Image Matching Algorithm Based on Topology Consistency of Bidirectional Optimal Matching Point Pairs

Aihua Wu,[1] Weizheng Chen,[2] Yijie Bian,[1] and Song Xue[1*]

[1]Business School, Hohai University, Nanjing 211100, China
[2]Nanjing XuanNing Information Technology Co., Ltd.,
Huashen Science and Technology Park, Yuhuatai District, Nanjing 210000, China

The random sample consensus (RANSAC) algorithm is commonly used to estimate the parameters of the image transformation model based on matching point pairs in the feature-based image matching field. If the dataset of matching point pairs contains outliers, the conventional RANSAC algorithm may take a large number of iterations to obtain the desired model. To reduce mismatching, we propose the bidirectional optimal matching method, aiming to find robust parameters within a short time. The topology-consistency-based sampling method is introduced to divide the dataset into certain consensus sets, and sampling from each of them can reduce randomness. Then, all point pairs from a consensus set are used to estimate a model, and a point pair unsuitable for the model is deleted in each iteration, which is demonstrated to be faster than the conventional RANSAC. The superiority of the proposed method in fingerprint matching based on the scale-invariant feature transform is shown in experiments.

1. Introduction

Fingerprint recognition technology has been widely used in electronic consumer terminals. In these application scenarios, there are three main forms of fingerprint modules: capacitive, optical, and ultrasonic. Owing to the cost of the hardware, the above methods only collect small finger areas. Usually, the above fingerprints do not contain enough details, and it is difficult for the conventional matching methods based on the details to achieve ideal results. Although most electronic consumer terminals cannot provide enough computing power for AI-based image matching algorithms,[1–6] image matching based on local feature points is widely used in the above scenes, such as the well-known scale-invariant feature transform (SIFT) algorithm,[7] the speeded-up robust features (SURF) algorithm,[8] binary robust independent elementary features (BRIEF) algorithm,[9] and the features from accelerated segment test (FAST) and rotated BRIEF (ORB) algorithm.[10]

However, differences between the various image matching algorithms based on feature points, such as orientation, coordinates, and sub-vector description, may lead to diverse effects on the same image.[11] It is difficult to choose the correct image matching method based on

---

feature points to complete the image matching task. Therefore, more attempts have been made to propose image matching methods based on different feature point algorithms, such as SIFT, SURF, and ORB.[12]

This method presented in this paper has a higher matching accuracy and a higher calculation speed than SIFT, SURF, and ORB. The following are the key points of the image matching method:

(1) Bidirectional optimal matching point pair (BOMP): A matching point pair contains two matching points, one from the original image and one from the target image. For an optimal matching point pair, the point in the target image is the optimal matching of the point in the original image based on some appraisal criterion. In a BOMP, each point is the optimal matching point of the other. There might be a mismatch between two points if image matching methods were only based on single-direction feature point matching as before. In the meantime, the mismatching points are retained. Therefore, to a certain extent, it would not be beneficial to calculate the mapping relation between two images. The method proposed in this paper matches points from two images in two ways to ensure that its ratio of mismatch is lower than those of SIFT, SURF, and ORB in retained point pairs, which is beneficial for the calculation of the mapping relation.

(2) Topology consistency: There are a large number of matching point pairs between two matching images. Each feature point in every matching point pair contains orientation and coordinate information, and each matching point pair has a transformation angle determined by the orientations of its two matching points. Two matching point pairs are consistent only if they are subject to a mapping relation with the condition that the transformation angles of the two matching point pairs are close and the relative position of the two points in the target image is similar to that in the original image. All matching point pairs can be divided into some consensus sets. This relative position is described by the distance and azimuth. The distance denotes the length of the vector formed between the two points in the target image and thus is in the original image, where the azimuth denotes the angle between the orientation of the starting point and the vector.

(3) Homography matrix estimation: A homography matrix is used to describe the mapping relation between the target image and the original image, and it is estimated by an iterative process. The least squares method is used to calculate the initial homography matrix based on all the point pairs in a consensus set. Thereby, the projection deviation of all matching point pairs in the consensus set is calculated from the homography matrix to find the maximum deviation. If the maximum deviation is larger than the tolerance limit, the corresponding matching point pair can be removed. If a matching point pair is deleted and the number of remaining matching point pairs in the consensus set is not less than four, the iteration process returns to the first step. Otherwise, it can be terminated. Theoretically, a consensus set can produce at most one homography matrix, which is the desired model, but in practice if more than one homography matrix is generated, none of the matrices should be accepted.

The method presented in this paper completes the task of image matching through three steps, BOMP, topology consistency, and homography matrix estimation, to address the problem of image matching. The BOMP decreases the ratio of mismatching point pairs to reduce the

amount of calculation required for the mapping relation and the sample space of the matching point pairs. The topology consistency determines whether two matching point pairs can be subject to one mapping model, and the homography matrix estimation calculates the mapping relation from some consensus sets.

This paper is organized as follows. A brief review of related works is given in Sect. 2, while BOMPs are described in detail in Sect. 3. The topology consistency is presented in Sect. 4 and the homography matrix estimation is presented in Sect. 5. Then, in Sect. 6, comprehensive experiments are shown to verify the performance of the proposed method. Finally, we conclude our study in Sect. 7.

2. Related Works

With the rapid development of image matching methods based on feature points, Lowe proposed the SIFT algorithm, which fully considers the variations of rotation and scale, the effects of lighting and noise, and so forth, in image transformation.[7] Firstly, the difference of Gaussians (DoG) is applied to estimate a scale space, and the key feature points in the image to be matched are next positioned through locating and refining candidate key feature points by eliminating low-contrast points. Then, the orientation of each key feature point in the image is assigned on the basis of the local image gradient. Finally, the descriptor of each key feature point is calculated on the basis of its orientation and local image gradient.

Bay *et al.* proposed the SURF algorithm based on SIFT, which greatly improved the efficiency of computation. Unlike the Gaussian averaging of the image, SURF approximates the DoG using box filters. SURF can perform image matching tasks on images of different scales by searching local interest feature points using a blob detector based on the Hessian matrix.[13] In addition, it uses wavelet responses to describe feature points by selecting a neighborhood around the key point and dividing it into sub-areas. The potential points of interest in the image are searched for through Laplace symbols that distinguish bright blobs on dark backgrounds from the reverse case. In the matching stage, SURF only compares the features with the same type of contrast, resulting in high-speed image matching.

BRIEF is another alternative to SIFT with a lower complexity than SIFT but a similar matching performance.[9] Rublee *et al.* achieved a breakthrough by developing the ORB algorithm, which has a higher computational efficiency and superior real-time performance to SIFT and SURF.[10] The matching process of ORB is divided into three steps: feature point extraction, feature point descriptor generation, and feature point matching. A flow chart of the method is shown in Fig. 1. Firstly, ORB is a fusion of some modified rapid feature detectors and brief descriptors. The main points are initially rapidly determined and used, then the Harris angle measure is explored to find the first N feature points. Secondly, ORB generates feature point descriptors through the data obtained by calculating feature points of interest. Lastly, ORB matches the feature points with the same descriptors to match the images.

Choosing the key matching point pairs is important to calculate the mapping relation for the image matching task. RANSAC, as a classical algorithm, is widely used to estimate the parameters of the mapping relation for image matching and stitching.[14] The mapping relation
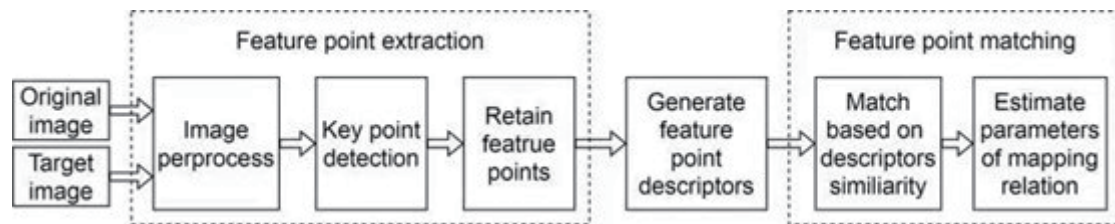
Fig. 1. Image matching flow chart based on feature points.

between the original image and the target image is calculated by selecting a set of matching point pairs randomly. RANSAC searches for the optimal feature point set by the iterative calculation of feature points, and the feature point sets are spliced to form a feature point matrix. Assuming that there are both correct and false data in the set of sample feature points, the optimal homography matrix is calculated by their respective algorithms for precise matching.[15,16] However, as the number of iterations is uncontrollable, RANSAC requires a large amount of computing resources, and if the maximum number of iterations is limited, the robustness and efficiency of the method will be very low.

To solve the above problems, many researchers have proposed many variants of the RANSAC algorithm for 2D feature point matching.[17–21] Chum and Matas proposed the progressive sample consensus algorithm, which exploited the linear ordering defined on the set of correspondences by using a similarity function to establish tentative correspondences.[17] Shi *et al.* proposed an improved RANSAC algorithm that removed parts of the error feature points before the iterations by two methods: the elimination of features not belonging to the target area and the removal of crossing points.[18] Fotouhi *et al.* proposed a robust and efficient method named SC-RANSAC, which detected gross outliers to increase the inlier-to-outlier ratio in a reduced set of corresponding image points by using a new hypothesis and verification scheme that utilizes spatial relations between extracted corresponding points in two images.[19] Liu and Bu proposed an improved RANSAC algorithm to improve the sample probability of correct matching points by sorting pairs of matching points by the Euclidean distance and sampling only the middle matching point pairs.[20] Rahman *et al.* proposed descendant likelihood sampling RANSAC (DL-RANSAC), which reduced randomness by introducing a descending likelihood that provided prior knowledge of the random population before selecting the hypothesis set.[21] The methods proposed in Refs. 17–19 can be considered as a preprocessing step for RANSAC, while the methods in Refs. 20 and 21 change the sampling method during iterations.

The previous works proposed improved RANSAC algorithms to reduce the number of iterations by introducing some external information. However, these algorithms did not fully consider the topological constraints between pairs of matching feature points, which is the basis of our proposed method. Also, in this paper, the number of remaining matching point pairs is reduced in the feature matching stage, which can significantly reduce the sampling space required for RANSAC. We compare the performance characteristics of different image matching techniques based on SIFT in three small-area fingerprint libraries. Our experiments show that the method proposed in this paper converges to the desired result faster than the conventional RANSAC.

3.   Bidirectional Optimal Matching Point Pairs

3.1   Feature points

The feature point set extracted from the original image is defined as

$$T = \{t_i\}, i = 1, 2, 3, \ldots, m, \tag{1}$$

where t_i is the feature point extracted from the original image.

The feature point set extracted from the target image is defined as

$$V = \{v_i\}, i = 1, 2, 3, \ldots, n, \tag{2}$$

where v_i is the feature point extracted from the target image. Each feature point is described by four properties, α, x, y, and D:

$$t_i = (\alpha, x, y, D), \tag{3}$$

where α is the orientation of the feature point, x and y are its abscissa and ordinate, respectively, and D is its descriptor vector.

3.2   Optimal matching point pairs

The similarity between t_i in the original image and v_j in the target image is denoted as s_{ij}. This similarity is measured using the distance between the descriptors of two feature points, such as the cosine or Euclidean distance. The formula using the cosine distance for S_{ij} is

$$S_{ij} = \frac{D_i D_j}{\|D_i\| \times \|D_j\|}. \tag{4}$$

For each feature point t_i in the original image, its similarity with all the feature points in the target image can be denoted as

$$S_{l*} = \{s_{ij}\}, j = 1, 2, 3, \ldots, n. \tag{5}$$

There are two methods of defining the optimal matching point in the target image for t_i, the exhaustion method and Lowe's method. The exhaustion method defines v_j as the optimal matching point for t_i if the similarity between t_i and v_j is the maximum of S_{l*}, expressed as

$$s_{ij} = \max_{j=1,\ldots,n} S_{l*}. \tag{6}$$

Lowe's method defines v_j as the optimal matching point for t_i if the similarity between t_i and v_j is the maximum of S_{l*} and the second-largest similarity in S_{l*} is less than a predetermined percentage of the maximum.

In matching the original image to the target image, $(t_x, \rightarrow, v_y)$ denotes the optimal matching point pair, where v_y is the optimal matching point in the target image for t_x from the original image, and $(t_x, \leftarrow, v_y)$ denotes the optimal matching point pair, where t_x is the optimal matching point in the original image for v_y from the target image.

3.3 Ambiguous matching scenarios

In the point matching stage, the previous methods only calculated optimal matching from the original image to the target image, leading to the generation of multiple ambiguous matching scenarios.

The first ambiguous matching scenario is presented in Fig. 2, where t_3 is a feature point of the original image T and v_4 is a feature point of the target image V. We assume that the optimal matching point of t_3 from T is the feature point v_4 in the target image V. However, there may be another feature point t_6 or t_7 whose optimal matching point is also the feature point v_4. The ambiguous matching generates a mapping relation so that multiple feature points from the original image correspond to one feature point in the target image, known as one-to-many matching.

The matching point pairs obtained by the previous method contain a large number of mismatches, which increases computation time and reduces accuracy when calculating the mapping relation using RANSAC.

The second ambiguous matching scenario is presented in Fig. 3. If the feature point t_3 in the original image T corresponds to the feature point v_4 in the target image V, denoted as $(t_3, \rightarrow, v_4)$, the matching point pair may be used to calculate the mapping relation.

In the other case, the original and target images may switch their roles. As a feature point from the original image V, the optimal matching point of v_4 may be the feature point t_7. It is obvious that the mapping relation calculated using $(t_3, \rightarrow, v_4)$ is not the inverse transformation of the mapping relation calculated using $(t_7, \leftarrow, v_4)$; this situation is known as inconsistent matching.
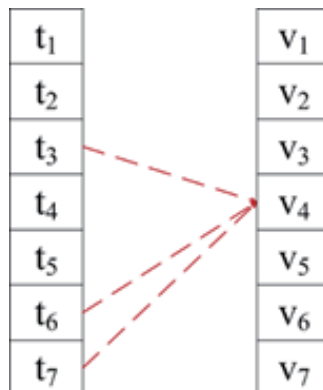

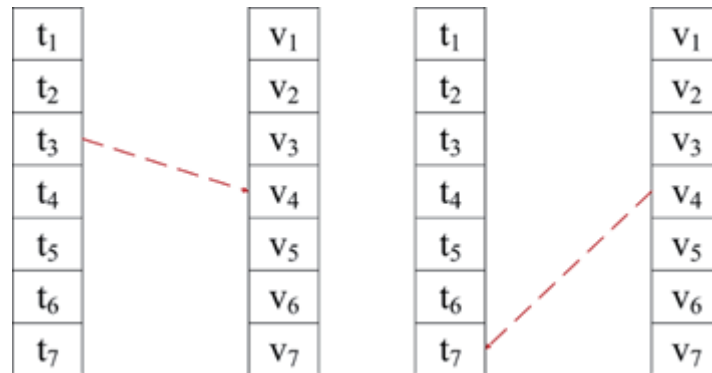Fig. 2.   (Color online) Ambiguous matching: one-to-many matching.

Fig. 3. (Color online) Ambiguous matching: inconsistent matching.

In general, the previous method set the feature points t_3 and v_4 as an optimal matching point pair because of the assumption that the similarity calculation between two feature points is correct, and it was not verified that the correspondence is correct in reverse. Thus, the asymmetry of matching feature points leads to the second ambiguous matching scenario.

3.4 Bidirectional optimal matching

To decrease the ratio of mismatches in the retained matching point pairs, for each matching point pair $(t_x, \rightarrow, v_y)$, we check whether t_x is the optimal matching point for v_y. $(t_x, \leftrightarrow, v_y)$ denotes the bidirectional optimal matching point pair, which satisfies both constraint conditions of $(t_x, \rightarrow, v_y)$ and $(t_x, \leftarrow, v_y)$, as shown in Fig. 4.

In the stage of feature point matching, for each feature point t_x in the original image, we calculate its similarity with all the feature points in the target image, and the optimal matching point pair $(t_x, \rightarrow, v_y)$ can be obtained if it exists. For each optimal matching point pair $(t_x, \rightarrow, v_y)$, we check whether $(t_x, \leftarrow, v_y)$ exists. The bidirectional optimal matching point pairs will contain a small percentage of mismatches.

4. Topology Consistency

RANSAC can sample from the matching point pairs retained from the above method. However, it is a waste of computing resources to estimate parameters using the subset consisting of matching point pairs not subject to the same mapping relation. Topology consistency is proposed to divide the matching point pairs into several consensus sets, in which the matching point pairs may be subject to the same image transformation.

4.1 Orientation consistency

Fundamental assumption 1: If $(t_i, \leftrightarrow, v_k)$ and $(t_j, \rightarrow, v_l)$ are two matching point pairs subject to the same image transformation, the rotation angle of $(t_i, \leftrightarrow, v_k)$ should be close to that of
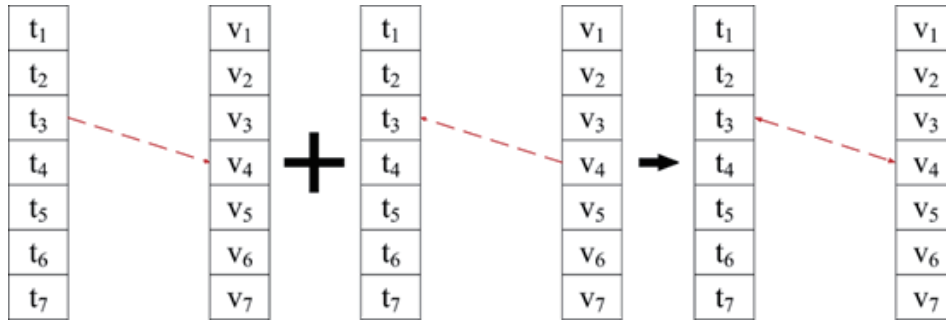
Fig. 4.   (Color online) Bidirectional optimal matching.

$(t_j, \rightarrow, v_l)$. The feature point sets of the original image are denoted as T and those of the target images are denoted as V, as shown in Fig. 5. Supposing that there are two matching point pairs $(t_i, \leftrightarrow, v_k)$ and $(t_j, \rightarrow, v_l)$, t_i and t_j are feature points from the original image, whereas v_k and v_l are feature points from the target image.

First, we calculate the rotation angle between the orientations of the two feature points for each bidirectional optimal matching point pair, denoted as Δ. If Δ is similar for two pairs, the two matching point pairs may satisfy the same transformation, and we consider that the two matching point pairs are consistent in terms of the rotation angle.

$$\Delta_{oik} = \alpha_i - \alpha_k \bmod 360 \qquad (7)$$

$$\Delta_{tjl} = \alpha_j - \alpha_l \bmod 360 \qquad (8)$$

As shown in Fig. 6, if the difference in Δ calculated from two matching point pairs is very large, the two matching point pairs cannot be generated by the same transformation, and we consider that the two matching point pairs are not consistent in terms of the rotation angle. Then, we consider the difference histogram of rotation angles for all matching point pairs between the original image and the target image. The orientation histogram has 36 bins covering the 360° range of rotation angles. Each rotation angle of one matching point pair can cast a vote for the nearest angle bin. Then, we select the peak of the orientation histogram, denoted as Δ_{top}. This peak corresponds to the rotation angle with the maximum probability of rotating the target image to the original image. A parabola is fit to three values in the difference histogram, namely, the peak value and the two values of the bins closest to the peak, to interpolate the peak position for a higher accuracy.

We set a threshold of the orientation difference, denoted as d_Δ. Then, the difference in orientation is limited to $(\Delta_{top} - d_\Delta, \Delta_{top} + d_\Delta)$. For each matching point pair, if the rotation angle satisfies $\Delta \in (\Delta_{top} - d_\Delta, \Delta_{top} + d_\Delta)$, the matching point pair is treated as a candidate sample of the consensus set corresponding to Δ_{top}. Other local peaks can also be treated as the rotation angles of consensus sets.
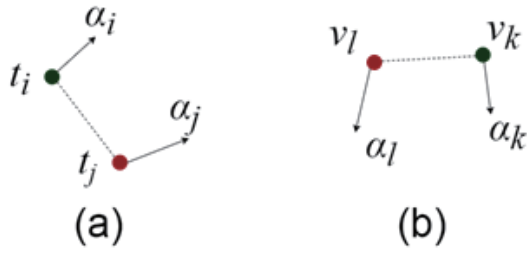
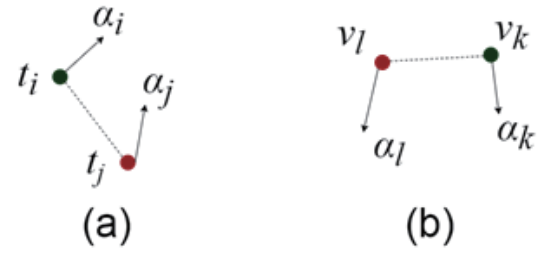Fig. 5.   (Color online) (a) Original and (b) target images with similar rotation angles.

Fig. 6.   (Color online) (a) Original and (b) target images with dissimilar rotation angles.

4.2   Distance consistency

Fundamental assumption 2: If $(t_i, \leftrightarrow, v_k)$ and $(t_j, \rightarrow, v_l)$ are two matching point pairs subject to the same image transformation, the distance between t_i and t_j should be close to that between v_k and v_l.

We calculate the distance set of the original and target images. We assume that there are two matching point pairs $(t_i, \leftrightarrow, v_k)$ and $(t_j, \rightarrow, v_l)$, the coordinates of t_i and t_j in the original image are (x_i, y_i) and (x_j, y_j), and the coordinates of v_k and v_l in the original image are (x_k, y_k) and (x_l, y_l), respectively.

As shown in Fig. 7, L_{ij} denotes the distance between t_i and t_j in the original image, and L_{kl} denotes the distance between v_k and v_l in the target image. L_{ij} is calculated as

$$L_{ij} = \sqrt{\left(x_i - x_j\right)^2 + \left(y_i - y_j\right)^2}, \tag{9}$$

and L_{kl} is calculated as

$$L_{kl} = \sqrt{\left(x_k - x_l\right)^2 + \left(y_k - y_l\right)^2}. \tag{10}$$

Given the threshold T_L, if $|L_{ij} - L_{kl}| < T_L$, the two matching point pairs $(t_i, \leftrightarrow, v_k)$ and $(t_j, \rightarrow, v_l)$ are subject to the consistency constraint of their distance.

4.3   Azimuth consistency

Fundamental assumption 3: If $(t_i, \leftrightarrow, v_k)$ and $(t_j, \rightarrow, v_l)$ are two matching point pairs subject to the same image transformation, the relative position between t_i and t_j should be close to that between v_k and v_l.

The azimuth angle from t_i to t_j is defined as the rotation angle from the orientation of t_i to the vector from t_i to t_j. For two coincident matching point pairs, their azimuth angles are close to each other.
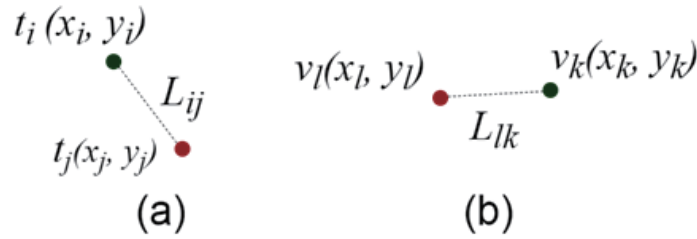
Fig. 7.   (Color online) (a) Original and (b) target images with similar distances.

Supposing that there are two matching point pairs $(t_i, \leftrightarrow, v_k)$ and $(t_j, \rightarrow, v_l)$, the coordinates of t_i and t_j in the original image are (x_i, y_i) and (x_j, y_j), and those of v_k and v_l in the original image are (x_k, y_k) and (x_l, y_l), respectively. The orientations of these four feature points are denoted as α_i, α_j, α_k, and α_l, respectively. The vector in the original image is denoted as $<t_i, t_j>$ and that in the target image is denoted as $<v_k, v_l>$. The direction angle of the vector $<t_i, t_j>$, denoted as θ_{ij}, is calculated as

$$\theta_{ij} = \tan^{-1} \frac{y_j - y_i}{x_j - x_i}. \tag{11}$$

The azimuth angle from t_i to t_j, denoted as β_{ij}, is calculated as

$$\beta_{ij} = \left(\theta_{ij} - \alpha_i\right) \bmod 360. \tag{12}$$

Similarly, the direction angle of the vector $<v_k, v_l>$, denoted as θ_{kl}, is calculated as

$$\theta_{kl} = \tan^{-1} \frac{y_l - y_k}{x_l - x_k}. \tag{13}$$

The azimuth angle from v_k to v_l, denoted as β_{kl}, is calculated as

$$\beta_{kl} = \left(\theta_{kl} - \alpha_k\right) \bmod 360. \tag{14}$$

Given the threshold T_β, the two matching point pairs are consistent if

$$\left|\beta_{ij} - \beta_{kl}\right| < T_\beta. \tag{15}$$

The scenarios for similar and dissimilar matching point pairs are shown in Figs. 8 and 9, respectively.

Finally, the matching point pairs can be divided into several consensus sets, each of which only contains matching point pairs that may be subject to the same image transformation. Each consensus set is considered as a sample space for RANSAC that seldom contains any outliers.
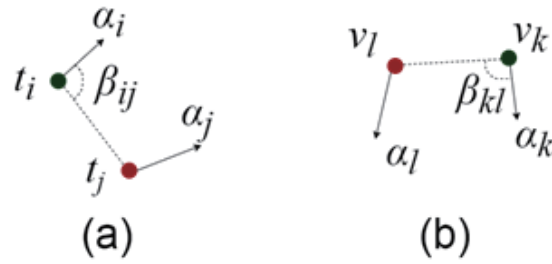
Fig. 8.   (Color online) (a) Original and (b) target images with similar azimuths.
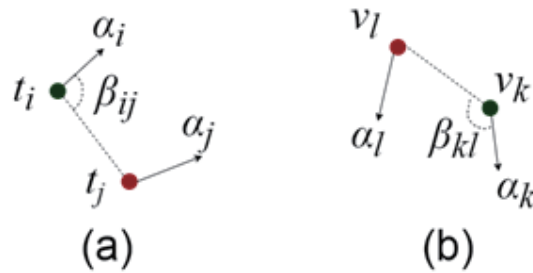

Fig. 9.   (Color online) (a) Original and (b) target images with dissimilar azimuths.

As a result, only a few iterations are needed to obtain the optimal mapping relation model in each consensus set or to prove that a consensus set does not correspond to any mapping relation.

5.   Homography Matrix Estimation

5.1   Calculation of homography matrix

We assume that the homography matrix can be described as

$$
\boldsymbol{H} = \begin{bmatrix} h_{11} & h_{12} & h_{13} \\ h_{21} & h_{22} & h_{23} \\ h_{31} & h_{32} & h_{33} \end{bmatrix}.
\tag{16}
$$

For each matching point pair, we use the homography matrix to map the point from the target image to the original image.

$$
\begin{bmatrix} x^{(v)} \\ y^{(v)} \\ 1 \end{bmatrix} = \begin{bmatrix} h_{11} & h_{12} & h_{13} \\ h_{21} & h_{22} & h_{23} \\ h_{31} & h_{32} & h_{33} \end{bmatrix} \begin{bmatrix} x^{(t)} \\ y^{(t)} \\ 1 \end{bmatrix}
\tag{17}
$$

Homogeneous coordinates [i.e., $(x, y, 1)$] are used to represent points on the plane, so that h_{33} can be set to 1 without loss of generality. The above equation can be expressed as

$$\begin{bmatrix} -x^{(t)} & -y^{(t)} & -1 & 0 & 0 & 0 & x^{(t)}x^{(v)} & y^{(t)}x^{(v)} & x^{(v)} \\ 0 & 0 & 0 & -x^{(t)} & -y^{(t)} & -1 & x^{(t)}y^{(v)} & y^{(t)}y^{(v)} & y^{(v)} \end{bmatrix} h = 0, \tag{18}$$

where

$$\boldsymbol{h} = (h_{11}, h_{12}, h_{13}, h_{21}, h_{22}, h_{23}, h_{31}, h_{32}, 1)^T. \tag{19}$$

For the set

$$A = \begin{bmatrix} -x^{(t)} & -y^{(t)} & -1 & 0 & 0 & 0 & x^{(t)}x^{(v)} & y^{(t)}x^{(v)} & x^{(v)} \\ 0 & 0 & 0 & -x^{(t)} & -y^{(t)} & -1 & x^{(t)}y^{(v)} & y^{(t)}y^{(v)} & y^{(v)} \end{bmatrix}, \tag{20}$$

we can form the following linear system of equations from a set of corresponding points:

$$A\boldsymbol{h} = 0, \tag{21}$$

where $A \in R^{2N \times 9}$ with N being the number of matching pairs (e.g., $N = 1$). There are only eight variables, so only four pairs of matching points are needed to calculate $\boldsymbol{H}$ by solving systems of linear equations. In each iteration of the conventional RANSAC, four matching point pairs are selected randomly to calculate $\boldsymbol{H}$.

Different from the conventional RANSAC, we take all the remaining pairs of matching points in the consensus set to calculate $\boldsymbol{H}$; thus, in each iteration, N may be greater than four. We consider that $Ax = 0$, which is a homogeneous linear least squares problem similar in appearance to the inhomogeneous linear least squares problem $Ax = b$. Then, we solve the problem by singular value decomposition (SVD).

$$A = U \sum V^T \tag{22}$$

From the SVD, we take the last column of V as the solution $\boldsymbol{h}$, which contains the coefficients of the homography matrix that best fit the points. We reshape $\boldsymbol{h}$ into the matrix $\boldsymbol{H}$.

5.2 Calculation of projection deviations

Assuming that $(t_i, \leftrightarrow, v_k)$ is a pair of matching points remaining in the current consensus set when the matrix $\boldsymbol{H}$ is obtained, the feature point t_i $(x_i^{(t)}, y_i^{(t)})$ from the original image can be projected to the target image, denoted as t_i' (x_i', y_i').

$$\begin{bmatrix} x_i' \\ y_i' \\ 1 \end{bmatrix} = \begin{bmatrix} h_{11} & h_{12} & h_{13} \\ h_{21} & h_{22} & h_{23} \\ h_{31} & h_{32} & h_{33} \end{bmatrix} \begin{bmatrix} x_i^{(t)} \\ y_i^{(t)} \\ 1 \end{bmatrix} \tag{23}$$

The projection deviation is calculated as

$$pd_{ik} = \sqrt{(x_k^{(v)} - x_i')^2 + (y_k^{(v)} - y_i')^2}. \tag{24}$$

If the maximum of $\{pd_{ik}\}$ is greater than a given threshold, the corresponding matching point pair is removed from the current consensus set and the homography matrix should be recalculated on the basis of the consensus set. If the number of matching point pairs is less than four, the current consensus set is abandoned.

If the maximum of $\{pd_{ik}\}$ is less than the given threshold, the obtained homography matrix H can be the transformational model for the consensus set.

5.3   Iterative process

On the basis of the bidirectional optimal matching points and topology consistency, we reserve a sufficient number of feature point pairs. This means that only a few feature points need to be calculated. To reduce the search space and speed up the computation, we perform iterations, deleting the feature point with the greatest error in each iteration, after which the direct mapping relation between the remaining feature points sets is calculated.

As shown in Fig. 10, the steps of our method are as follows:

Step 1:   Through the bidirectional optimal matching points and topology consistency, there is a large set of feature points that meet the condition of consistency.

Step 2:   Select a group of feature points randomly and use all matching point pairs in the set to calculate the homography matrix H through the least squares method.

Step 3:   Using the homography matrix H, calculate the projection deviation of all matching point pairs in the set.

Step 4:   If the maximum deviation of matching point pairs is greater than the given threshold, delete the corresponding matching point pair.

Step 5:   Recalculate the model parameters with the remaining matching point pairs in the current set.

Step 6:   Iterate the process from Step 2 to Step 5 until the algorithm terminates.

There are two conditions for the termination of our method. One is that the number of matching point pairs in the current set is less than four, indicating that no mapping relation can be obtained from the current set; the other is that the projection deviation of the remaining matching point pairs is less than the given threshold, which means that the optimal mapping relation of the current set has been obtained.
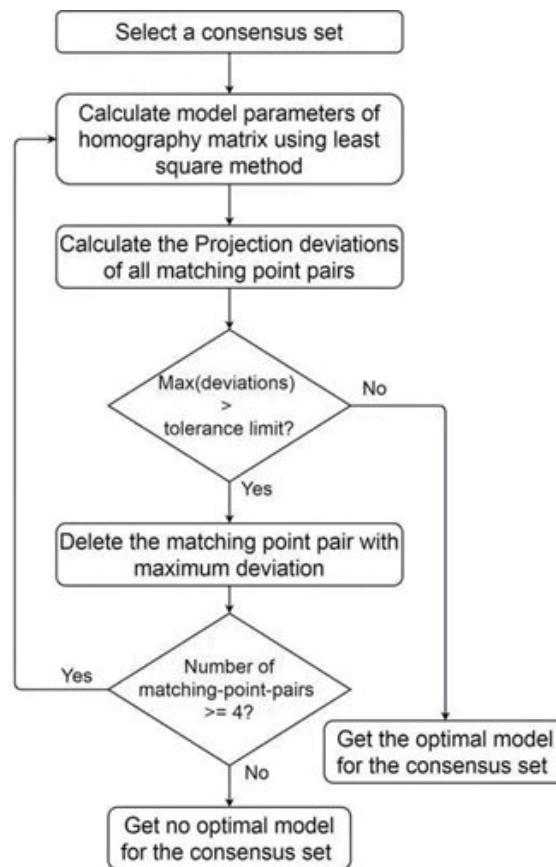
Fig. 10.   Our proposed method.

There are multiple sets meeting the condition of consistency between matching point pairs in the two images, one of which represents the homography matrix ***H***. The key difference from the conventional RANSAC is that there is only one mapping relation between the two images.

Once the homography matrix between images is obtained, one image can be projected onto another through this matrix to complete the basic mosaic. In RANSAC, the mapping relation of matching point pairs with the maximum inlier ratio is selected, while the fact that there is only one homography matrix between two of the same images is ignored.

In Fig. 11, the original and target images are different but have the same mapping relation. When using RANSAC to calculate the homography matrix, the optimal mapping relation is that located at the bottom of the original image. Nevertheless, the original image is different from the target image. Because more than one mapping relation exists between the original image and the target image, the method proposed in this paper considers them to be different.

6.   Results and Discussion

We tested the proposed method on three datasets of local area fingerprints provided by mobile phone and fingerprint module manufacturers, which were collected with a capacitive sensor, as shown in Table 1. Owing to the cost of the hardware, each collected fingerprint image
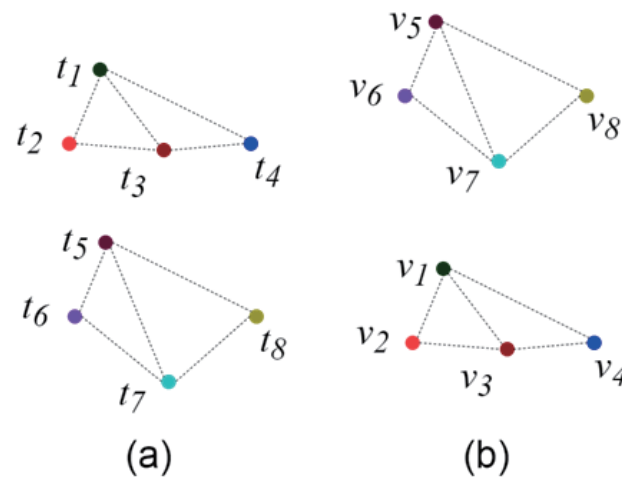
Fig. 11.   (Color online) The conventional RANSAC can calculate multiple mapping relations for (a) original and (b) target images.

Table 1
Three datasets of small-area fingerprints.

	Persons	Fingers for each person	Image for each finger	Image size
Dataset 1	50	6	28	112 × 88
Dataset 2	35	6	136	132 × 114
Dataset 3	30	10	60	64 × 80

generally covers a small area and contains only the local information of a fingerprint, which is often called a small-area or local fingerprint.

In dataset 1, fingerprints of 50 people are collected from six fingers for each person. From each finger, 28 images are collected in different regions (with overlapping between different regions to enable small fingerprints to be stitched into large fingerprints). A total of 8400 (50 × 6 × 28) images with a size of 112 × 88 are obtained. In the algorithm test, two images are randomly matched once, and the total number of matchings is $C_{8400}^{2} = 70551600$.

In dataset 2, fingerprints of 35 people are collected from six fingers for each person. From each finger, 136 images are collected in different regions (with overlapping between different regions to enable small fingerprints to be stitched into large fingerprints). A total of 28560 (35 × 6 × 136) images with a size of 132 × 114 are obtained. In the algorithm test, two images are randomly matched once, and the total number of matchings is $C_{28560}^{2} = 815645040$.

In dataset 3, fingerprints of 30 people are collected from ten fingers for each person. From each finger, 60 images are collected in different regions (with overlapping between different regions to enable small fingerprints to be stitched into large fingerprints). A total of 18000 (30 × 10 × 60) images with a size of 64 × 80 are obtained. In the algorithm test, two images are randomly matched once, and the total number of matchings is $C_{18000}^{2} = 323982000$.

6.1   Sample from dataset 1

From dataset 1, we choose two images as shown in Fig. 12. The matching point pairs obtained by the exhaustion method and Lowe's method are shown in Fig. 13. The matching point pairs obtained by the proposed method and classified by the consistency are shown in Fig. 14.
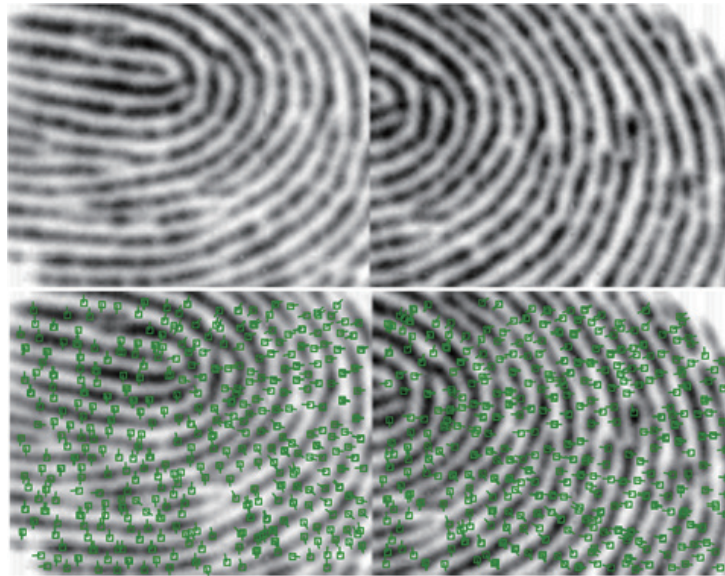

Fig. 12.   (Color online) Original image (upper left) and target image (upper right); feature points from original image (lower left) and target image (lower right) obtained by SIFT.
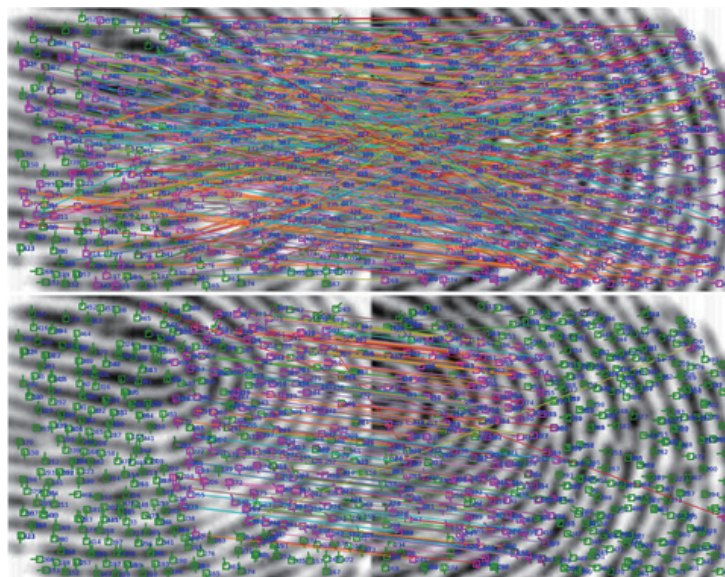

Fig. 13.   (Color online) Matching point pairs obtained by exhaustion method (upper) and Lowe's method (lower).
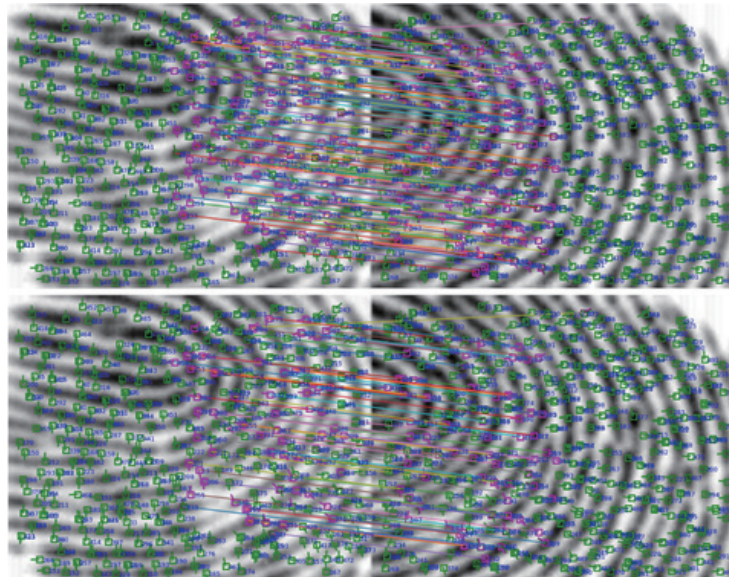
Fig. 14.   (Color online) Matching point pairs obtained by the proposed method (upper) and in all consensus sets (lower).
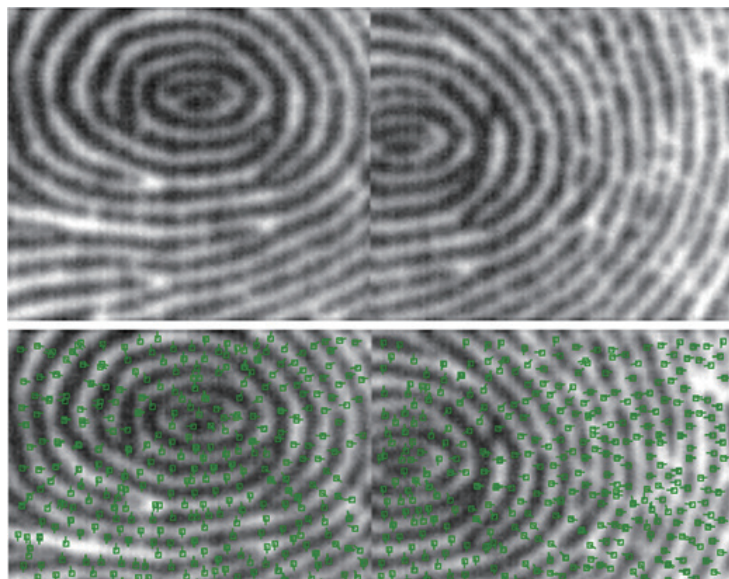

Fig. 15.   (Color online) Original image (upper left) and target image; feature points from original image (lower left) and target image (lower right) obtained by SIFT.

6.2   Sample from dataset 2

From dataset 2, we choose two images as shown in Fig. 15. The matching point pairs obtained by the exhaustion method and Lowe's method are shown in Fig. 16. The matching point pairs obtained by the proposed method and divided by the consistency are shown in Fig. 17.
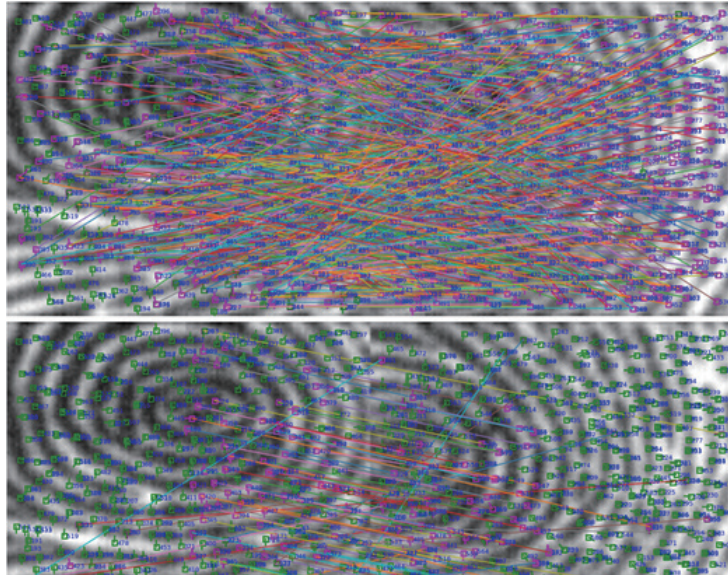
Fig. 16.   (Color online) Matching point pairs obtained by exhaustion method (upper) and Lowe's method (lower).
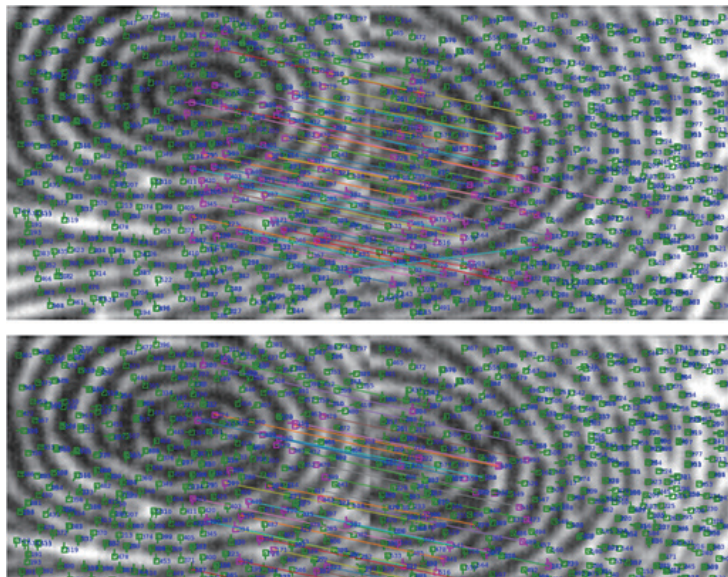

Fig. 17.   (Color online) Matching point pairs obtained by the proposed method (upper) and in all consensus sets (lower).

6.3   Sample from dataset 3

From dataset 3, we choose two images as shown in Fig. 18. The matching point pairs obtained by the exhaustion method and Lowe's method are shown in Fig. 19. The matching point pairs obtained by the proposed method and divided by the consistency are shown in Fig. 20.
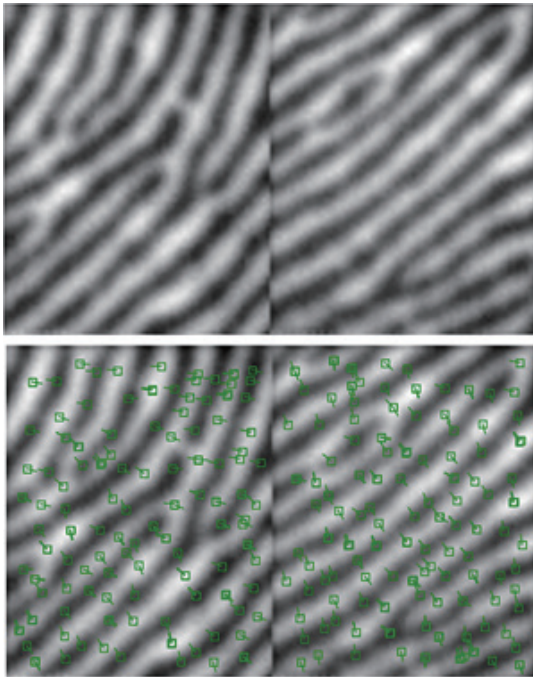
Fig. 18. (Color online) Original image (upper left) and target image (upper right); feature points from original image (lower left) and target image (lower right) obtained by SIFT.
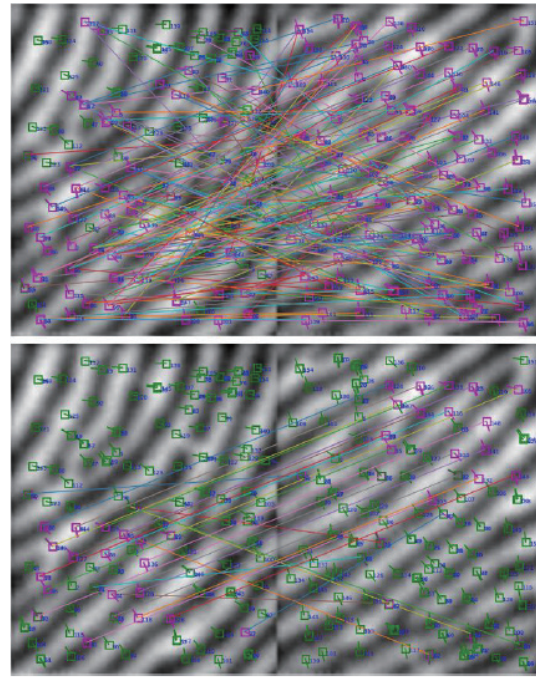

Fig. 19. (Color online) Matching point pairs obtained by exhaustion method (upper) and Lowe's method (lower).
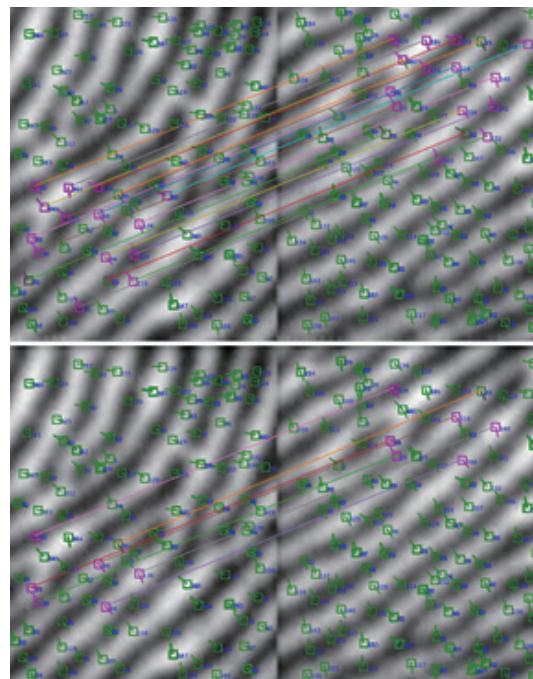

Fig. 20. (Color online) Matching point pairs obtained by the proposed method (upper) and in all consensus sets (lower).

Table 2
Matching point pairs obtained by different methods.

	Exhaustion method	Lowe's method	Bidirectional optimal match	Consensus set
Dataset 1	492	144	134	73
Dataset 2	582	90	61	32
Dataset 3	161	29	19	8

Table 3
Iteration times before obtaining the optimal homography matrix by different methods.

	RANSAC	Improved RANSAC in Ref. 16	DL-RANSAC	Proposed method
Dataset 1	26122320	70	63	1
Dataset 2	863040	29	21	1
Dataset 3	1680	4	7	1

The numbers of matching point pairs obtained by the different methods in the experiments are shown in Table 2. The last two columns are obtained by our proposed methods. Different methods for estimating parameters of the transformational model based on matching point pairs were compared. To ensure fairness, the consensus sets obtained by the proposed method were used for each experiment. The results are shown in Table 3.

As described in Sect. 2, the improved RANSAC in Refs. 13–15 can be considered as a preprocessing step for RANSAC. In the stage of estimating parameters based on our proposed consensus set, the improved methods in Refs. 13–15 have the same number of iterations as the conventional RANSAC.

7.   Conclusion

The ratio of mismatching in matching point pairs selected by the bidirectional optimal matching proposed in the paper is much lower than in the exhaustion method and Lowe's method. By sampling from each consensus set, which is divided on the basis of the consistency of matching point pairs, RANSAC requires only a few iterations to obtain the desired model or eliminate a subset. The number of iterations of the proposed method is much less than that of RANSAC, and no more than one mapping relation between two images can be accepted, which makes the method presented in this paper superior to RANSAC in accuracy and computation speed.

Availability of Data and Materials

Some or all of the data, models, and code generated or used during the study are proprietary or confidential in nature and may only be provided with restrictions (e.g., anonymized data).

Competing Interests

The authors declare no conflicts of interest in this paper.

Acknowledgments

References

1 H. AlShehri, M. Hussain, H. AboAlSamh, and M. AlZuair: Sensors **18** (2018) 1008. https://doi.org/10.3390/s18041008

2 A. Alrashidi, A. Alotaibi, M. Hussain, H. AlShehri, H. A. AboAlSamh, and G. Bebis: Sensors **21** (2021) 3657. https://doi.org/10.3390/s21113657

3 Y. Chen, G. Li, S. Jhong, P. Chen, C. Tsai, and P. Chen: Sens. Mater. **32** (2020) 3157. https://doi.org/10.18494/SAM.2020.2838

4 C. Lin, C. Wu, C. Sun, and C. Lin: Sens. Mater. **32** (2020) 3137. https://doi.org/10.18494/SAM.2020.2771

5 C. Hsia, S. Yen, and J. Jang: Sens. Mater. **31** (2019) 1803. https://doi.org/10.18494/SAM.2019.2351

6 C. Hsia: IEEE Sens. J. **18** (2018) 790. https://doi.org/10.1109/JSEN.2017.2772799

7 D. G. Lowe: Int. J. Comput. Vision **60** (2004) 91. https://doi.org/10.1023/B:VISI.0000029664.99615.94

8 H. Bay, T. Tuytelaars, and L. V. Gool: Comput. Vision Image Understanding **110** (2008) 346. https://doi.org/10.1016/j.cviu.2007.09.014

9 M. Calonder, V. Lepetit, C. Strecha, and P. Fua: Proc. Computer Vision – ECCV 2010. (ECCV, 2010) 778–792.

10 E. Rublee, V. Rabaud, K. Konolige, and G. Bradski: Proc. 2011 Int. Conf. Computer Vision. (IEEE, 2011) 2564–2571.

11 A. C. Berg, T. L. Berg, and J. Malik: Proc. 2005 IEEE Computer Society Conf. Computer Vision and Pattern Recognition. (IEEE, 2005) 26–33.

12 E. Karami, S. Prasad, and M. Shehata: Proc. 2015 Newfoundland Electrical and Computer Engineering Conf. (NECEC, 2015) 1710–1714.

13 J. Liu, J. M. White, and R. M. Summers: Proc. 2010 IEEE Int. Conf. Image Processing. (IEEE, 2010) 841–844.

14 M. A. Fischler and R. C. Bolles: Readings in Computer Vision, M. A. Fischler and O. Firschein, Eds. (Morgan Kaufmann, San Francisco, 1987) Chap. 7.

15 J. Zhao, H. Liu, Y. Feng, S. Yuan, and W. Cai: Proc. 2015 IEEE Int. Conf. Computer and Information Technology; Ubiquitous Computing and Communications; Dependable, Autonomic and Secure Computing; Pervasive Intelligence and Computing, Liverpool (IEEE, 2015) 568–574.

16 B. Zhong and Y. Li: Proc. 2019 IEEE 4th Int. Conf. Image, Vision and Computing (ICIVC, 2019) 489–493.

17 O. Chum and J. Matas: Proc. 2005 IEEE Computer Society Conf. Computer Vision and Pattern Recognition (CVPR'05) 220–226.

18 G. Shi, X. Xu, and Y. Dai: Proc. 2013 5th Int. Conf. Intelligent Human-Machine Systems and Cybernetics (IEEE, 2013) 474–477.

19 M. Fotouhi, H. Hekmatian, M. A. Kashani-Nezhad, and S. Kasaei: Multimed. Tools Appl. **78** (2019) 9429. https://doi.org/10.1007/s11042-018-6475-6

20 J. Liu and F. Bu: J. Eng. **2019** (2019) 9118. https://doi.org/10.1049/joe.2018.9198

21 M. Rahman, X. Li, and X. Yin: Proc. 2019 IEEE 4th Int. Conf. Image, Vision and Computing (ICIVC, 2019) 463–468.

About the Authors

Aihua Wu is a Ph.D. student of Hohai University and his research interest is in pattern recognition.

Weizheng Chen received his MA.Sc degree from Tsinghua University and his research interests include fingerprint matching, pattern recognition, and computer vision.

Yijie Bian is a professor of Hohai University and his research interests are in information management and information services.

Song Xue is an associate professor of Hohai University and his research interest is in information management.