated Fs in virtue of the ground truth corner correspondences provided by the dataset. Several error criteria have been proposed, a survey of which is studied in (Fathy et al., 2011). Without loss of generality, we choose Symmetric Epipolar Distance (SED) as our error criterion, which is defined as:

$$\sqrt{\frac{\sum_{i=1}^{p}\frac{\left(\bar{\mathbf{x}}_i'^T\widehat{\mathbf{F}}\bar{\mathbf{x}}_i\right)^2}{(\widehat{\mathbf{F}}\bar{\mathbf{x}}_i)_1^2+(\widehat{\mathbf{F}}\bar{\mathbf{x}}_i)_2^2}+\frac{\left(\bar{\mathbf{x}}_i'^T\widehat{\mathbf{F}}\bar{\mathbf{x}}_i\right)^2}{\left(\widehat{\mathbf{F}}^T\bar{\mathbf{x}}_i'\right)_1^2+\left(\widehat{\mathbf{F}}^T\bar{\mathbf{x}}_i'\right)_2^2}}{2p}}, \tag{11}$$

where $\bar{\mathbf{x}}$ denotes for the ground truth of $\mathbf{x}$, $\widehat{\mathbf{F}}$ is the estimated F and p stands for the number of used correspondences provided by the dataset.

[3] http://www.csse.uwa.edu.au/pk/Research/MatlabFns.
[4] http://www.robots.ox.ac.uk/vgg/data2.html.

Since the estimation is based on RANSAC, the estimated fundamental matrix may slightly vary for each run. So we repeat the estimation procedure 100 times for each category and average the SED as the final errors. Table 1 gives the error comparison between our method and OMM. The results reveal that our method is more robust and suitable than OMM. Visual results can be found in Fig. 10. The ground truth, our result and the result by OMM are plotted in red solid line, blue dashed line and green dotted line, respectively.

5. Conclusion

Similarity measurement is the crucial part for numerous tasks built on local image features. There are two factors for a good measurement, *i.e.* the matching accuracy and the number of correct matches. Unfortunately, existing methods hardly take in consideration both two factors. We have proposed a new measurement to improve the matching accuracy and the number of correct matches simultaneously. Experimental results have demonstrated that our method outperformed OMM in terms of matching relative and irrelative image pairs.

Acknowledgements

This work was supported by National Natural Science Foundation of China (No. 60905019), Program for New Century Excellent Talents in University, Tianjin Key Technologies R&D program (No. 11ZCKFGX00800), Tsinghua – Tencent Joint Laboratory for Internet Innovation Technology, SKL of CG&CAD, and Open Projects Program of NLPR.

References

Bay, H., Tuytelaars, T., Van Gool, L., 2006. Surf:speeded up robust features. In: European Conf. Comput. Vision, pp. 404–417.
Bosch, A., Zisserman, A., Muñoz, X., 2008. Scene classification using a hybrid generative/discriminative approach. IEEE Trans. Pattern Anal. Machine Intell. 30 (4), 712–727.
Brown, M., Lowe, D., 2003. Recognising panoramas. In: Internat Conf. Comput. Vis., pp. 1218–1227.
Chesi, G., Garulli, A., Vicino, A., Cipolla, R., 2002. Estimating the fundamental matrix via constrainted least-squares: A convex approach. IEEE Trans. Pattern Anal. Machine Intell. 24 (3), 397–401.
Cho, M., Shin, Y., Lee, K., 2010. Unsupervised detection and segmentation of identical objects. In: IEEE Conf. Comput. Vision Pattern Recog., pp. 1617–1624.
Dalal, N., Triggs, B., 2005. Histograms of oriented gradients for human detection. In: IEEE Conf. Comput. Vision Pattern Recog., pp. 886–893.
Fathy, M., Hussein, A., Tolba, M., 2011. Fundamental matrix estimation: A study of error criteria. Pattern Recognition Lett. 32 (2), 383–391.
Ferrari, V., Tuytelaars, T., Van Gool, L., 2006. Simultaneous object recognition and segmentation from single or multiple model views. Internat. J. Comput. Vision 67 (2), 159–188.
Harris, C., Stephens, M.J., 1988. A combined corner and edge detector. In: Alvey Vision Conf., vol. 20, pp. 147–152.
Hartley, R., Zisserman, A., 2004. Multiple View Geometry in Computer Vision. Cambridge University Press.
Jegou, H., Douze, M., Schmid, C., 2008. Hamming embedding and weak geometric consistency for large scale image search. In: European Conf. Comput. Vis., pp. 304–317.
Jiang, H., Yu, S., 2009. Linear solution to scale and rotation invariant object matching. In: IEEE Conf. Comput. Vision Pattern Recog., pp. 2474–2481.
Kadir, T., Zisserman, A., Brandy, M., 2004. An affine invariant salient region detector. In: European Conf. Comput. Vision, pp. 228–242.
Kannala, J., Brandt, S., 2007. Quasi-dense wide baseline matching using match propagation. In: IEEE Conf. Comput. Vision Pattern Recog., Minnesota, USA.
Lee, S., Liu, Y., 2009. Curved glide-reflection symmetry detection. In: IEEE Conf. Comput. Vision Pattern Recog., pp. 1046–1053.
Leordeanu, M., Hebert, M., 2005. A spectral technique for correspondence problems using pairwise constraints. In: Internat. Conf. Comput. Vision, pp. 1482–1489.
Lowe, D., 2004. Distinctive image features from scale-invariant keypoints. Internat. J. Comput. Vision 60 (2), 91–110.
Luong, Q., Faugeras, O., 1996. The fundamental matrix: Theory, algorithms and stability analysis. Internat. J. Comput. Vision 17 (1), 43–76.
Mads Nielsen, Martin Lillholm., 2001. What do features tell about images? In: Internat. Conf. Scale-Space Morphol. Comput. Vision, pp. 39–50.
Matas, J., Chum, O., Urban, M., Pajdla, T., 2004. Robust wide-baseline stereo from maximally stable extremal regions. In: British Machine Vision Conf., pp. 761–767.
Mikolajczyk, K., Schmid, C., 2001. Indexing based on scale invariant interest points. In: Internat. Conf. Comput. Vision, pp. 525–531.
Mikolajczyk, K., Schmid, C., 2002. An affine invariant interest point detector. In: European Conf. Comput. Vision, pp. 128–142.
Mikolajczyk, K., Schmid, C., 2005. A performance evaluation of local descriptors. IEEE Trans. Pattern Anal. Machine Intell. 27 (10), 1615–1630.
Rabin, J., Delon, J., Gousseau, Y., 2008. Circular earth mover's distance for the comparison of local features. In: Internat. Conf. Pattern Recog., Florida, USA.
Schaffalitzky, F., Zisserman, A., 2002. Multi-view matching for unordered image sets. In: European Conf. Comput. Vis., pp. 414–431.
Scovanner, P., Ali, S., Shah, M., 2007. A 3-dimensional sift descriptor and its application to action recognition. In: ACM Internat. Conf. Multimedia, pp. 357–360.
Ta, D., Chen, W., Gelfand, N., Pulli, K., 2009. Surftrac: Efficient tracking and continuous object recognition using local feature descriptors. In: IEEE Conf. Comput. Vision Pattern Recog., pp. 2937–2944.
Tola, E., Lepetit, V., Fua, P., 2008. A fast local descriptor for dense matching. In: IEEE Conf. Comput. Vision Pattern Recog., pp. 1–8.
Tuytelaars, T., Van Gool, L., 1999. Content-based image retrieval based on local affinely invariant regions. In: Internat. Conf. Visual Inform. Syst., pp. 493–500.
Tuytelaars, T., Van Gool, L., 2000. Wide baseline stereo matching based on local, affinely invariant regions. In: British Machine Vision Conf., pp. 412–425.
Tuytelaars, T., Van Gool, L., 2004. Matching widely separated views based on affine invariant regions. Internat. J. Comput. Vision 1 (59), 61–85.
Vedaldi, A., Soatto, S., 2006. Local features, all grown up. In: IEEE Conf. Comput. Vision Pattern Recog., pp. 1753–1760.
Wang, X., Han, T., Yan, S., 2009. An hog-lbp human detector with partial occlusion handling. In: Internat Conf. Comput. Vision, pp. 32–39.