Pattern Recognition 150 (2024) 110293

Contents lists available at ScienceDirect

Pattern Recognition

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

Local topology similarity guided probabilistic sampling for mismatch removal

Zaixing He, Chentao Shen, Xinyue Zhao *

School of Mechanical Engineering, The State Key Lab of Fluid Power & Mechatronic Systems, Zhejiang University, Hangzhou 310058, PR China

ARTICLE INFO

Keywords:
Topological stability
Local topology similarity
Probability sampling consensus
Mismatch removal

ABSTRACT

Feature point matching between two images is a fundamental and important process in machine vision. In many cases, mismatches are inevitable, and removing mismatches is an indispensable task. The existing methods attempt to find comprehensive constraints or sampling model to achieve better performance, which results in the increasingly complexity and may cause the weakness of the generality and scalability. To address this issue, a method called Local Topology similarity guided probabilistic Sampling consensus (LTS) is proposed. It constructs a topological network, then quantifies the mismatch probability in a concise approach based on comparing the topological relationship with neighbourhoods. Then, it detects and removes the mismatches by sampling guided by the mismatch probability. Compared with the state-of-the-art methods, LTS has an excellent performance in accuracy and robustness.

1. Introduction

Feature point matching is an important topic in machine vision, which is widely used in object detection, image registration, tracking, pose estimation, etc. In the task of feature point matching, it firstly extracts the feature points to describe local information, for which many methods have been proposed such as SIFT (scale invariant feature transform) [1], SURF (Speeded up robust features) [2], ORB (Oriented FAST and Rotated BRIEF) [3], or deep learning-based as LIFT (Learned Invariant Feature Transform) [4], SuperPoint [5], and then it establishes a one-to-one matching relationship between feature points in different images. However, though these methods are becoming more and more accurate, it is inevitable to produce mismatches, especially in some complex cases, such as the texture of the object is not obvious or locally similar, or in the cases with strong external interference. The correctness of matching is the key to the realization of these visual tasks. Therefore, the removal of mismatching is of great importance.

The existing methods to remove mismatches can be mainly divided into resampling-based, geometry-based, learning-based methods, combining methods. In resampling-based methods, points are sampled to obtain a transformation model. Geometry-based methods use geometry constraints between matching points, they are always popular and the performance of them is progressively improving. Learning-based method make judgements on matches by machine learning, which could deal with challenging issues better. The combining methods try to take benefits of both, instead of simply compose two kinds of method, which is not easy to achieve.

In order to get more excellent result, the current study of resampling-based methods mainly focus on promotion of convergence efficiency and probability to proposed more complex sampling framework, and geometry-based methods also exhibit a tendency to entail numerous intricate geometric constraints to adapt more conditions. However, with the increasing complexity, the methods may lose the generality and scalability. Further, in pragmatic scenarios, we often prefer approaches that are easy to execute with uncomplicated principle, and also yield excellent results at the same time.

For this case, a method called Local Topology similarity guided probabilistic Sampling consensus (LTS) is proposed. The proposed method calculates the mismatch probability based on topological similarity of each matching point in a concise way, and weights the possibility into the process of sampling to obtain the transform matrix, which makes the sample process converge quickly and also improves the efficiency and accuracy. Experiments show that the proposed method is of more robustness and accuracy compared with other state-of-art methods.

The contributions of this paper can mainly be summarized as follows:

(1) We propose a lightweight method for computing mismatch probability of each point pair based on comparing the local topology

* Corresponding author.
E-mail address: zhaoxinyue@zju.edu.cn (X. Zhao).

https://doi.org/10.1016/j.patcog.2024.110293
Received 13 January 2023; Received in revised form 3 August 2023; Accepted 16 January 2024
Available online 19 January 2024

relationship in two images. Different from the increasingly complicated existing methods of describing characteristics of local consensus, the proposed method describes the local topology using the simplest approach to extract the more essential features, avoiding complex operations by angle or distance constraints in traditional geometry.

The topological structure of correct matched points exhibits local consensus, which is widely accepted as an important characteristic for mismatch removal. Utilizing this characteristic becomes the focus of many methods. Most methods attempt to find comprehensive constraints, which results in the increasingly complexity of describing local consistency, in turn becoming a trend in the field. However, we consider that the more complex the method and the more constraints involved, the weaker its generalization. Therefore, we propose a method using a simple algorithm to describe the local topological structure to extract essential features but with stronger generalization.

(2) We propose a sampling-refined method for mismatch removal by combining local consensus with sampling tightly. Instead of the simply combining topology and sampling, our method has a strong complementary dependency between the local topology and sampling.

Although directly combining existing geometry-based and sampling-based methods can improve the effectiveness, it has not achieved the optimal combination effect. These methods consist of complicated calculation of mismatching probability, leading to the limited exploitation of the advantages of the sampling method, which also limits the overall performance of the method. In our method, we simplify the mismatching probability calculation based on local consensus, even though which results in a decrease in accuracy of mismatching probability, it can be fully compensated with sampling. Therefore, our method can achieve simplification of local consensus description, improved generalization with high performance in mismatching removal.

2. Related works

In this section, the existing mismatching removal theory and methods are summarized. The proposed mismatch removal methods can be divided into four categories as following.

2.1. Resampling-based methods

The basis of the resampling-based methods is the graphics principle that the correct matching points conform to a transform model, the outliers (the points do not conform to the model) are considered to be mismatches and be removed.

The resampling-based method is to find a transform model that makes the maximum number or more than a certain number of matches that meet which. The most popular method in the area of mismatching removal is using RANSAC (Random Sample Consensus) [6,7], it estimates the optimal transform matrix of two images, the outliers of the model is considered the mismatches. However, it will be inefficient when there are a large number of mismatches.

Therefore, various methods are proposed to improve RANSAC, including the optimization of sampling, optimization of loss function and optimization of model estimation typically.

Optimization of loss function: it proposes new loss function instead of the function of RANSAC, such as MLESAC (maximum likelihood SAC), which enhances the accuracy of the calculation [8], but it provides a limited improvement. Optimization of model estimation: such as R-RANSAC (Randomized RANSAC) and SPRT-RANSAC (Randomized RANSAC with sequential probability ratio test), it will judge whether it is the correct model first after finding the model, and will continue to sample and iterate if not [11,12].Optimization of sampling: it focuses on more effective ways to sample, such as PROSAC (progressive sampling consensus), which firstly obtains the probability of each data being an inlier, then preferentially extracts the data with high probability [9]. GroupSAC, it firstly groups all of matches, the group with more matching points is preferred when in the sampling [10]. These two methods are well known, which accelerate convergence to some extent. SESAC (Sequential Evaluation on Sample Consensus) sorts the matches based on the similarity of the corresponding features, then selects the samples sequentially, which perform better than PROSAC [14]. Gao et al., improved the RANSAC by taking pre-validation and resampling during iterations, which accelerates efficiency [15]. As for the optimization in the various aspects mentioned above, Rahul et al. summarized and proposed a universal framework, USAC, which obtains the advantages of some previous methods [37].

Recently, more advanced sampling methods are proposed. DL-RANSAC (Descendant Likelihood-RANSAC) introduces descending likelihood to reduce the randomness so that it converges faster [13]. MAGSAC (Marginalizing Sample Consensus) proposed new quality function with no need for a user-defined noise scale [36].

The resampling-based method is widely used in rigid feature matching. Generally, it performs well in ordinary scenes, where the mismatch ratio is not high. However, it is challenging for it to deal with the high mismatch cases.

2.2. Geometry-based methods

Many researchers focus on the geometry of the matching points to construct the geometric or topological constraint between the matching points to remove mismatch.

GTM (Graph Transformation Matching) [16], proposed by Aguilar et al., is a typical method based on geometry, which continually constructs the KNN undirected graph based on matching points and removes the mismatches until two images have similar graph, however it will be inefficient with large number of matches. GMS (Grid-based motion statistics) [17], based on theory of motion statistics, focus on the image pairs with high-speed transform.

In [18], we proposed a robust method based on comparing triangular topology and distance constraint of feature points. Luo et al., analyzed the relationship of Euclidean distance between the matching points then correct the mismatch based on angular cosine [20]. Zhao et al., remove the mismatches according to the constraints that matching distances tend to be consistent [21].

These geometry-based methods can detect mismatches more efficiently compared with resampling-based method because it avoids the iteration of sample. However, it may fail to construct the correct constraints in the cases of high mismatch ratio.

Additionally, some have focused on studying local topological consensus to remove mismatches. LPM (Locality Preserving Matching) [25,40], proposed by Ma et.al., representative methods of geometry-based also, which uses the principle that the local neighborhood structures of true matches will maintain in different images, they interpolate vector fields between the two-point sets and removes mismatches according to consensus of local points, which achieves good performance and also has low complexity. LGSC (Local Graph Structure Consensus) [28], PSC (Progressive Smoothness Consensus) [30] LOGO (Locality-guided Global-preserving Optimization) [29] are proposed based on the idea of preserving of local structure, which further improved the effectiveness of mismatch removal. And in [42], a method based on topological clustering for wide-baseline mismatch removal is proposed, which concludes the topological consensus in wide-baseline match, which performs well in that case. Especially in image registration, the pervious work [39] extends the LPM [25] for firstly computing a set of good matches to guide removal in whole matching set. In [41] transformation is constrained with manifold regularization and used to learning to remove mismatches. Generally, most of methods based on local topological consensus have achieved better performances using more comprehensive constraints, which results in the increasingly complexity of describing local consistency.

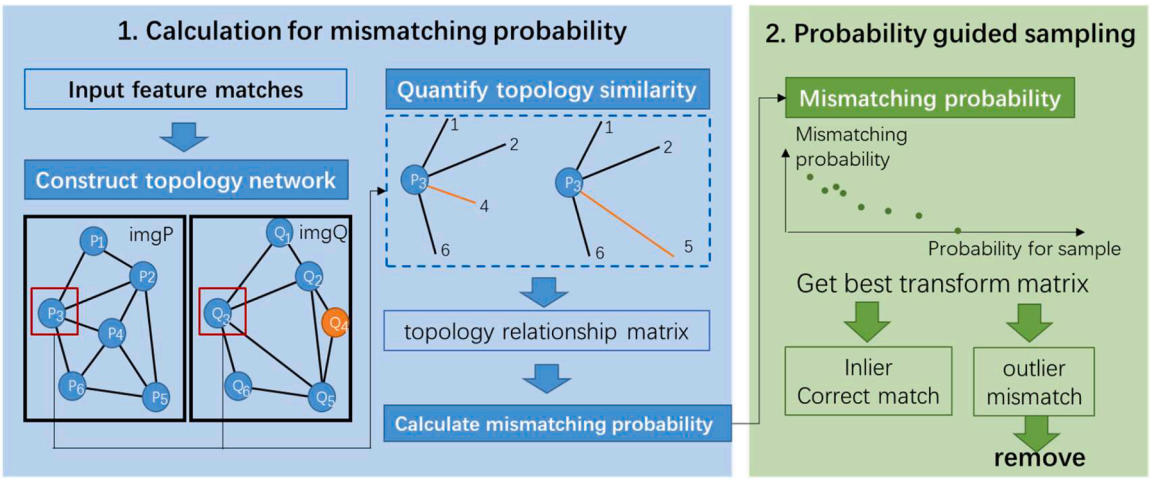

Fig. 1. The summarization of proposed method.

2.3. Learning-based methods

Learning-based matching approaches utilize the benefits of machine learning to generate geometric features and enhance the efficacy of mismatch removal.

Yi et al. [31] initially undertook an effort to remove mismatches with depth neural network, which is based on multilayer perception for binocular vision. Nonetheless, the effectiveness of this relies upon the input of camera intrinsic parameters. SuperGlue [32] combines the inferred matching and outlier removal based on Graphic Neural Network (GNN), it has a good result combined with SuperPoint [5]. Recently, PGFNet [33] introduce a novel iterative filtering structure while MS2DG-net [43] capture sparse semantics similarity between two given images to remove mismatches and get the camera position. Zhang et al. [34] used order-aware networks (OAnet) to get probabilities of being correctly corresponded and regresses relative pose.

However, these learning-based methods rely on training data greatly, which impacts the generality. Ma et al. [35] proposed a two-class classifier LMR for removing mismatch data with linear time complexity, which has higher generality.

In spite of the effectiveness of learning-based methods, practical visual tasks often meet with indefinite issues. In addition, these methods essentially belong to extraction and processing of geometric information, the approaches in Section 2.2 have a more precise computation process, while computation of learning-based methods is done in a network with numerous parameters.

2.4. Geometry and resampling-combining methods

Recently, few geometry and resampling-combining methods have been proposed, which aim to integrate sampling and geometry constraints to take both advantages, such as Zhu et al. [19] proposed improved RANSAC and Lan et al. proposed GMS-RANSAC [26]. However, such a simple combination of geometry constraints and sampling shows a limited improvement.

GC-RANSAC (Graph-Cut RANSAC), it induces the graph-cut algorithm to RANSAC, which proposed Graph Cutting algorithm based on spatial consistency to improve the accuracy of the RANSAC [38]. Our previous work [27] proposed an effective combination of geometry and resampling, which calculates the mismatch probability of each matches through triangulation constraints and calculates the transformation model of the image pairs through probability sampling. The method shows a good performance especially in a high mismatch rate compared to the existing methods. However, its calculation of network distortion is of large complexity.

3. Methodology

In practical, a compact and effective method to eliminate mismatches is in crucial need. However, most of existing techniques, either sampling-based or geometry-based, involve complex and thorough calculations or sampling conditions. Learning-based methods, on the other hand, typically require a significant amount of data training and its generalization is affected by the training data and process.

The combined methods are a better choice to achieve this, with process of sampling, a simple probability calculation approach can achieve comparable outcomes to existing state-of-art methods, with no need for accurate mismatch probability estimates. By combining geometry and sampling, a more promising approach can be achieved with a limited prior probability. Considering this, we propose a simple yet effective method that leverages prior probability to guide sampling and improve the overall results.

3.1. Framework

The proposed method can be visualized by flowchart shown in Fig. 1

In this paper, we first construct triangular topological network of matching points, then describe and quantify the local topology similarity of matching points for each matching points by comparing their local topological relations, and subsequently quantifies the correctly matched probability according to the topology similarity. The higher the local topology similarity of a pair of matching points, the more likely it is to be a correct match. Secondary, introduce the mismatch probability of each pair of matching points into the sampling progress, so that matching points pairs with a high mismatch probability are more likely to be selected by the sampling. After sampling, we obtain matrix with the best results and thus filter the correct matches.

In addition, we can also improve the sampling consistency to speed up the convergence, such as MLESAC, which optimizes the loss function.

3.2. Quantification of mismatching probability

Generally, the position and shape of an object in two images conform to a certain transformation relationship, so we can suppose one image as the initial state and another as the result of the transformation. After transformation, the position of the object in the image changes, but the topological relationships of its components are preserved. As a result, for the matches between two images, they can be filtered according to the topological relationship. The correctly matched points will have similar topological relationship in the two images, while the mismatched points not because they do not conform to the transformation.

Fig. 2. The topology stability of point cloud. (a) and (b) show the topology stability of 1D and 2D point cloud respectively.

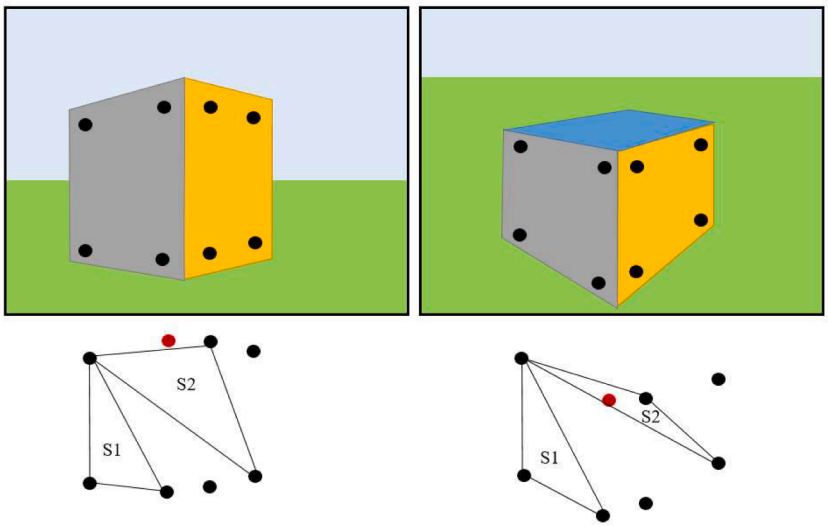

Fig. 3. Local topology stability, the vertex of S1 is on the local same plane with the red point, while S2 not. As a result, S1 remain the topology relation with red point, and S2 not.

3.2.1. Topology stability and local topology similarity

The topology relationships mentioned in this paper are related to location relationship and inclusion relationship between points and their surrounding points, ignoring the quantitative distance and azimuth. Topology stability means that these relationships preserved after transformation.

For example, after 1D (one-dimensional) scaling transformation, 1D point clouds still maintain the same topology, as shown in Fig. 2(a), the relative proximity and local relationship between the points have not changed. For 2D planar point cloud, after homographic transformation, the relative position of each point and the inclusion relationship between each point and the topological area formed by other points preserve. For example, point P_4 is always in the area formed by pointsP_2,P_3, P_5 andP_7, only the absolute distance and angle has changed, as shown in Fig. 2(b).

When a point cloud existed in a high-dimensional space originally, being transformed in high dimension and projected to a low-dimensional point cloud will cause loss of global topology stability, but retains local topology in the same low dimension space.

For instance, 3-D points are projected onto a 2-D image plane and switching camera perspective is equivalent to the 3-D transformation for point clouds. Although there may be some loss of topology due to 3-D transformations, local point clouds within the same plane still retain their original topology relationships, conforming to the homography transformation as shown in Fig. 3.

However, for some points with the same low dimension in a local area, such as the points on a plane in the 3-D space, the origin transform on this area is equivalent to a low-dimensional transformation. Consequently, the points in locally low-dimension still conform to topology stability after high-dimensional transformation.

Therefore, for each point in a point cloud, its topological relationship before and after transformation is not always preserved same, but has some topological similarity, which means, there will always be neighbor points in or near the same low-dimensional space in high-dimensional space with this point, such as the same plane in 3-D space. Generally, as a part of the neighbor points around a point retains the topological relationship, the point has a topological similarity through the transformation.

The more topological relationships preserved with neighbor points, the higher the local topology similarity, the more likely the point is correct matched, so we propose a method for calculating mismatching probability based on this. However, it is impossible to establish accurate relationships between correctness and local topology similarity, because it is impacted by various conditions such as number of neighborhoods, the position of the point in the network, etc.

3.2.2. Quantification of local topology similarity

In this paper, we construct triangular topological network of matching points, and quantify the local topology similarity through the connection between the points in the network then calculate the mismatching probability.

Here we define the first-order neighborhood, second-order

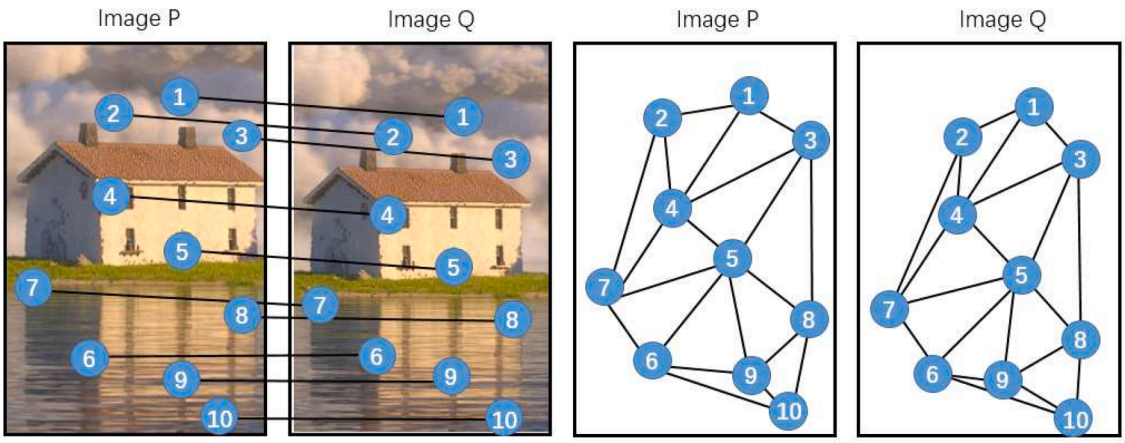

Fig. 4. The example of feature matching and topology relationship between matching points.

neighborhood …. *K*-order neighborhood of each matching point as the set of all the neighbor points which are connected through one topology edge, two topology edges …. *K* topology edges.

Suppose that two images P and Q are matched and *m* pair of matches is produced, and we record $P_1, P_2, P_3 \ldots P_m$ as the matching points in image P and $Q_1, Q_2, Q_3 \ldots Q_m$ in image Q. For $i \in \{1, 2, 3 \ldots m\}$, the matching point P_i and Q_i are matched, we record this match as $P_i - Q_i$. Then construct triangular topology network of the two images.

We firstly consider the topological relationship of first-order neighborhood, which is based on the direct connection of the topological edges. The local topological relationship between a point i and its neighborhood can be derived from the topological edges between them. And we use an m-D vector V_i to indicate the local topological relationship. The kth element of the vector indicates whether matching point k connect to this matching point (point *i*) based on a topological edge.

$$\mathbf{V}_i = [connect(i,1)\ connect(i,2) \ldots\ connect(i,m)]^T, \tag{1}$$

where

$$\text{connect(i, j)} = \begin{cases} 1\ connected\ by\ topo\ edge \\ 0\ not\ connected\ or\ i = j \end{cases} \tag{2}$$

For the matching point P5 in Fig. 4 above, the corresponding vector is

$$\mathbf{V}_5 = [0\ 0\ 1\ 1\ 0\ 1\ 1\ 1\ 1\ 0]^T$$

As the results of triangular topology, the calculation of each vector corresponding to each matching point can be accomplished. For the local topological relationship of all matching points in the image, we can use an m-D matrix **M** to represent it, the matrix is a symmetric matrix composed of vectors of each matching point.

$$\mathbf{M} = [\mathbf{V}_1\ \mathbf{V}_2\ \mathbf{V}_3 \ldots\ \mathbf{V}_m] \tag{3}$$

After obtaining the topological relationship matrices $\mathbf{M}_p$ and $\mathbf{M}_q$ of the two images, we can compare the two matrices and get the edges which preserve the topological relationship in the two images. These edges can reflect the local topology similarity of the matching points in the two images directly.

Then we multiply these two matrices by elements.

$$\mathbf{M_{pq}} = \mathbf{M_p}\ .*\ \mathbf{M_q} \tag{4}$$

For two pair of matching points $P_i - Q_i$ and $P_j - Q_j$, only if both connected in images P and Q, both $\mathbf{M_p}(i,j) = 1$ and $\mathbf{M_q}(i,j) = 1$ are valid with the result of 1 after operation, which indicates that the two pair of matching points have the same topology relationship in both images and they are neighbors of each other, so that it can be inferred that they are more likely to be the correct matching points.

The operation above gets topology edges with the same neighborhood topology in the picture P and Q. As for matching points, it considered that the more such edges around, the more likely the local area is to be correctly corresponded in the picture P and Q, the lower the mismatch probability of the match $P_i - Q_i$. Then we quantify this probability. Since the ith column in the matrix M_{pq} presents the first-order neighborhood topology similarity of the matching point P_i, we can sum over each column. The ith element in the vector SUM represents the number of local topological relationships of the pair of matching points $P_i - Q_i$. The larger the sum, the more likely a pair of matching points to be correct.

$$\mathbf{SUM} = [1 \quad 1 \ldots 1 \quad 1] \cdot \mathbf{M_{pq}} \tag{5}$$

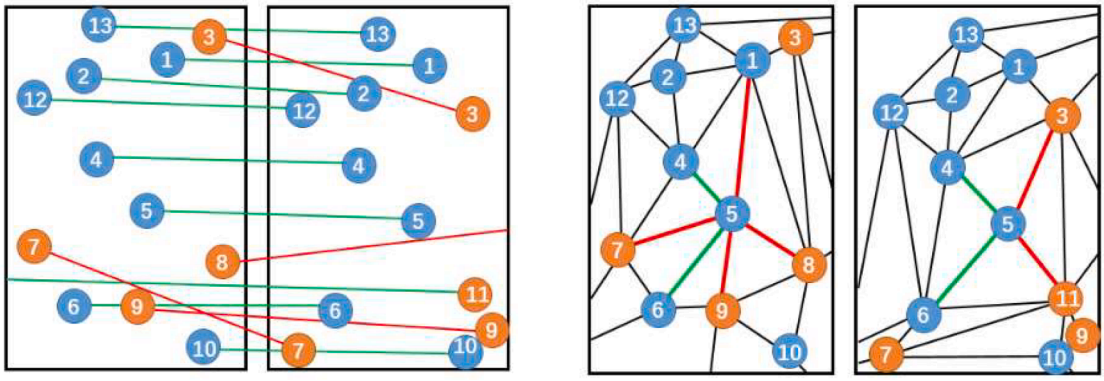

Fig. 5. Some cases that first neighborhood cannot deal with.

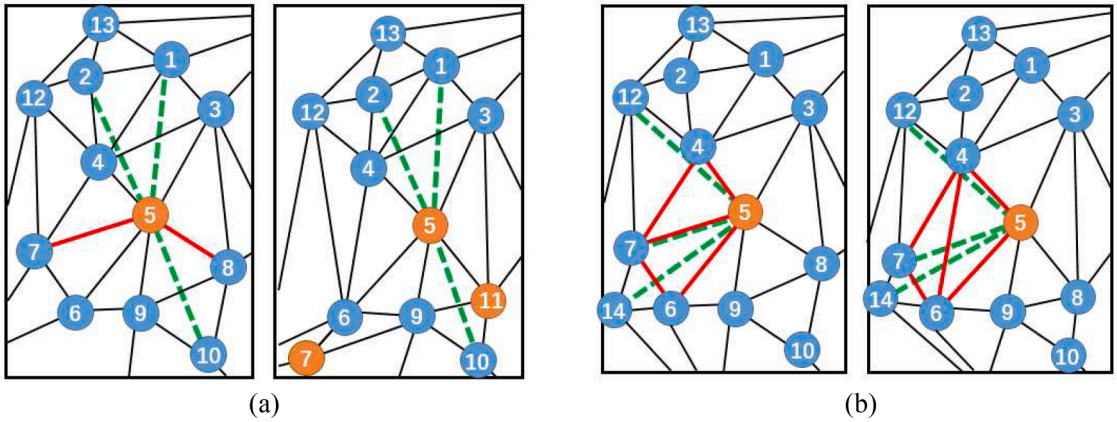

Fig. 6. Some cases based on first-order neighborhood will lead to inconsistent in two images (such as the point P_5) while second-order neighborhood not be affected, the red lines indicate the difference between two topology network, the green lines indicate the second-order neighborhood relationship will not change in these cases. (a) indicates the cases that the neighbor points are mismatched. (b) indicates the cases that the different ways of triangulation.

Generally, the first-order neighborhood can reflect the neighborhood topological relationship well, but it may preserve less topology relationship in some case. As shown in Fig. 5, when there are more mismatches, such as $P_3 - Q_3$, $P_8 - Q_8$,etc., the topology relationship around matching point P_5 is different in two images, only P_4 and P_6 preserve the topology relationship with P_5.

In these cases, we consider the second-order neighbourhood topology, that is, the topological relationship between the matching point and its neighbour points' neighbour points. In Fig. 6, match point P_1, P_2, P_3, P_4, P_6, P_7, P_8, P_9, P_{10}, P_{12}, P_{12} are second-order neighbourhood of P_5. As shown in Fig. 6(a), it can be seen that even if mismatches occur around matching point P_5, the topological relationship between matching point P_5 and P_{10} will not be affected, which remains a second-order neighbourhood topological relationship. For matching point P_7 which is mismatched, the second-order neighbourhoods around it have changed dramatically. On the other hand, when under a huge perspective transformation between two images, there is another way of triangulation in a topological area, which also results in inconsistent topological relations between some correct matching points and their neighbourhood. As shown in Fig. 6(b), the first-order neighbourhood topology between matching points P_5 and P_7 is not consistent, while both matching points are correct matches. As for second-order neighbourhood, the topological relationship between matching points P_5 and P_7 is consist in two images.

Assume that the first-order neighborhood topological relationship matrix M of the matching points in the images has been obtained above, the topological vector for a match $P_i - Q_i$ is $\mathbf{V}_i = [v1\ v2\ v3 \ldots vm]^T$.As for second-order topology, we can also use a topological vector $\mathbf{Vp}_i$ denotes, the kth element of the vector represents whether a pair of matching point $P_k - Q_k$ has a second-order neighborhood relationship to this pair of matching points $P_i - Q_i$. Here, instead of limiting the value to 1 or 0, we give the second-order neighborhood relationship a weight of k, which meets $0<k<1$. Taking the first-order neighborhood topology relationship into account, topological vector Vp_ican be calculated by the formula 2.

$$\mathbf{Vp}_i = \mathbf{V}_i + k(v1 \times \mathbf{V}_1 + v2 \times \mathbf{V}_2 + \ldots + vm \times \mathbf{V}_m) \quad ((6))$$

The vector can also be derived from the fist-order topological relationship matrix **M** above:

$$\mathbf{V}'_i = \mathbf{V}_i + k\cdot\boldsymbol{M}\mathbf{V}_i = (\boldsymbol{E} + k\boldsymbol{M})\boldsymbol{V}_i \quad ((7))$$

Similarly, we define a second-order topological relationship matrix **Mp**.

$$\mathbf{Mp} = [\mathbf{Vp}_1\ \mathbf{Vp}_2\ \mathbf{Vp}_3 \ldots\ \mathbf{Vp}_m] \quad ((8))$$

After obtaining the topological relationship matrices of the two images, we also multiply these two matrices by elements, and then calculate the sum of each column. At the same time, as the second-order topology is given the weight of k, it is indicated that the second-order topological relationship is weaker than the first-order topological relationship as the basis for judging the matching points.

For the third-order neighborhood or more, the principle is similar, while the effect cannot be improved more, so it will not be described here.

3.2.3. Calculation of mismatching probability

After obtaining the topological relations for each matching point, we assign the mismatch probability p_i to a matching $P_i - Q_i$ which is corresponding to the matching point i. And it obeys the following distribution with the number of topological relations SUM(i). As shown in Eqs. (3)–(6), the larger the number of topological relations, the lower its probability of being a mismatch, and it will be exactly 0.5 when number achieve the average value μ. σ represents the second moment, which is used to normalize the input parameters. The value of σ becomes larger when the topology network is sparse, the difference of topological relations between matching points is larger.

$$p_i = 1 - \frac{1}{1 + e^{-\frac{SUM(i)-\mu}{\sigma}}}, where\ \mu = \frac{1}{n}\sum_{1}^{n} SUM(i),\ \sigma = \sqrt[2]{\frac{1}{n}\sum_{1}^{n}(\mathrm{SUM(i)} - \mu)^2} \quad (9)$$

However, we cannot directly use this probability to judge whether a match is correct or not, this probability is a reflect of its local topology similarity, which is not so accurate but enough for guiding sampling to remove mismatches.

3.3. Probability sampling consensus

The RANSAC calculates a model by sampling a small amount of data, and then substitutes transform matrix between two graphs. The transform matrix refers to fundamental matrix F for all of rigid transformations in 3D space, and it can also simplify to homography matrix H when the main content of images is on a plane.

Specifically, the algorithm randomly samples n pairs of matching points and calculates the transform matrix, the matching points on the image should meet to the following equation,

$$\boldsymbol{x}'\boldsymbol{F}\boldsymbol{x}^{\boldsymbol{T}} = 0\ or\ \boldsymbol{x}' = \boldsymbol{H}\boldsymbol{x} \quad (10)$$

Exactly, and the error should be in a certain threshold.

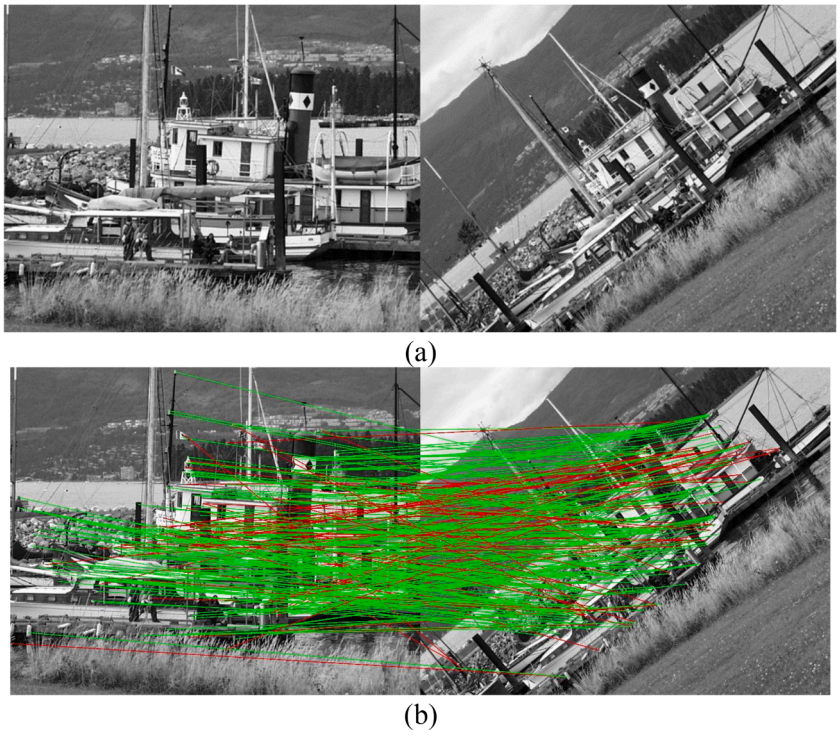

Fig. 7. The results of image matching. (a) used for image matching, and (b) shows the result of image matching, the green matches indicate correct matches while the red matches indicate incorrect.

$$\boldsymbol{x}^{'}\boldsymbol{F}\boldsymbol{x}^{T} < threshold \ or \ |\boldsymbol{x}^{'} - \boldsymbol{H}\boldsymbol{x}| < threshold \tag{11}$$

After calculating the transform matrix, if a pair of matching points meet the condition above, it is considered as correctly matched, and if not, it will be removed.

In this paper, according to the mismatch probability p of each pair of matching points, we assign the probability ps for each pair to be sampled. The relationship can be expressed by Formula 8. That is, the lower the mismatch probability p, the higher the probability of to be sampled.

$$ps_i = (1 - p_i) \Big/ \sum_{1}^{n} (1 - p_j) \tag{12}$$

After obtaining the sampling probability, we sample the matching point pairs according to the probability ps. We sample 4 pairs of matching points each time for calculating the homography matrix H between image P and image Q, and sample 7 pairs of matching points to calculate the fundamental matrix F for image pairs with greater perspective transformation.While sampling, we calculate the error of each point under transformation matrix. The point with a low error is considered to be inlier, so that we find the homography matrix with the largest number of inliers.

Finally, combining the transformation matrix obtained above, we can filter the inlier of the model (the matches that have a low reprojection error under this transformation matrix), those we consider to be the correct matches.

The time complexity of the algorithm is determined by two aspects, mismatching probability calculation and sampling.

As for the probability calculation, it consists of the construction of network and the quantification of topological similarity. The former varies between $O(n)$and $O(nlgn)$ in different approaches, where n is number of matches. The latter is implemented by traversing all the feature points, resulting in a linear time complexity (n) . The overall time complexity of this aspect can be considered between $O(n)$and $O(nlgn)$ depending on approaches to construct networks.

As for sampling, in each sampling process, it contains probability sampling with complexity of $O(n)$ and the transform model calculation with complexity of $O(1)$. So the time complexity of sampling exhibits linear correlation to maximum iterations and number of matches $O(kn)$, (k is maximum iterations), which is the primary factor affecting algorithmic efficiency.

Therefore, the complexity of proposed method is between $O(kn)$and $O(nlgn + kn)$. With topological similarity facilitating convergence, fewer iterations can be utilized to attain more precise results, leading to the promotion of the efficiency. As for second or higher order of neighborhood, it will be slightly slower for extra matrix operation.

4. Experiment

The proposed method is compared with the state-of-art mismatching removal method.

The experiments are performed on Windows 10 operating system of a Macbook Air (13-inch, 2017) computer with an Intel Core i5-5350K processor and 8-GB RAM. The algorithms in this paper are written in Python. And in all experiments, the correspondences are computed from the ORB and SIFT keypoints.

4.1. Examples of mismatching removal

Firstly, we extract the feature points and match them in a pair of images in Fig. 7. As shown in the figure, the green matches indicate correct matches which are judged based on the ground truth provided by the dataset, while the red ones are the incorrect. In this figure, there are 320 matches, among them 233 are correct.

We construct topological networks for both two images, and here we choose Delaunay triangular, as shown in Fig. 8(a). The topological network in Fig. 8(b) represents the first-order neighbourhood topology, where the red topology edge represents that the end points of them preserving the first-order neighbourhood topology in both graphs, and it shows that the matching points retaining the first-order neighbourhood topology are mostly the correct matching points;

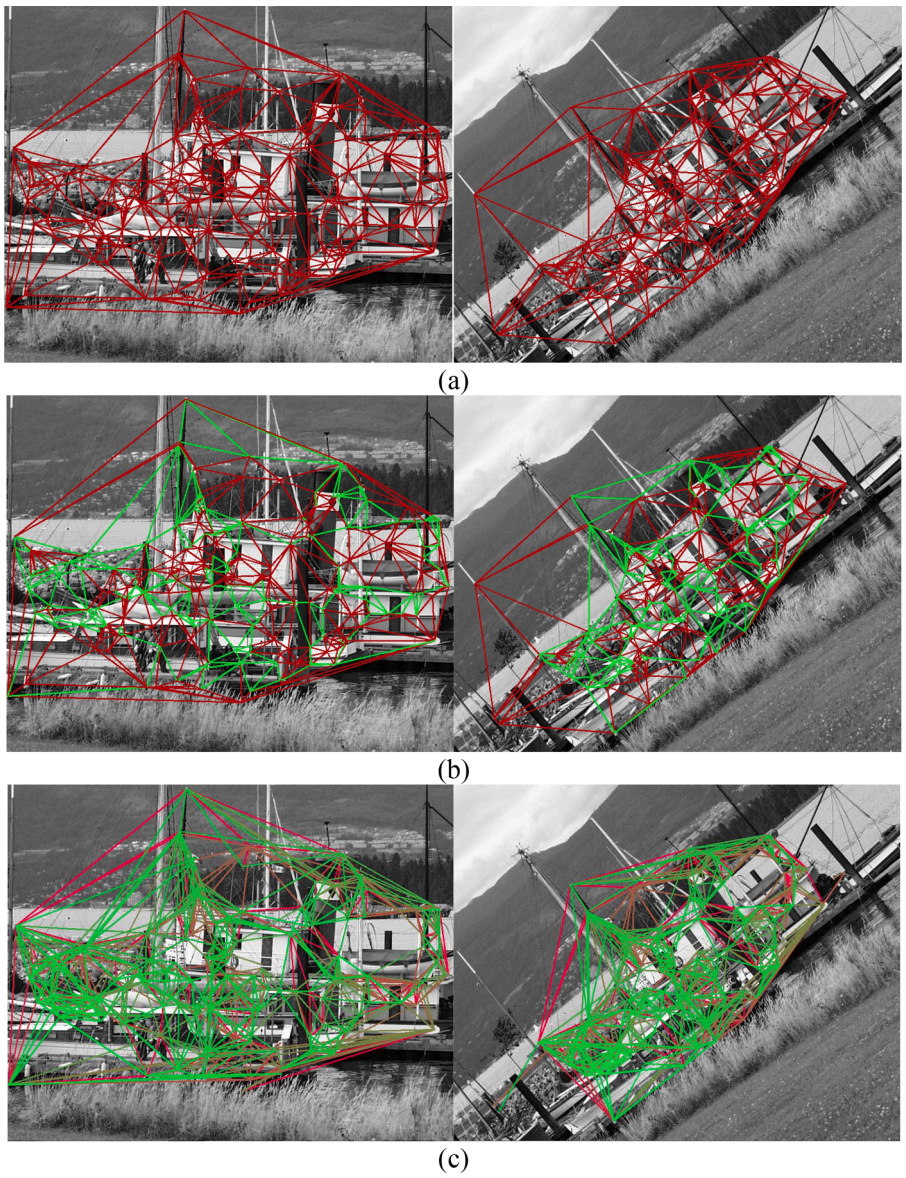

Fig. 8. Topology network and neighborhood of points in two images. (a) shows topology network of two images, (b) the green edges represent the first-order neighborhood which preserved, (c) the green edges represent the second-order neighborhood which preserved.

In Fig. 8(c) the topological network represents a second-order neighborhood topology, the topology edges in the figure represent the matching points at the two ends of these edge preserving the second-order neighborhood topology in two figures, which is much more than edges in first-order neighborhood, and reduces the cases where some of the topological relationships are preserved but not identified. Since there will be more matching points preserved local topology relationship as the order of neighborhood more than 3, the edges in image will be denser, we do not present here.

Subsequently, calculate the local topology similarity of each matching point in the two images, and calculate the mismatch probability of matching point based on Eq. (9).For the first-order neighbourhood topology, the probability of correct matching points is relatively discrete, while second-order neighbourhood topology is more detailed and differentiated compared to the first-order since the neighbourhood are considered more thoroughly.

Finally, we import the probability of each match into the sampling process of probability sampling consensus and calculate the transformation matrix of the two images. Since the two images are approximate to the affine transformation, the model of the sampling calculation can be set as the homography matrix between the two images. After calculating, the homography matrixes based on first-order and second-order neighborhood are obtained, which were:

$$\mathbf{H} = \begin{bmatrix} 0.572 & 0.472 & 25.25 \\ -0.469 & 0.569 & 349.2 \\ \sim 0 & \sim 0 & 1 \end{bmatrix} \quad \mathbf{H}' = \begin{bmatrix} 0.568 & 0.471 & 25.08 \\ -0.468 & 0.566 & 347.7 \\ \sim 0 & \sim 0 & 1 \end{bmatrix}$$

For the ground truth given in the database, the homography matrix between two images is:

$$\mathbf{Htruth} = \begin{bmatrix} 0.569 & 0.471 & 25.51 \\ -0.468 & 0.565 & 348.2 \\ \sim 0 & \sim 0 & 1 \end{bmatrix}$$

The error of each parameter in the matrix is within 1 %, which indicates that we have achieved quite good results both in the first-order and the second-order neighborhood. Therefore, the LTS in following experiments is refer to first-order neighbourhood. After filtering the matches according to the homography matrix calculated above, we obtain the following results. For both the methods based on the first-order or the second-order neighborhood topology result in a good

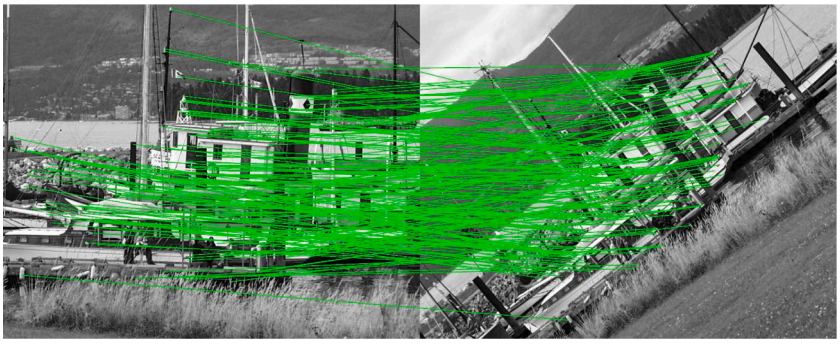

Fig. 9. Result of mismatching removal.

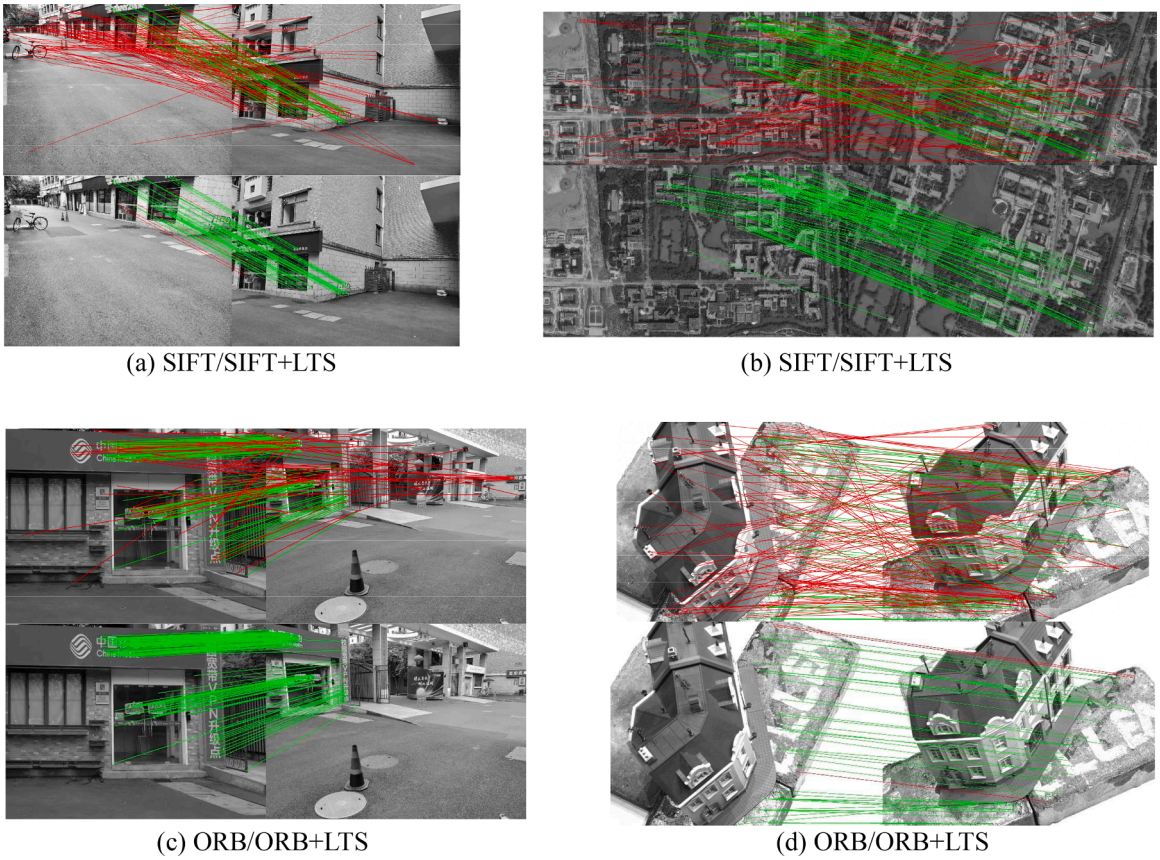

Fig. 10. Some examples for mismatching removal. (a) and (b) shows the result of mismatching removal for SIFT feature matches. (c) and (d) shows the result of mismatching removal for ORB feature matches.

effect. The algorithm obtained 232 matches after mismatching removal, and all of them are correct, which represents that the proposed method has very high precision and recall, as shown in Fig. 9.

Here we take more examples on the images. As shown in Fig. 10, the matches are produced by SIFT and ORB, each Figure consists of 2 pairs of images, containing the matches before and after mismatching removal.

4.2. Comparative experiment

4.2.1. Experimental setting

In this Section, we conduct comparative experiments to compare the proposed method (with first-order neighborhood) with the all kinds of methods, RANSAC (abbreviated as RAN), LPM [25], GTM [16], GMS [17], TSAC [27], GC-RANSAC (abbreviated as GCR) [38], OA-Net [34]. And for the methods with process of sampling, we set the max iterations for sampling to be 100.

To test the proposed method, we used the dataset of Mikolajczyk [22], HPatches [23], and hannover [24], which contain several scenes of images and their groundtruth transform matrix.

The main indicators of the experiments are precision, recall, F-score, and runtime. We define the precision as the proportion of correct matches in the matches extracted by mismatching removal method; And Recall is defined as the proportion of correct matches after extraction in whole correct matches. Generally, precision and recall are always negatively correlated. Therefore, F-score is used to measure the whole performance of mismatching removal, which is define as Eq. (13). And runtime is tested with methods writing and running under Python.

$$F-\text{score} = \frac{2 * precision * recall}{precision + recall} \tag{13}$$

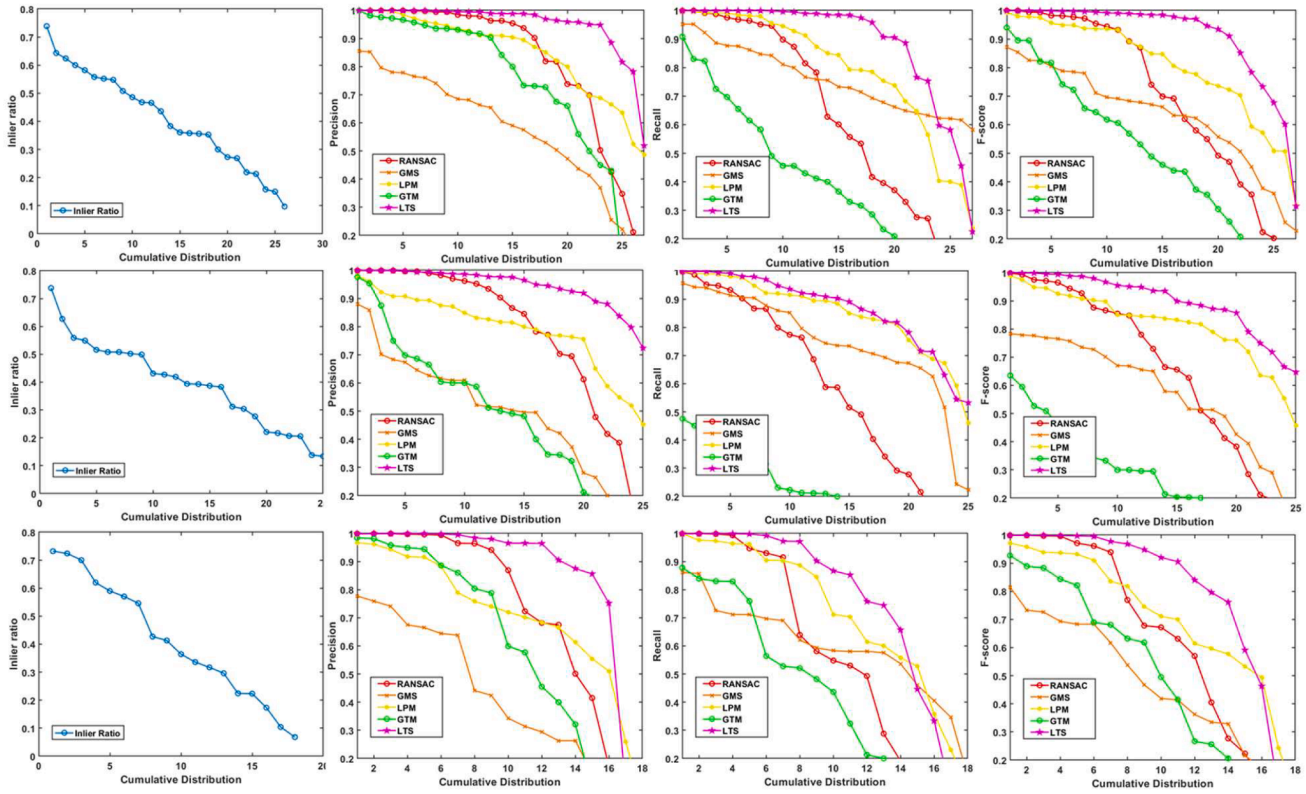

Fig. 11. The experimental results on the datasets. Each row from up to below are the results of the dataset Mikolajczyk [22], Hpatches [23], and hannover [24], each consists of the cumulative distribution graph of the inlier rate, precision, recall, and F-score in the dataset.

Table 1
The average results on three datasets.

outlier	Indicator	RANSAC	GMS	LPM	GTM	**LTS**
0.594	precision	0.774	0.498	0.784	0.607	**0.916**
	recall	0.599	0.699	0.789	0.353	**0.831**
	F-score	0.643	0.559	0.785	0.427	**0.865**
	Runtime/ms	68	10(C++)	**40**	190	51

We use cumulative distribution curves to reflect the characteristics of each method in the dataset, the slower the corresponding curve declines means that the method has better performance in the dataset. In the graph, for a method, the larger the horizontal ordinate (cumulative proportion) under a certain vertical ordinate (value of the indicator), the larger proportion of images have reached the value of the indicator.

4.2.2. *Experiments I*

We first carried out the experiments compared with some typical sampling-based and geometry-based methods. In order to reduce the contingency of random processes, we carried out experiments for ten times on each pair of images. The results are shown in Fig. 11 and Table 1.

From the results, we can observe that GMS performs not so accurate, while RANSAC have a better performance on precision but tends to have a low recall, GTM performs not so well and also have a low speed, LPM did better on recall especially in change of viewpoints, and executes fast, but the precision is lower than the resampling-based method. Compared with these methods, the proposed method LTS obtains the superior performance, and behaves more accurate. As for execution time, the proposed method faster than RANSAC even it contains the process of construction of topology network, which indicates that the proposed method accelerates the process of sampling, and compared with most other methods, it also shows a higher speed.

4.2.3. *Experiments II*

To better assess our approach, we compared it with our peer methods, geometry and sampling combine methods, and typical deep learning method. Additionally, we combined the typical geometry-based methods, LPM and GTM with sampling for comparison. The results are shown in Fig. 12 and Table 2.

From the results, we can observe that the performances of GMS+SAC and LPM+SAC are better than GMS and LPM, which indicates that the sampling step indeed improved the results to some extent, especially in precision. However, though these combinations achieve better results, their first stage to make judgements spend more time, and they are more inefficient combining with sampling. As for GC-RANSAC, it performs well with limited mismatches, and it becomes worse rapidly in high mismatching ratio. For our previous research TSAC, it performs better than GC-RANSAC. Learning-based method OA-Net achieves the best performance especially in recall. Compared with OA-Net, the proposed method LTS obtains the comparable performance, and behaves more accurate. As for execution time, the proposed method ran faster than any other method in this experiment, and almost 2 times faster than simple combination of sampling and geometry methods (e.g. LPM+SAC).

4.2.4. *Experiment under increasing mismatching ratio*

Then we compared the robustness of different methods under different mismatching ratios, which means whether the methods can adapt to worse mismatching situations. In these experiments, we produced a certain proportion of mismatches by shuffling the certain proportion of correct matches, the results of the experiments are presented in the graphs in Fig. 13.

Generally, as mismatches increases, the performance of all kinds of methods tend to be worse. Specifically, the recall of GMS remains high but its precision decreases rapidly, which shows that GMS aims to extract more but has a low accuracy. RANSAC performs well in precision in low mismatching ratio, as the ratio increases, it performs worse in

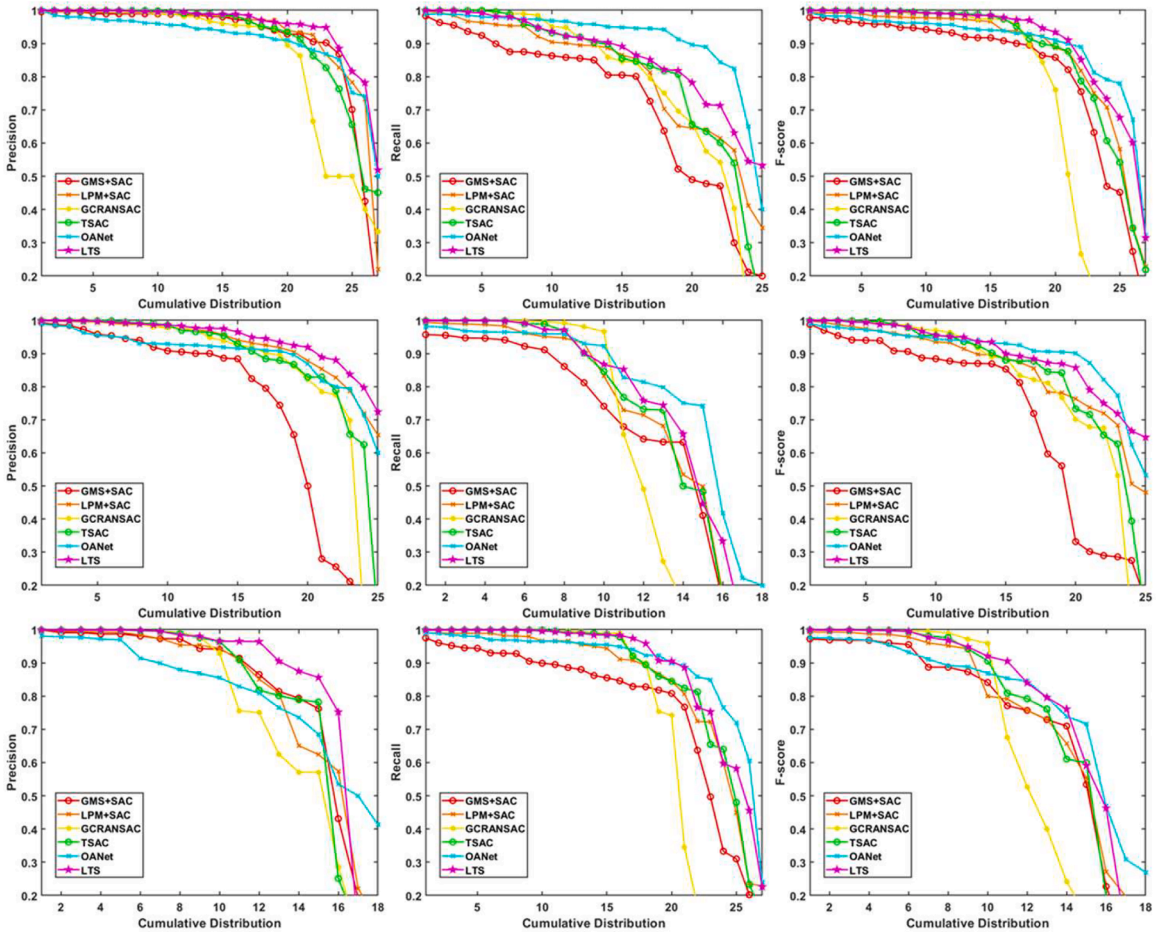

Fig. 12. The experimental results on three indicators with ome combination methods and deep learning method. The inlier ratio of dataset is similar to Fig.11 (a).

Table 2
The average results on three datasets.

outlier	Indicator	GMS+SAC	LPM+SAC	GCRANSAC	TSAC	OANet	**LTS**
0.594	precision	0.821	0.885	0.827	0.866	0.872	**0.916**
	recall	0.724	0.795	0.731	0.803	**0.872**	0.831
	F-score	0.756	0.829	0.792	0.827	**0.866**	**0.865**
	Runtime/ms	>100	109	15(C++)	76	56	**51**

precision, and the recall of RANSAC is also low, which shows the lower robustness of RANSAC in high proportion of mismatching. OA-Net shows highest performance, which remains high precision, recall and F-score with increasing proportion of mismatches. GC-RANSAC perform well when mismatching rate is not so high, and it becomes invalid quickly when mismatching ratio increase, which shows a weak robustness. And the method LTS we proposed perform better than other method mentioned expect OA-Net, regardless of precision, recall or F-score, LTS declines more slowly. In summary, our method LTS is significantly excellent in accuracy and robustness. Although our LTS performs slightly worse than OA-Net, a representative deep learning-based method, it is still a good choice for mismatching removal with no training process and a stronger generalization ability.

5. Conclusion

In this paper, a method LTS for mismatch removal is proposed. It first simplifies the mismatching probability calculation based on local topology similarity in an extremely simplest way to extract the more essential features, and imports the probability into process of the sampling to fully compensate the advantage of sampling. It is proven by the experiment that it has high accuracy and robustness, with a good adaptability to high mismatching ratio conditions.

However, the proposed approach relies on the establishment of a local topology network and becomes ineffective as the number of feature points in the image are limited, which prevents the construction of a stable topology. Furthermore, sampling requires the prior determination of a transformation model, rendering it unsuitable for non-rigid matching applications.

Generally, the proposed method LTS has achieved good results as a geometry and resampling combining method, which shows the combining method have a huge potential in the fields of mismatch removal.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

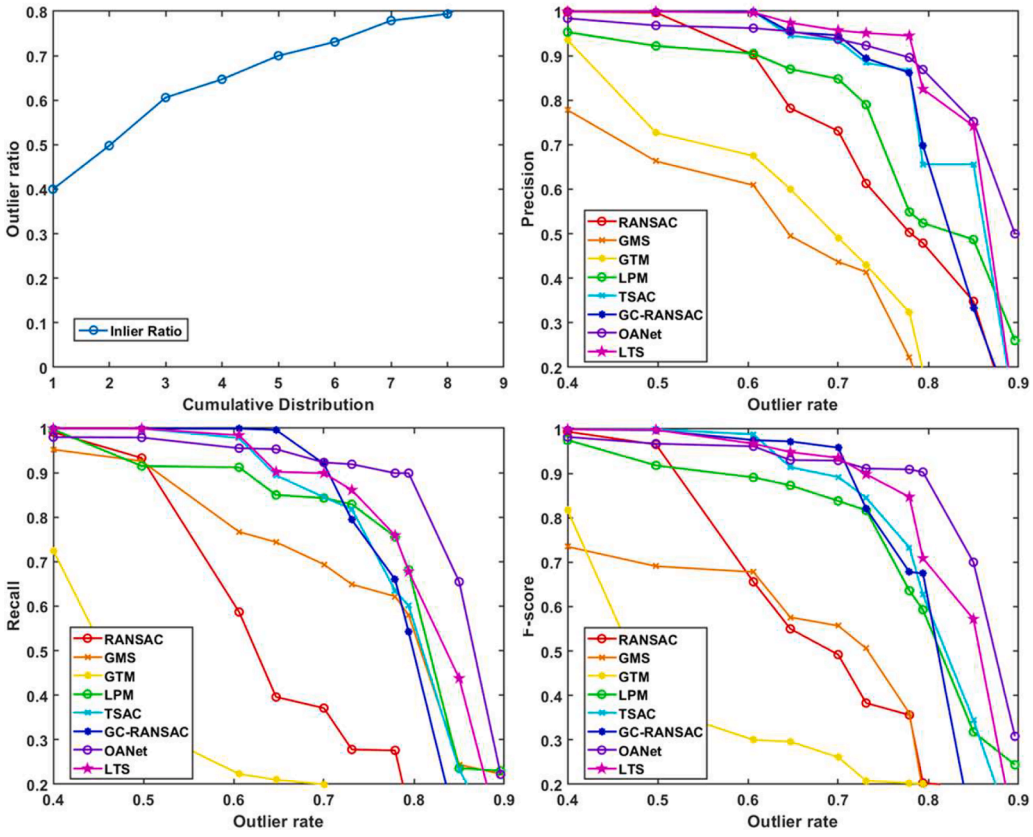

Fig. 13. The performance on different proportion of mismatches. The Figures respectively present the mismatch proportion of each image pair, and precision, recall, F-score of each method.

Data availability

Data will be made available on request.

Acknowledgments

This work was supported by the National Natural Science Foundation of China [Grant No's 52275514 and 52275547], and the Zhejiang Provincial Natural Science Foundation of China [Grant No. LY21E050021].

References

[1] D.G. Lowe, Distinctive image features from scale-invariant keypoints, Int. J. Comput. Vis. 60 (2) (2004) 91–110.
[2] H. Bay, T. Tuytelaars, L.V. Gool, SURF: Speeded up Robust Features (2006).
[3] C. Minchael, L. Vincent, S. Christoph, F. Pascal, BRIEF: binary robust independent elementary features, in: Proceedings of the 11th European conference on Computer vision: Part IV, 2010, pp. 778–792. Pages.
[4] K.M. Yi, E. Trulls, V. Lepetit, et al., LIFT: learned invariant feature transform, in: Proceedings of the European Conference on Computer Vision, 2016.
[5] D. Detone, T. Malisiewicz, A. Rabinovich, SuperPoint: self-supervised interest point detection and description, in: Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), 2018. IEEE.
[6] H. Richard, Z. Andrew, Multiple View Geometry in Computer Vision, Cambridge University Press, Cambridge, 2000, pp. 123–125.
[7] M. Fischler, R. Belles, Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography, CACM 24 (6) (1981) 381–395.
[8] P.H.S. Torr, A. Zisserman, MLESAC: a new robust estimator with application to estimating image geometry, Comput. Vis. Image Underst. 78 (1) (2000) 138–156.
[9] O. Chum, J. Matas, Matching with PROSAC-progressive sample consensus, in: Proceedings of the IEEE Computer Society Conference on Computer Society, 2005, pp. 220–226.
[10] N. Kai, J. Hailing, F. Dellaert, GroupSAC: efficient consensus in the presence of groupings, in: Proceedings of the 12th IEEE International Conference on Computer Vision, 2009, pp. 2193–2200.
[11] O. Chun, J. Matas, Randomized RANSAC with Td,d test, in: Proceedings of the 13th British Machine Vision Conference, Springer, Berlin, 2002, pp. 448–457.
[12] J. Matas, O. Chun, Randomized RANSAC with sequential probability ratio test, in: Proc of the 10th IEEE International Conference on Computer Vision, IEEE Press, 2005, pp. 1727–1732.
[13] M. Rahman, X. Li, X. Yin, DL-RANSAC: an improved RANSAC with modified sampling strategy based on the likelihood, in: Proceedings of the 2019 IEEE 4th International Conference on Image, Vision and Computing (ICIVC), 2019, pp. 463–468.
[14] C. Shi, Y. Wang, H. Li, Feature point matching using sequential evaluation on sample consensus, in: Proceedings of the International Conference on Security, IEEE, 2017, pp. 302–306.
[15] B. Gao, S. Liu, J. Zhang, B. Zhang, Pose estimation algorithm based on improved RANSAC with an RGB-D camera, in: Proceedings of the 2018 Chinese Control And Decision Conference (CCDC), 2018, pp. 5024–5029.
[16] W. Aguilar, Y. Fraud, F. Escolano, et al., A robust graph transformation matching for non-rigid registration, Image Vis. Comput. 27 (7) (2009) 897–910.
[17] J. Bian, W. Lin, Y. Matsushita, S. Yeung, T. Nguyen, M. Cheng, GMS: grid-based motion statistics for fast, ultra-robust feature correspondence, in: Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR, 2017, pp. 2828–2837.
[18] X. Zhao, Z. He, et al., Improved keypoint descriptors based on Delaunay triangulation for image matching, Optik Int. J. Light Electronopt. 125 (13) (2014) 3121–3123.
[19] W. Zhu, W. Sun, Y. Wang, S. Liu, K. Xu, An improved RANSAC algorithm based on similar structure constraints, in: Proceedings of the 2016 International Conference on Robots & Intelligent System (ICRIS), 2016, pp. 94–98.
[20] Y. Luo, R. Li, J. Zhang, et al., Research on correction method of local feature descriptor mismatch, in: Proceedings of the 2019 IEEE 4th Advanced Information Technology, Electronic and Automation Control Conference (IAEAC), IEEE, 2019.
[21] M. Zhao, H. Chen, T. Song, S. Deng, Research on image matching based on improved RANSAC-SIFT algorithm, in: 2017 16th International Conference on Optical Communications and Networks (ICOCN), Wuzhen, China, 2017, pp. 1–3, https://doi.org/10.1109/ICOCN.2017.8121270.
[22] K. Mikolajczyk, T. Tuytelaars, C. Schmid, et al., A comparison of affine region detectors, Int. J. Comput. Vis. 65 (1) (2005) 43–72.
[23] V. Balntas, K. Lenc, A. Vedaldi, K. Mikolajczyk, HPatches: a benchmark and evaluation of handcrafted and learned local descriptors, in: Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017, pp. 3852–3861.
[24] K. Cordes, B. Rosenhahn, J. Ostermann, High-resolution feature evaluation benchmark, in: Proceedings of the CAIP, 2013, pp. 327–334, pages.

[25] J. Ma, J. Zhao, J. Jiang, et al., Locality Preserving Matching, Int. J. Comput. Vis. 127 (5) (2019) 512–531.
[26] X. Lan, B. Guo, Z. Huang, S. Zhang, An improved UAV aerial image mosaic algorithm based on GMS-RANSAC, in: Proceedings of the 2020 IEEE 5th International Conference on Signal and Image Processing (ICSIP), 2020, pp. 148–152.
[27] Z. He, C. Shen, Q. Wang, X. Zhao, H. Jiang, Mismatching removal for feature-point matching based on triangular topology probability sampling consensus, Remote Sens. 14 (2022) 706.
[28] X. Jiang, Y. Xia, X.P. Zhang, J. Ma, Robust image matching via local graph structure consensus, Pattern Recognit. 126 (2022) 108588.
[29] Y. Xia, J. Ma, Locality-guided global-preserving optimization for robust feature matching, IEEE Trans. Image Process. 31 (2022) 5093–5108.
[30] Y. Xia, J. Jiang, Y. Lu, W. Liu, J. Ma, Robust feature matching via progressive smoothness consensus, ISPRS J. Photogramm. Remote Sens. 196 (2023) 502–513.
[31] K.M. Yi, E. Trulls, Y. Ono, V. Lepetit, M. Salzmann, P. Fua, Learning to find good correspondences, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2018, pp. 2666–2674, https://doi.org/10.1109/CVPR.2018.00282.
[32] P.E. Sarlin, D. DeTone, T. Malisiewicz, A. Rabinovich, SuperGlue: learning feature matching with graph neural networks, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2020, pp. 4938–4947.
[33] X. Liu, G. Xiao, R. Chen, J. Ma, PGFNet: preference-guided filtering network for two-view correspondence learning, IEEE Trans. Image Process. 32 (2023) 1367–1378, https://doi.org/10.1109/TIP.2023.3242598.
[34] J. Zhang, et al., OANet: learning two-view correspondences and geometry using order-aware network, IEEE Trans. Pattern Anal. Mach. Intell. 44 (6) (2022) 3110–3122, https://doi.org/10.1109/TPAMI.2020.3048013.
[35] J. Ma, X. Jiang, J. Jiang, J. Zhao, X. Guo, LMR: learning a two-class classifier for mismatch removal, IEEE Trans. Image Process. 28 (8) (2019) 4045–4059.
[36] D. Barath, J. Matas, J. Noskova, MAGSAC: marginalizing sample consensus, in: Proceedings of the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Long Beach, CA, USA, 2019, pp. 10189–10197, https://doi.org/10.1109/CVPR.2019.01044.
[37] R. Raguram, O. Chum, M. Pollefeys, J. Matas, J.-M. Frahm, USAC: a universal framework for random sample consensus, IEEE Trans. Pattern Anal. Mach. Intell. 35 (8) (2013) 2022–2038.
[38] D. Barath, J. Matas, Graph-cut RANSAC, in: Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 2018, pp. 6733–6741.
[39] J. Ma, J. Jiang, H. Zhou, J. Zhao, X. Guo, Guided locality preserving feature matching for remote sensing image registration, in: Proceedings of the IEEE Transactions on Geoscience and Remote Sensing 56, 2018, pp. 4435–4447, https://doi.org/10.1109/TGRS.2018.2820040.
[40] J. Ma, J. Zhao, J. Tian, A.L. Yuille, Z. Tu, Robust point matching via vector field consensus, IEEE Trans. Image Process. 23 (4) (2014) 1706–1721.
[41] J. Ma, J. Wu, J. Zhao, J. Jiang, H. Zhou, Q.Z. Sheng, Nonrigid point set registration with robust transformation learning under manifold regularization, IEEE Trans. Neural Netw. Learn. Syst. 30 (12) (2019) 3584–3597.
[42] Y. Wang, D. Zhang, J. Tian, Topological clustering and its application for discarding wide-baseline mismatches, Opt. Eng. 47 (5) (2024) 057202.
[43] L. Dai, et al., MS2DG-Net: progressive correspondence learning via multiple sparse semantics dynamic graph, in: Proceedings of the 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), New Orleans, LA, USA, 2022, pp. 8963–8972.

Zaixing He received his B.Sc. and M.Sc. degrees in Mechanical Engineering from Zhejiang University, China in 2006 and 2008, respectively. He received his Ph. D. degree in 2012 from the Graduate School of Information Science and Technology, Hokkaido University, Japan.

He is currently an associate professor in the School of Mechanical Engineering, Zhejiang University. His-research interests include robotic vision, visual intelligence of manufacturing equipment, and optical-based measurement. He has published over 40 peer reviewed papers in prestigious journals such as Pattern Recognition, IEEE TIE, TII, TIM, IEEE/ASME TMech, Neurocumputing. He served as Lead Guest Editor or Guest Editor of several journals including IEEE TCE and Mathematics, Program Chair or TPC of more than 10 international conferences. He is a senior member of IEEE.

Chentao Shen received the B.Sc. Degree in process Mechanical Engineering from Zhejiang University, China, in 2022.

He is currently a M.D. candidate in the Department of Mechanical Engineering, Zhejiang University, China. His-research interests include machine vision and image process.

Xinyue Zhao received her M.S. degree in Mechanical Engineering from Zhejiang University, China in 2008, and her Ph.D degree in Graduate School of Information Science and Technology from Hokkaido University, Japan in 2012.

She is currently an associate professor in the Department of Mechanical Engineering, Zhejiang University, China. Her research interests include machine learning and machine vision. She has published nearly 50 peer reviewed journal papers.