Homography Estimation Based on Order-Preserving Constraint and Similarity Measurement

HAIJIANG ZHU[1], XIN WEN[1], FAN ZHANG[1], (Senior Member, IEEE),
XUEJING WANG[2], AND GUANGHUI WANG[3], (Senior Member, IEEE)

[1]College of Information Science and Technology, Beijing University of Chemical Technology, Beijing 100029, China
[2]Center for Information Technology, Beijing University of Chemical Technology, Beijing 100029, China
[3]Department of Electrical Engineering and Computer Science, University of Kansas, Lawrence, KS 66045, USA

Corresponding authors: Haijiang Zhu (zhuhj@mail.buct.edu.cn) and Guanghui Wang (ghwang@ku.edu)

ABSTRACT Homography is an important concept that has been extensively applied in many computer vision applications. However, accurate estimation of the homography is still a challenging problem. The classical approaches for robust estimation of the homography are all based on the iterative RANSAC framework. In this paper, we explore the problem from a new perspective by finding four point correspondences between two images given a set of point correspondences. The approach is achieved by means of an order-preserving constraint and a similarity measurement of the quadrilateral formed by the four points. The proposed method is computationally efficient as it requires much less iterations than the RANSAC algorithm. But this method is designed for small camera motions between consecutive frames in video sequences. Extensive evaluations on both synthetic data and real images have been performed to validate the effectiveness and accuracy of the proposed approach. In the synthetic experiments, we investigated and compared the accuracy of three types of methods and the influence of the proportion of outliers and the level of noise for homography estimation. We also analyzed the computational cost of the proposed method and compared our method with the state-of-the-art approaches in real image experiments. The experimental results show that the proposed method is more robust than the RANSAC algorithm.

INDEX TERMS Homography estimation, order-preserving constraint, similarity measurement.

I. INTRODUCTION

Homography estimation has wide applications in processing multiple images and video sequences, such as robot navigation [2], tracking and mapping [3], augmented reality [4], [27], image mosaics [5], [6], [23]–[25], finding point correspondences [7]–[9] and image registration [28]. In general, a homography model can be computed from a set of corresponding feature points extracted from the image pairs or two frames in a video sequence [1]. However, accurate estimation of the homography is still a challenging problem in computer vision.

The most common homography estimation technique is the RANSAC method [10], which is adopted in most of the reported works [4]–[9], [23]–[25], [27], [28]. On the android platform Cheng et al. [4] developed mobile

augmented reality (MAR) application based on planar natural features that are registered using the 2D homography matrix Xiong and Pulli [5] implemented fast image stitching on a mobile phone through homography estimating between two images. Sun et al. [8] proposed an efficient line matching algorithm based on planar homography from a pair of calibrated aerial photogrammetric images. Zhang et al. [9] investigated a homography estimation algorithm by employing a contour model to track and locate the texture-less object in highclutter environments. Du and Padir [24] studied the infinite homography method for an intelligent portable aerial surveillance system in order to develop a low-cost, light-weight unmanned aerial vehicle. Lim and Park [27] presented an efficient method from singleview to multiple-view conversion using the estimated infinite homography.

Lou and Gevers [28] explored an image alignment method by incorporating piecewise local geometric models in which the homography is employed for matching the planar regions.

Although homography estimation based on the RANSAC method [10] is effectively utilized in many reported studies, the RANSAC method is very computationally expensive due to its iterative nature. To reduce its complexity numerous extensions have been proposed to improve the RANSAC method [11]–[17]. Chum and Matas [13] first presented a randomized RANSAC algorithm to increase the speed of model parameter estimation under a broad range of conditions Bhattacharya and Gavrilova [16] utilized the topological information to improve RANSAC feature matching. Márquez-Neila *et al.* [17] introduced a procedure to reduce the number of samples required for fitting a homography to a set of noisy correspondences using the RANSAC method. This was achieved by means of a geometric constraint that detects invalid minimal sets. However, all these improvements are still based on the RANSAC framework.

Generally speaking, finding the correct correspondence point between two images is a key problem for homography estimation. To reduce the iteration times in RANSAC, Bhattacharya and Gavrilova [16] and Márquez-Neila *et al.* [17] investigated an order-preserving constraint on correspondence points. The image point order-preserving constraint, which was previously proposed in Ferrari *et al.* [18], was applied to remove mismatches for triples of regions. However, Márquez-Neila's method [17] presented three-point order-preserving in selecting random four-point correspondences.

Inspired by Márquez-Neila *et al.* [17] and Ferrari *et al.* [18], we explored the four-point order restriction to estimate the homography between two images. Supposing that the initial matches between two views are given, we determine the relative order of the four correspondence points using the normalized data and the similarity distance. Then, the homography is estimated by minimizing the sum of the squared Euclidean distance between the detected points and the reprojected points [1]. In the synthetic experiments, we investigated the accuracy of the three types of methods and the influence of the proportion of outliers and the level of noise for homography estimation. In real image experiments, we compared our method with the RANSAC algorithm. The experimental results show that the proposed method is more robust than the RANSAC algorithm. In addition, our method reduces the computational cost of RANSAC.

The rest of this paper is organized as follows: Section II briefly describes some related work. In Section III we present the proposed method. Data normalization is first introduced, and the order-consistent constraint is proposed by the relative order of two sets of feature points and the minimum of the similarity distance using the normalized data. The proposed homography estimation method is evaluated using synthetic data and real images in Section IV and the conclusions are given in Section V.

II. RELATED WORK

Suppose the projection matrices of two images are $[I, 0]$ and $[A, a]$, and a space plane is defined by $\pi = [v^T, 1]^T$, the homography between the two images of the planar surface π in the scene is expressed by an explicit function [1]

$$H = A - av^T$$

In general, the plane π can not contain either of the two camera centers [1]. That is to say, At least four pairs of corresponding points (non-collinear) are needed to estimate the homography between the two images.

Under the framework of RANSAC algorithm, hypothetical correspondences are necessary, and non-collinear four point correspondences are selected randomly to estimate an initial homography between two views or an image and a space plane [1].

To reduce the repetitive fitting model, Márquez-Neila *et al.* [17] presented a geometric constraint on three points order-preserving in each random sample (selecting four point correspondences). As illustrated in Fig. 1, the corresponding points A′, B′ and C′ in the second image have the same relative order as points A, B and C in the first image. If this geometric constraint does not hold, this set of correspondences selected in each random sample should be discarded. Therefore, when the homography is fitted, this geometric constraint is applied up to four times. A similar constraint was previously used to remove mismatches in Ferrari *et al.* [18], and it was applied to estimate homography.
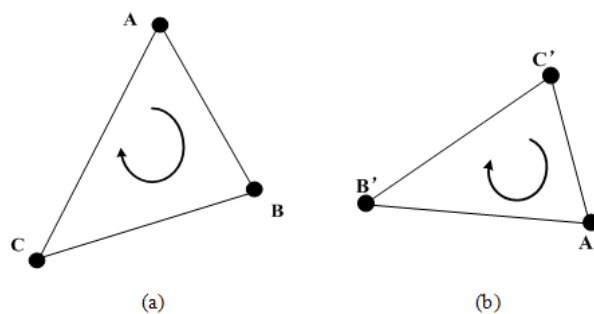

FIGURE 1. Paradigm of geometric constraint in two images: (a) Points A, B and C in first image, (b) Corresponding points A′, B′ and C′.

However, this constraint only considers the order-preserving relationship between points A, B and C and corresponding points A′, B′ and C′ in two images.

Because of the importance of shape recognition in computer vision and multimedia applications involving shape analysis, image retrieval, and visualization, polygon similarity measurement has attracted the attention of some researchers in recent years. Werman and Weinsahll [19] presented the expression for the 2D similarity distance, which is invariant to affine transformation or similarity transformation, assuming a known correspondence between the point sets. Nguyen and Hoang [20] proposed a generic approach for the polygonal measurement by digitization and similarity transformation in the projection space. Nacera *et al.* [21]

employed the polygonal approximation to represent the outline shape of an object and proposed a recognition method based on polynomial curve similarity measurement. Most such methods carry out polynomial curve fitting to establish shape similarity or to identify object shape.

Therefore, the feature point order-preserving constraint and the similarity measurement are investigated to find four point correspondences between two images in this paper. Compared with the RANSAC method the proposed method does not need many iterations for the estimation of the homography and thus requires less computation time.

III. PROPOSED METHOD
In this paper, we assume that the camera follows a pinhole model, and only points in front of the camera are visible. If a set of point correspondences between two images is known, we can select four point correspondences to compute the homography. The configuration of the four-point is usually constructed by a set of ordered points in two views (see in Fig.2).
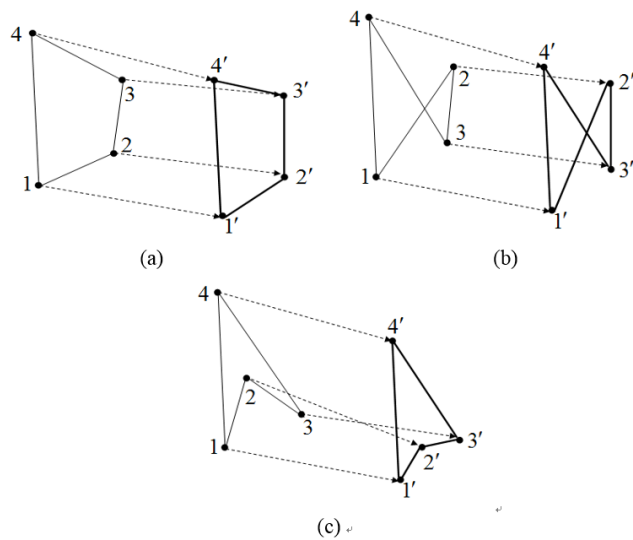

FIGURE 2. Paradigm of the changed four-point shape in two images: (a) Convex (b) Cross shape and (c) Concave.

The numbers 1, 2, 3, 4 in Fig. 2 represent the four feature points in one image, and the four feature points in another image are numbered $1'$, $2'$, $3'$, $4'$, respectively. The dashed line represents the corresponding path of each point. It can be seen from Fig. 2 that the three cases occur during the different camera positions or the different corner detection between two images. Fig. 2(a) depicts the convex-order for four correspondence points; Fig. 2(b) depicts the twist-order for four correspondence points; and Fig. 2(c) depicts the concave-order for four correspondence points between the two images.

In fact, the proposed method, in essence, addresses two problems in this paper. The first problem is whether the relative order of points 1, 2, 3, 4 and that of points $1'$, $2'$, $3'$, $4'$ is consistent; and the second is whether the fitting error

of the homography approaches a minimum. We describe the proposed method in the following.

A. OUTLINE OF THE PROPOSED METHOD
The overall flow of the proposed method is shown in Algorithm 1, where $\{\mathbf{P}_1, \mathbf{P}_2\}$ are the set of the initial feature point correspondences between two images I_1 and I_2 using the SURF operator in OpenCV [21], and $\{\hat{\mathbf{P}}_1, \hat{\mathbf{P}}_2\}$ are the normalized data of $\{\mathbf{P}_1, \mathbf{P}_2\}$.

Algorithm 1 The Proposed Method

1: **input**: I_1, I_2, $\{\mathbf{P}_1, \mathbf{P}_2\}$
2: **procedure**: *NormalizedData*($\{\mathbf{P}_1, \mathbf{P}_2\}$)
3: $\{\mathbf{P}_1, \mathbf{P}_2\} \leftarrow \{\hat{\mathbf{P}}_1, \hat{\mathbf{P}}_2\}$;
4: **while** $Q(k) > \varepsilon$ or $Q(k) < Q(k-1)$ **do**
5:    random selected $\{\hat{p}_{1i} \leftrightarrow \hat{p}_{2i}\}$ ($i = 1, 2, 3, 4$);
6:    estimated order-consistent constraint;
7:    $d_s \leftarrow norm(\hat{p}_{1i}, \hat{p}_{2i})$;
8:    if $d_s < T$
9:      $H \leftarrow$ homography (p_{1i}, p_{2i});
10:    $Q(k) \leftarrow \sum_{j=1}^{n} \left(d^2(p_{2j}, Hp_{1j}) + d^2(p_{1j}, H^{-1}p_{2j})\right)$
11:    end
12:    $k \leftarrow k+1$;
13:   **end while**
14:     **return** H;
15: **end procedure**

We select four initial matched points randomly $\{\hat{p}_{1i} \leftrightarrow \hat{p}_{2i}\}$ ($i = 1, 2, 3, 4$) in $\{\hat{\mathbf{P}}_1, \hat{\mathbf{P}}_2\}$ and determine the corrected point correspondences according to the order-consistent constraint and the similarity distance d_s. Suppose that the four point distances $\{\hat{p}_{1i} \leftrightarrow \hat{p}_{2i}\}$ ($i = 1, 2, 3, 4$) are the correctly matched points; a homography H can then be estimated. Furthermore, the homography may be optimized by minimizing the objective function $Q(k)$, where $Q(k)$ is the re-projective error between the matched points in the two images I_1 and I_2.

B. STAGE 1: NORMALIZATION
Most methods exploit data normalization to improve the accuracy of results and reduce the effect of coordinates changes of the measurement data [1], [31]. Here, we assume that space objects are composed of n three-dimensional points. The two images I_1 and I_2 are obtained by a rigid transformation. It is important to carry out the data normalization in the 2D homography estimation [1].

Let $\mathbf{P}_1 = \{p_{11}, p_{12}, \cdots, p_{1n}\}$ be the set of 2D image points in I_1, and let $\mathbf{P}_2 = \{p_{21}, p_{22}, \cdots, p_{2n}\}$ be the set of 2D image points in I_2, where $\{p_{1j} = (x_{1j}, y_{1j})\}_{j=1}^{n}$ and $\{p_{2j} = (x_{2j}, y_{2j})\}_{j=1}^{n}$. The set of 2D feature points in the image are normalized by the following steps.

1) Calculate the centers $\boldsymbol{p}_{1c} = \left(x_1^*, y_1^*\right)$ and $\boldsymbol{p}_{2c} = \left(x_2^*, y_2^*\right)$ of two point sets $\boldsymbol{P}_1$ and $\boldsymbol{P}_2$ in terms of equation (1).

$$x_i^* = \sum_{j=1}^{n} x_{ij} \Big/ n, \ y_i^* = \sum_{j=1}^{n} y_{ij} \Big/ n, \quad (i = 1, 2) \quad (1)$$

2) Normalize the position of the image point set by equations (2) and (3).

$$\tilde{\mathbf{P}}_1 = \left\{\boldsymbol{p}_{11} - \boldsymbol{p}_{1c}, \boldsymbol{p}_{12} - \boldsymbol{p}_{1c}, \cdots, \boldsymbol{p}_{1n} - \boldsymbol{p}_{1c}\right\} \quad (2)$$
$$\tilde{\mathbf{P}}_2 = \left\{\boldsymbol{p}_{21} - \boldsymbol{p}_{2c}, \boldsymbol{p}_{22} - \boldsymbol{p}_{2c}, \cdots, \boldsymbol{p}_{2n} - \boldsymbol{p}_{2c}\right\} \quad (3)$$

3) Compute the 2-norm of the normalized data $\tilde{\mathbf{P}}_1$ and $\tilde{\mathbf{P}}_2$ by equations (4) and (5).

$$\boldsymbol{D}_1 = \left\{\left\|\boldsymbol{p}_{11} - \boldsymbol{p}_{1c}\right\|_2, \left\|\boldsymbol{p}_{12} - \boldsymbol{p}_{1c}\right\|_2, \cdots, \left\|\boldsymbol{p}_{1n} - \boldsymbol{p}_{1c}\right\|_2\right\}$$
$$(4)$$
$$\boldsymbol{D}_2 = \left\{\left\|\boldsymbol{p}_{21} - \boldsymbol{p}_{2c}\right\|_2, \left\|\boldsymbol{p}_{22} - \boldsymbol{p}_{2c}\right\|_2, \cdots, \left\|\boldsymbol{p}_{2n} - \boldsymbol{p}_{2c}\right\|_2\right\}$$
$$(5)$$

4) Determine the max 2-norm in $\boldsymbol{D}_1$ and $\boldsymbol{D}_2$ by $d_1 = \max\left(\boldsymbol{D}_1\right)$ and $d_2 = \max\left(\boldsymbol{D}_2\right)$

5) Normalize the scale of the image point set from equations (6) and (7).

$$\hat{\mathbf{P}}_1 = \tilde{\mathbf{P}}_1 / d_1 \quad (6)$$
$$\hat{\mathbf{P}}_2 = \tilde{\mathbf{P}}_2 / d_2 \quad (7)$$

The above normalization enables us to compare the two sets of image points in the different images under the same position and scale.

C. STAGE 2: ORDER-CONSISTENT CONSTRAINT AND SIMILARITY MEASUREMENT

The correct point matches between the two images are very important for accurate homography estimation. In this section, an order-consistent constraint for four points in the two images is introduced. The order-consistent constraint consists of the relative order of two sets of feature points and the minimum of the similarity distance using the normalized data.

As illustrated in Fig. 3, $\hat{\boldsymbol{p}}_{11}, \hat{\boldsymbol{p}}_{12}, \hat{\boldsymbol{p}}_{13}, \hat{\boldsymbol{p}}_{14}$ represent the normalized feature points inI_1, and points $\hat{\boldsymbol{p}}_{21}, \hat{\boldsymbol{p}}_{22}, \hat{\boldsymbol{p}}_{23}, \hat{\boldsymbol{p}}_{24}$ are the correspondences in I_2. The relative order of two sets of feature points is determined via the following steps.

First, we compute the vectors $\overrightarrow{\hat{\boldsymbol{p}}_{11}\hat{\boldsymbol{p}}_{12}}, \overrightarrow{\hat{\boldsymbol{p}}_{12}\hat{\boldsymbol{p}}_{13}}, \overrightarrow{\hat{\boldsymbol{p}}_{13}\hat{\boldsymbol{p}}_{14}}, \overrightarrow{\hat{\boldsymbol{p}}_{14}\hat{\boldsymbol{p}}_{11}}$ and $\overrightarrow{\hat{\boldsymbol{p}}_{21}\hat{\boldsymbol{p}}_{22}}, \overrightarrow{\hat{\boldsymbol{p}}_{22}\hat{\boldsymbol{p}}_{23}}, \overrightarrow{\hat{\boldsymbol{p}}_{23}\hat{\boldsymbol{p}}_{24}}, \overrightarrow{\hat{\boldsymbol{p}}_{24}\hat{\boldsymbol{p}}_{21}}$ for four points in the two images, respectively.

Next, the upper and lower directions of these vectors along the x axis direction are determined using the normalized coordinates. If the vector direction is upper, the vector is marked by "1". Otherwise, the vector is marked by "0". Let the vector direction sets be S_1 and S_2 which are given by $S_1 = \{1, 1, 1, 0\}$ and $S_2 = \{1, 1, 1, 0\}$, as shown in Fig. 3.

Finally, the relative order between the two sets of the normalized points is definitively established from the vector direction.
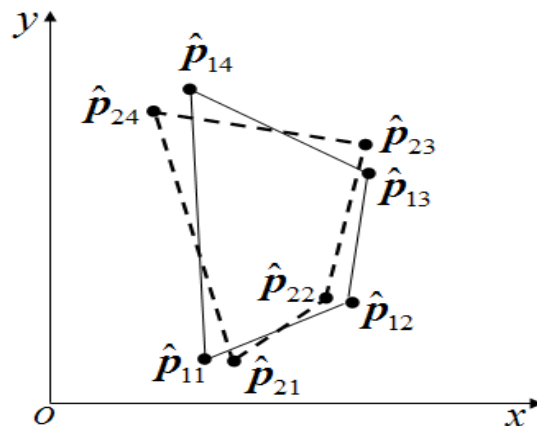

FIGURE 3. An illustration of four feature point normalization. Points $\hat{p}_{11}, \hat{p}_{12}, \hat{p}_{13}$, and $\hat{p}_{14}$ are on the first image, and points $\hat{p}_{21}, \hat{p}_{22}, \hat{p}_{23}$, and $\hat{p}_{24}$ are on the second image.

The similarity measurement is extensively applied in shape recognition, image retrieval, and visualization [19]–[21], [29], [30]. Nacera *et al.* [21] proposed a two-step matching algorithm using a shape similarity between the input image and its reference image. Luo *et al.* [29], [30] proposed a stacked extreme learning machine with sparse autoencoder and a quantized kernel least mean square scheme using the similarity measurement of two arbitrary random variables for data analysis. In this paper, we aim to find the matches of adjacent frames in video sequences. Thus, the camera motion between the two adjacent frames is relatively small, and the similarity distance between the two sets of normalized points is reinforced for the feature point match and is defined by

$$d_s^2\left(\hat{\boldsymbol{p}}_{1j}, \hat{\boldsymbol{p}}_{2j}\right) = \sum_{j=1}^{4}\left(\left\|\hat{\boldsymbol{p}}_{1j} - \hat{\boldsymbol{p}}_{2j}\right\|\right) \quad (8)$$

where $\hat{\boldsymbol{p}}_{1j}$ $(j = 1, 2, 3, 4)$ are the normalized points in the first image, and $\hat{\boldsymbol{p}}_{2j}$ $(j = 1, 2, 3, 4)$ are the normalized points in the second image.

Therefore, the matching between the two normalized point sets $\left\{\hat{\boldsymbol{p}}_{1j}\right\}$ $(j = 1, 2, 3, 4)$ and $\left\{\hat{\boldsymbol{p}}_{2j}\right\}$ $(j = 1, 2, 3, 4)$ is obtained by comparing their vector direction sets S_1 with S_2 within a threshold T subject to

$$d_s^2\left(\hat{\boldsymbol{p}}_{1j}, \hat{\boldsymbol{p}}_{2j}\right) \leq T \quad (9)$$

D. STAGE 3: HOMOGRAPHY ESTIMATION

After selecting four point correspondences from the hypothetical point matches in two images, we can estimate the homography.

Suppose that four points $\boldsymbol{p}_{1j}$ $(j = 1, 2, 3, 4)$ are chosen randomly from the point set $\mathbf{P}_1$ in I_1 and that four corresponding points $\boldsymbol{p}_{2j}$ $(j = 1, 2, 3, 4)$ are obtained from the point set $\mathbf{P}_2$ in I_2. $\boldsymbol{H}$ is a regular matrix representing a homography and is

written by

$$H = \begin{pmatrix} h_{11} & h_{12} & h_{13} \\ h_{21} & h_{22} & h_{23} \\ h_{31} & h_{32} & h_{33} \end{pmatrix}$$

Thus, our homography is estimated from equation (10)

$$\lambda_j \cdot p_{2j} = H \cdot p_{1j} \tag{10}$$

where λ_j $(j = 1, 2, 3, 4)$ is a scalar.

Let a popular loss function [1] $Q\left(p_{1j}, p_{2j}\right)$ $(j = 1, \cdots, n)$ be

$$Q\left(p_{1j}, p_{2j}\right) = \sum_{j=1}^{n} \left(d^2\left(p_{2j}, Hp_{1j}\right) + d^2\left(p_{1j}, H^{-1}p_{2j}\right)\right) \tag{11}$$

where $d\left(\cdot, \cdot\right)$ stands for the squared Euclidean distance [1]. After the initial homography is estimated, we select the optimum homography H that minimizes the loss function (11) [1].

IV. EXPERIMENTAL RESULTS

In this section, we report the experiments on both synthetic data and real images to verify the proposed method. In the synthetic experiment, we study the influence of the proportion of outliers and noise levels on the homography estimation. In the real image experiments, we analyze the computational cost of the proposed method and compare our method with other related methods.

A. SYNTHETIC EXPERIMENTS
The synthetic data were generated as follows. First, the intrinsic matrix of the simulated camera is

$$K = \begin{pmatrix} 1000 & 0.03 & 320 \\ 0 & 960 & 240 \\ 0 & 0 & 1 \end{pmatrix}$$

The two rotation vectors and two rotation angles are

$$r_1 = (0.44, 0.21, 0.87)^T, \quad \theta_1 = \pi/8$$
$$r_2 = (0.82, 0.41, 0.40)^T, \quad \theta_2 = \pi/6$$

and the two translation vectors are

$$t_1 = (20, 4, 250)^T \ t_2 = (40, 50, 320)^T$$

Second, we generated 50 space points randomly and two simulated images were obtained by the simulated camera parameters. The simulated images are 800×800 pixels in size. Finally, we add a random Gaussian value with mean zero and standard deviation $\delta \in [0, 5.0]$ pixels to the 2D image points

The first synthetic experiment is to validate the order-consistent constraint. The image points were added to the Gaussian noise with a standard deviation of 5.0 pixels. The first 30 are inliers, and the remaining 20 are outliers. After the image points were normalized by the III.B method, the order-consistent constraint is validated using the normalized data.

TABLE 1. The influence of the threshold on the matching rate.

The similarity distance threshold T	The correct matches	The mismatches
0.2	49	1
0.3	48	2
0.4	46	4
0.5	39	11

In this experiment, we run many random tests for four correspondences points between the two synthetic images. The four red numbers in Fig. 4 represent the order of feature points in the two images, where the order of the convex shape is (6, 23, 17, 25), that of the cross shape is (6, 8, 30, 23), and that of the concave shape is (12, 22, 29, 25).

We also tested the matching rate by varying the similarity distance threshold. The result is listed in TABLE 1, where four thresholds are presented. At each threshold, we run 50 trials. This is a clear demonstration that when the similarity distance threshold is less than 0.5, the four point correspondences between two images are matched with a higher correct rate. The experimental results corroborate the proposed constraint.

We performed some comparative experiments based on the normalized points and non-normalized points for homography estimation between the two synthetic images. 50 points in the two synthetic images were generated, and each point was added 2.5 and 4.5 pixels Gaussian noises. We randomly chose corresponding points from 50 points to compute the homography. The homography was repeated 30 times, and the average of the reprojection errors were compared. The results are shown in TABLE 2. We can see that data normalization substantially improves the accuracy of the estimated homography.

TABLE 2. The average of the reprojection error using homography based on the normalized points and non-normalized points.

Noises	The number of the corresponding points	The reprojection errors using the non-normalized data	The reprojection errors using the normalized data
4.5	4	1763.50	13.86
	8	1076.92	10.72
	16	937.71	6.46
2.5	4	1421.43	11.65
	8	806.25	10.51
	16	791.14	6.52

In the second experiment, we added Gaussian noise with a standard deviation of 3.0 pixels to the image point. The proportion of inliers is varied from 30% to 60%. The proposed method and the RANSAC method [1, 10] were performed to compute the homography between the two images. In the proposed method the similarity distance threshold is set at 0.3. The experiment was repeated 50 times. The Euclidean distances between the computed
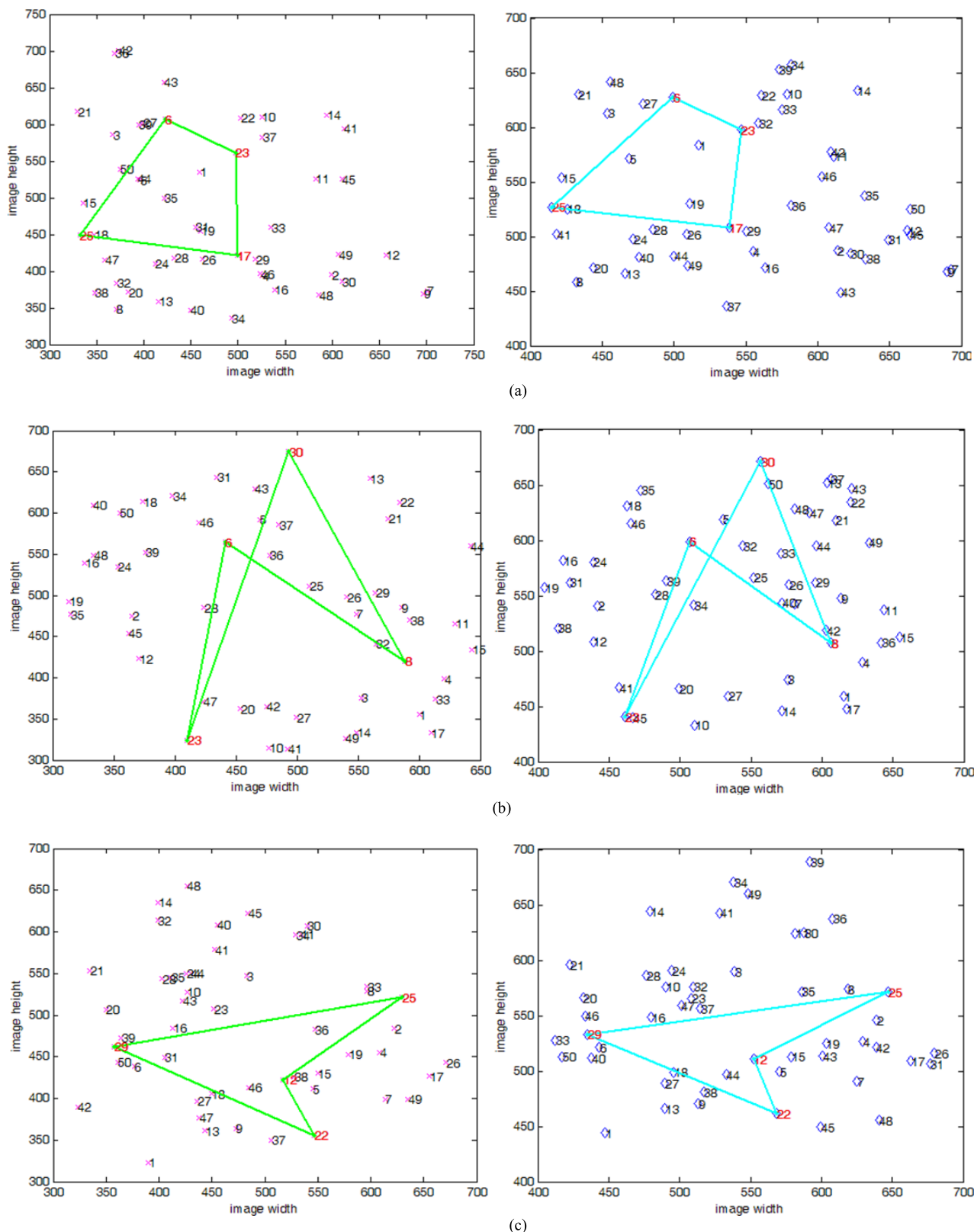
FIGURE 4. Experimental results with the order-consistent constraint: (a) Convex, (b) Cross shape and (c) Concave.

homography and the true homography were calculated at the different inlier proportions, and the average distance is illustrated in Fig. 5. We can see that when the proportion of the inliers is less than 45%, the RANSAC method [1, 10] has a higher fitting error for the estimated homography.
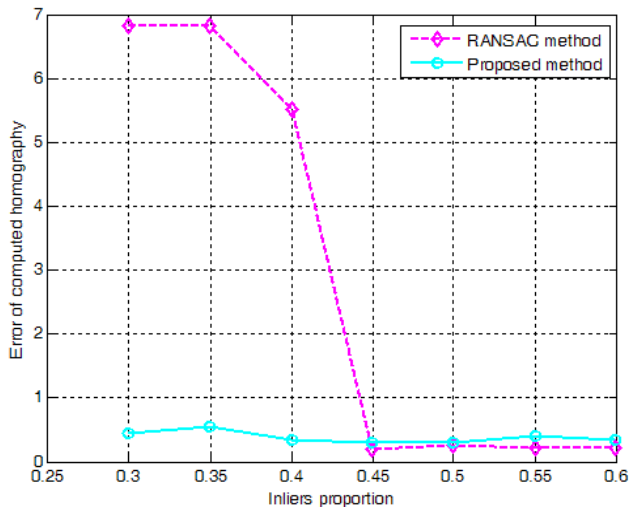
FIGURE 5. Average error of the estimated homography at different inlier proportions.

FIGURE 6. Average error between computed homography and true homography at different noise levels.

Finally, the proportion of outliers is chosen as 60%, and the image points have added Gaussian noise with standard deviations at 1.0, 2.0, 3.0, 4.0 and 5.0 pixels. We estimated the homography using the proposed method and the RANSAC method [1, 10], and the Euclidean distances [1] between the computed homography and the true homography were calculated at different noise levels. The similarity distance threshold is set at 0.3 in the proposed method. 50 trials were carried out in this experiment, and the average error at each noise level is plotted in Fig. 6. These results indicate that the proposed method is very effective and outperforms the RANSAC method at the higher noise levels (4.0 pixels and 5.0 pixels).

We also found that the RANSAC method needs to fit more than 260 homographies compared to the proposed method that needs to fit only approximately 10 homographies. These experimental results with the synthetic data prove that the proposed method reduces the number of fitted homographies and has better robustness.

B. REAL EXPERIMENTS

This section presents two experiments with the static real images and one experiment with a dynamic image sequence. Our goal here is to verify the correct rate of the true inliers and the computation cost using the proposed method.

In these experiments, the SURF key point detector [26] is first used, and the putative correspondence set between two images is obtained by taking the brute-force descriptor matcher in OpenCV [22]. The threshold value is less than 0.2 times the Euclidean distance between one feature point from the first image and the other feature point from the second image.

First, we performed the experiment to verify the correct rate of the true inliers. Fig. 7 shows an example, where figure (a) includes five inliers (marked by green lines)
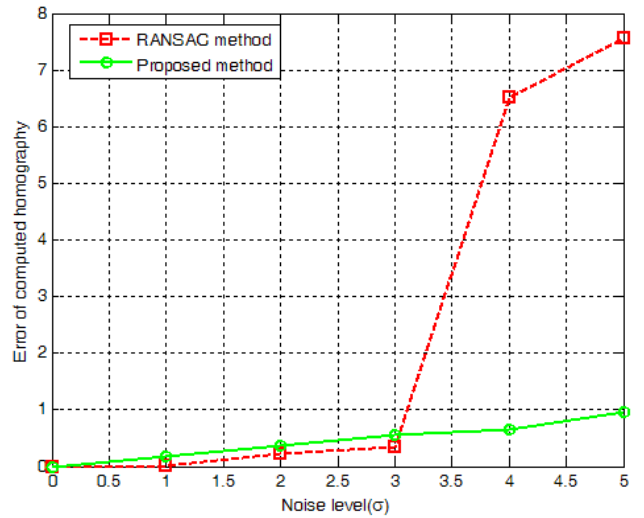
and seven outliers (marked by red dotted lines), and figure (b) includes six inliers (marked by green lines) and seven outliers (marked by red dotted lines). The RANSAC method [1] (*FindHomography* function in OpenCV [21]) and the proposed method were implemented to find the correct matches. When the sum of the squared Euclidean distance is set to 3, the number of correct matches is 5 using the RANSAC method and the proposed method (shown in Fig. 7(c) and (e)). We can see that there are two mismatches (marked by red ellipses in Fig.7 (c)) using the RANSAC method. Furthermore, the proposed method only needs to estimate the homographies twice, whereas the RANSAC method requires 173 times.

In another group of experiments, the RANSAC method found 5 matches by fitting 239 homographies along with two mismatches (marked by red ellipses), as depicted in Fig. 7 (d). The proposed method only determines four correct matches by fitting 7 homographies (shown in Fig.7 (f)). The proposed method thus greatly reduces the number of fitted homographies.

Next, we applied the proposed method, the RANSAC method and the Márquez-Neila method [17] to a panoramic image mosaic. The SURF detector [23] and brute-force matcher in OpenCV [22] were adopted in this experiment. Fig. 8 (a) and (b) compare the image mosaic results for a pair of images using three kinds of methods. Fig. 8 (c) shows the stitching result by the proposed method, Fig. 8 (d) illustrates the result using the RANSAC method [1], and Fig. 8 (e) is the result using Márquez-Neila's method [16]. From the stitching effect point of view, image (c) is better than images (d) and (e). We also find that the building in the top left corner of the stitching result is enlarged in Fig. 8 (d) and (e).

Furthermore, we measured the computation time of the proposed method, the RANSAC method and the Márquez-Neila method [17]. The experiment was run on an Intel(R)
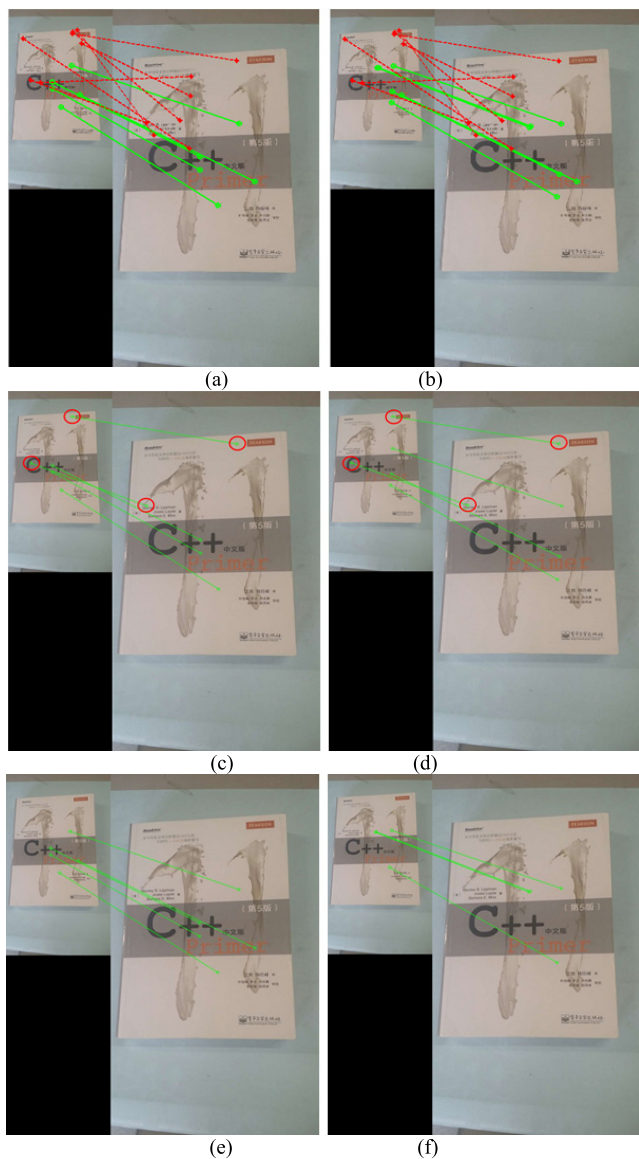
FIGURE 7. Real image experiments. (a) includes five inliers and seven outliers, and there are six inliers and seven outliers in (b); (c) and (d) are the matching results with the RANSAC method, (e) and (f) are the matching results using the proposed method. In these experiments, the size of the left image is 250×444 pixels, and that of the right image is 500×888 pixels. Green lines represent the inliers and red dotted lines stand for the outliers. The mismatches found by the RANSAC method are marked by red ellipses.

Core(TM) CPU i7-5500U @ 2.40GHz, 4GB RAM and Windows 7 Professional 32bit OS. We tested 50 pairs of images by fitting 100 homographies. Table 3 shows the computation times of the three methods. The computation of the Márquez-Neila method [17] is faster than that of the RANSAC method, whereas our method performs the fastest.

Finally, we present a real image sequence experiment. A video of an award posted on a wall is taken in this experiment. The putative correspondence points are achieved using the SURF detector and brute-force descriptor matcher in OpenCV [22] for each frame of the video. We compared the
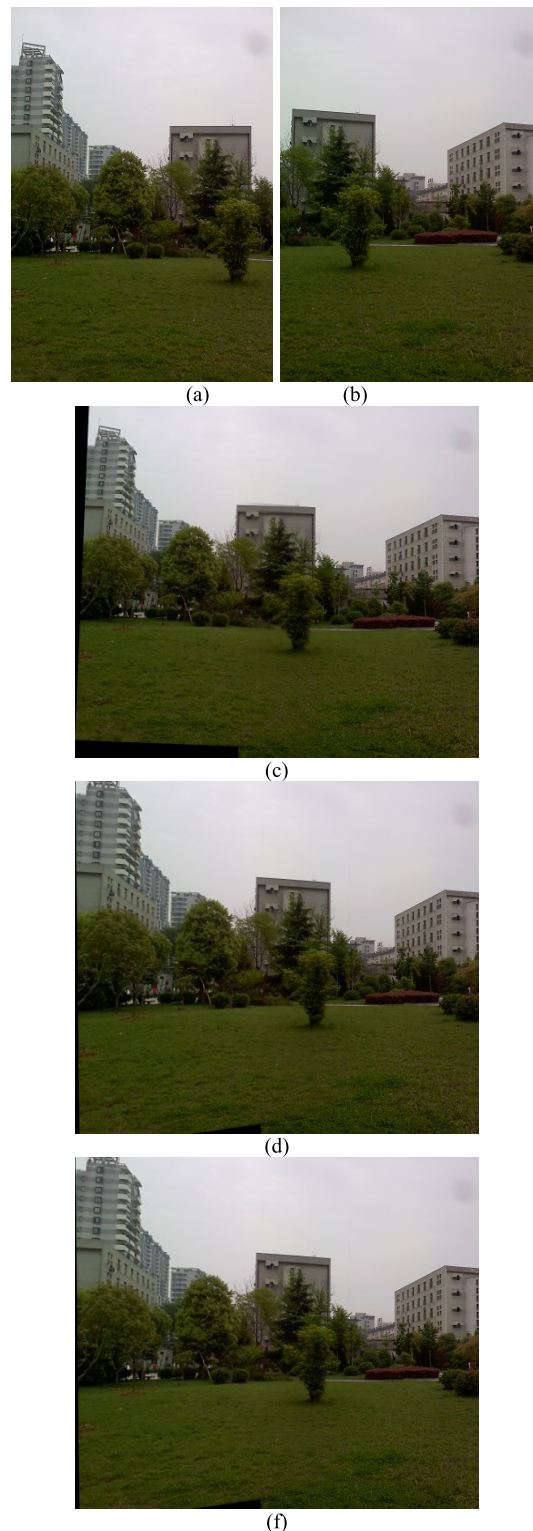

FIGURE 8. Image stitching experiments. (a) and (b) two original images, (c) the stitching result with the proposed method, (d) the stitching result using the RANSAC method [1], and (e) the stitching result using the Márquez-Neila method [17].

RANSAC method (implemented by adopting *FindHomography* in OpenCV [22]) and the Márquez-Neila method [17] with the proposed method. The putative feature point matches
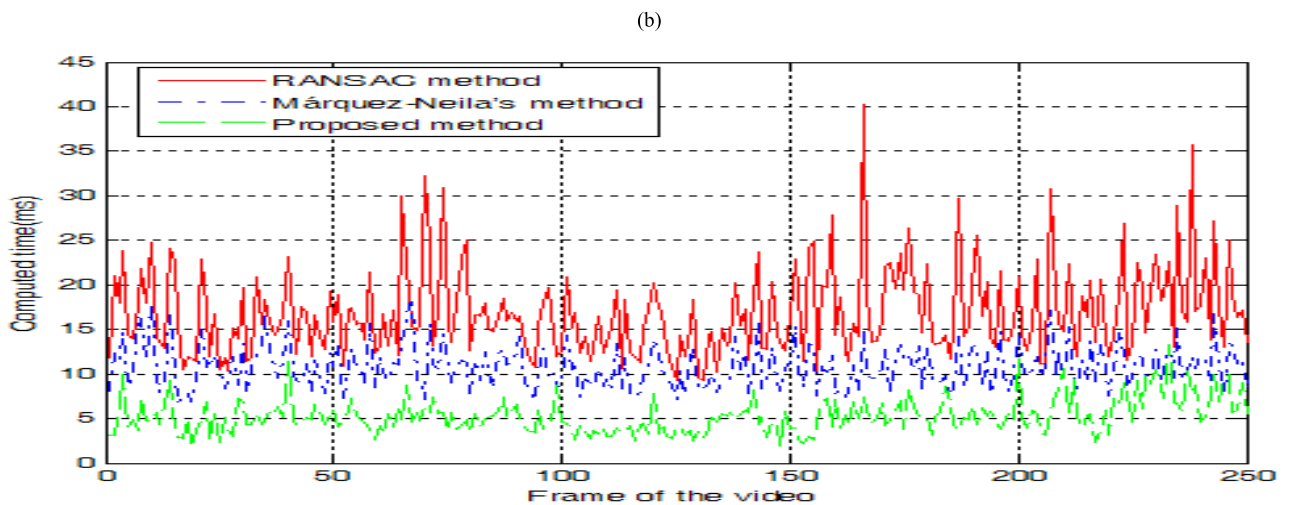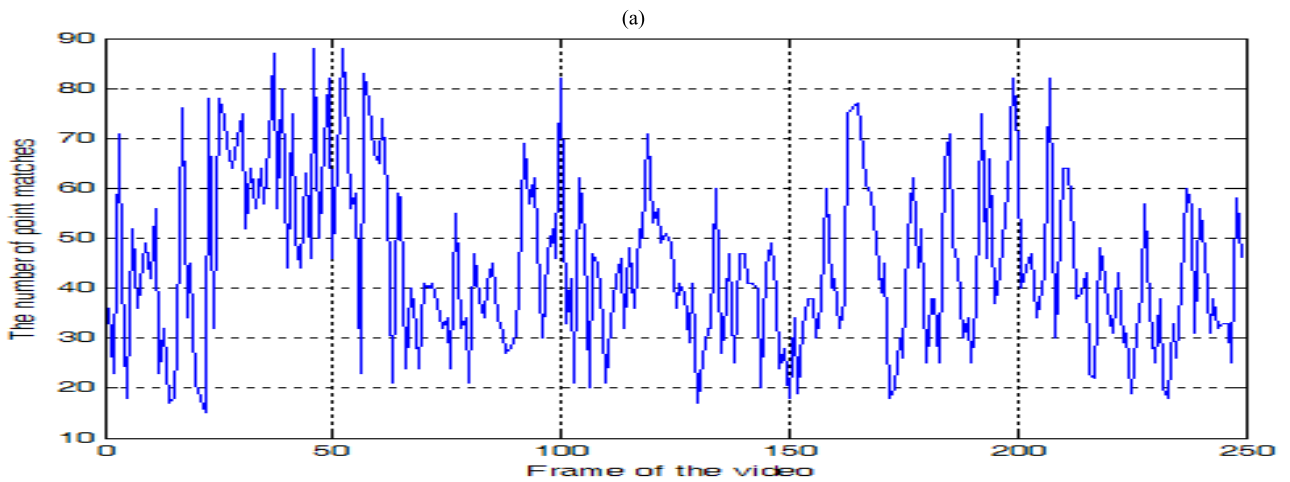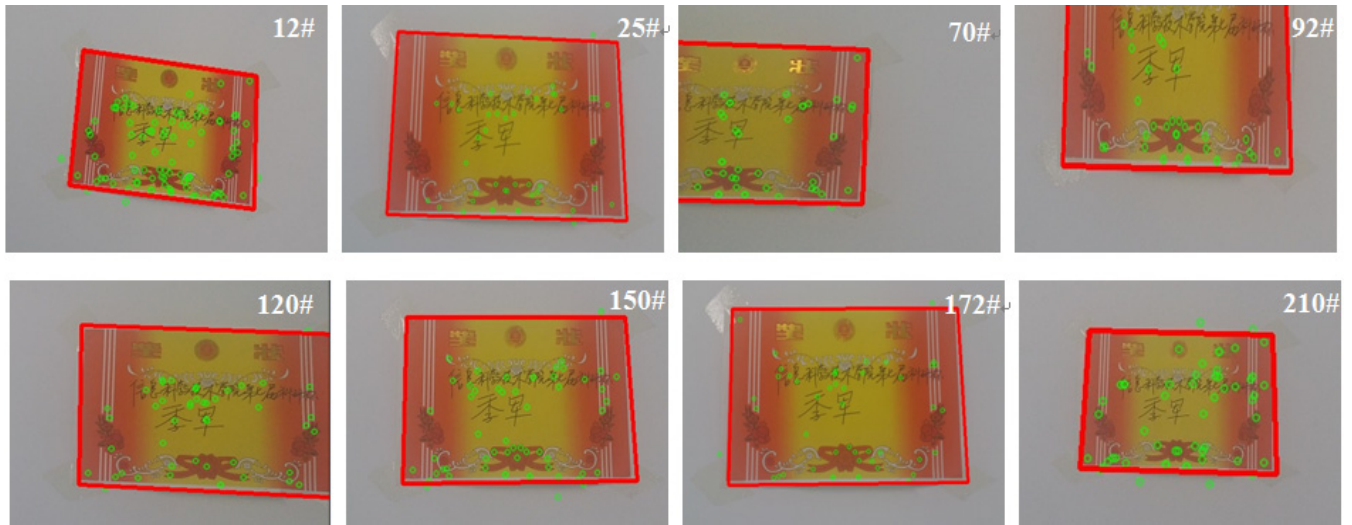
(a)

(b)

(c)

FIGURE 9. Real video image experiments. (a) Putative feature point matches in green "o" and the estimated homography that covered the outline of the awards on the per fame (marked in red line); (b) The number of putative matches with each video frame; (c) The computed time of homography for the RANSAC method with OpenCVŠs FindHomography, Márquez-NeilaŠs method [17] and the proposed method.

are marked by a green "o" in Fig. 9 (a), and the estimated homography overlaid on each frame in the video is marked by red lines. Fig. 9 (b) presents the number of putative matches on each frame, and Fig. 9 (c) displays the execution times of the homography estimation for the RANSAC method with OpenCV's *FindHomography*, the

TABLE 3. Time to fit 100 homography on 2.40GHz CPU and 4GB RAM.

Method	Total time (Unit: ms)
Proposed method	32.99
RANSAC method	90.59
Márquez-Neila method [17]	78.86

TABLE 4. Comparison of iterative times using the proposed method and the RANSAC method.

Method	The average number of iterations
Proposed method	12
RANSAC method	113

Márquez-Neila method [17] and the proposed method. These experimental results further indicate that the proposed method significantly reduces the computation time for videos.

In addition, we compared the number of iterations using the proposed method and the RANSAC method [1] while the homography was estimated using 30 frames of the sequence in Fig. 9. The number of iterations for calculating homography is listed in Table 4. The experimental result demonstrates that the proposed method requires much less iterations than the RANSAC algorithm for homography estimation. Here the RANSAC algorithm is implemented using the FindHomography function in OpenCV [21].

V. CONCLUSIONS

We have described a fast and accurate homography estimation in this paper. The main contributions of this work are the order-preserving constraint-based normalized data and the similarity measurement of four point correspondences between the two images. Under the presence of large noise and outlier proportions, the proposed method achieved a higher correct rate of point matches to estimate the homography.

In the synthetic experiments, larger noises and outlier proportions were added to the feature points. The proposed method requires fewer iterations and achieves higher accuracy, under different levels of noise and outlier ratios. When the proportion of inliers is above 45%, the correct rate of point matches of the proposed method is below that of the RANSAC method. The proposed method was clearly less computationally expensive than the RANSAC method and Márquez-Neila's method [17] in the experiment with real images. In the future, we will apply the proposed method to the mostpervasive digital apparatuses, such as mobile phones

REFERENCES

[1] R. I. Hartley and A. Zisserman, *Multiple View Geometry in Computer Vision*, 2nd ed. Cambridge, U.K.: Cambridge Univ. Press, 2004.

[2] G. López-Nicolás, N. R. Gans, S. Bhattacharya, C. Sagüés, J. J. Guerrero, and S. Hutchinson, "Homography-based control scheme for mobile robots with nonholonomic and field-of-view constraints," *IEEE Trans. Syst., Man, Cybern. B, Cybern.*, vol. 40, no. 4, pp. 1115–1127, Aug. 2010.

[3] G. Klein and D. Murray, "Parallel tracking and mapping on a camera phone," in *Proc. IEEE Int. Symp. Mixed Augmented Reality*, Orlando, FL, USA, Oct. 2009, pp. 83–86.

[4] P. Chen, P. Zhang, D. Li, and L. Yang, "An improved augmented reality system based on AndAR," *J. Vis. Commun. Image Represent.*, vol. 37, pp. 63–69, May 2015.

[5] Y. Xiong and K. Pulli, "Fast image stitching and editing for panorama painting on mobile phones," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. Workshops*, San Francisco, CA, USA, Jun. 2010, pp. 47–52.

[6] J. Zaragoza, T.-J. Chin, Q.-H. Tran, M. S. Brown, and D. Suter, "As-projective-as-possible image stitching with moving DLT," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 36, no. 7, pp. 1285–1298, Jul. 2013.

[7] S.-Y. Lee, J.-Y. Sim, C.-S. Kim, and S.-U. Lee, "Correspondence matching of multi-view video sequences using mutual information based similarity measure," *IEEE Trans. Multimedia*, vol. 15, no. 8, pp. 1719–1730, Dec. 2013.

[8] Y. Sun, L. Zhao, S. Huang, L. Yan, and G. Dissanayake, "Line matching based on planar homography for stereo aerial images," *ISPRS J. Photogram. Remote Sens.*, vol. 104, pp. 1–17, Jun. 2015.

[9] Y. Zhang, L. Zhou, Y. Shang, X. Zhang, and Q. Yu, "Contour model based homography estimation of texture-less planar objects in uncalibrated images," *Pattern Recognit.*, vol. 52, pp. 375–383, Apr. 2016.

[10] M. A. Fishler and R. C. Bolles, "Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography," *Commun. ACM*, vol. 6, pp. 381–395, Jun. 1981.

[11] P. H. S. Torr, C. Davidson, "IMPSAC: Synthesis of importance sampling and random sample consensus," in *Proc. Eur. Conf. Comput. Vis.*, Dublin, Ireland, 2000, pp. 819–833.

[12] B. J. Tordoff and D. W. Murray, "Guided-MLESAC: Faster image transform estimation by using matching priors," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 27, no. 10, pp. 1523–1535, Oct. 2005.

[13] J. Matas and O. Chum, "Randomized RANSAC with $T_{d,d}$ test," in *Proc. Brit. Mach. Vis. Conf.*, Cardiff, U.K., 2002, pp. 448–457.

[14] O. Chum and J. Matas, "Matching with PROSAC—Progressive sample consensus," in *Proc. Int. Conf. Comput. Vis. Pattern Recognit.*, San Diego, CA, USA, Jun. 2005, pp. 220–226.

[15] C.-M. Cheng and S.-H. Lai, "A consensus sampling technique for fast and robust model fitting," *Pattern Recognit.*, vol. 42, pp. 1318–1329, Jul. 2009.

[16] P. Bhattacharya and M. Gavrilova, "Improving RANSAC feature matching with local topological information," in *Proc. 9th Int. Symp. Voronoi Diagrams Sci. Eng.*, New Brunswick, NJ, USA, Jun. 2012, pp. 17–23.

[17] P. Márquez-Neila, J. López-Alberca, J. M. Buenaposada, and L. Baumela, "Speeding-up homography estimation in mobile devices," *J. Real-Time Image Process.*, vol. 11, no. 1, pp. 141–154, 2016.

[18] V. Ferrari, T. Tuytelaars, and L. Val Gool, "Wide-baseline multiple-view correspondences," in *Proc. CVPR*, Madison, WI, USA, Jun. 2003, pp. 718–725.

[19] M. Werman and D. Weinshall, "Similarity and affine invariant distances between 2D point sets," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 17, no. 8, pp. 810–814, Aug. 1995.

[20] T. P. Nguyen and T. V. Hoang, "Projection-based polygonality measurement," *IEEE Trans. Image Process.*, vol. 24, no. 1, pp. 305–315, Jan. 2015.

[21] L. Nacera, L. Slimane, L. Farouk, and A. Khadraoui, "Curve normalization for shape retrieval," *Signal Process., Image Commun.*, vol. 29, pp. 556–571, Apr. 2014.

[22] G. Bradski and A. Kaehler, *Learning OpenCV 3: Computer Vision in C++ With the OpenCV Library*. Newton, MA, USA: O'Reilly Media, Dec. 2016.

[23] R. Szeliski, *Computer Vision: Algorithms and Applications*. New York, NY, USA: Springer, 2010, pp. 425–462.

[24] R. Du and T. Padir, "Image stitching techniques for an intelligent portable aerial surveillance system," in *Proc. IEEE Int. Conf. Technol. Practical Robot Appl.*, Woburn, MA, USA, Apr. 2014, pp. 1–6.

[25] A. Hamza, R. Hafiz, M. M. Khan, Y. Cho, and J. Cha, "Stabilization of panoramic videos from mobile multi-camera platforms," *Image Vis. Comput.*, vol. 37, pp. 20–30, May 2015.

[26] H. Bay, A. Ess, T. Tuytelaars, and L. Van Gool, "Speeded-up robust features (SURF)," *Comput. Vis. Image Understand.*, vol. 110, no. 3, pp. 346–359, 2008.

[27] H. Lim and H. Park, "An efficient multi-view generation method from a single-view video based on affine geometry information," *IEEE Trans. Multimedia*, vol. 16, no. 3, pp. 726–737, Apr. 2014.

[28] Z. Lou and T. Gevers, "Image alignment by piecewise planar region matching," *IEEE Trans. Multimedia*, vol. 16, no. 7, pp. 2052–2561, Nov. 2014.

[29] X. Luo *et al.*, "Towards enhancing stacked extreme learning machine with sparse autoencoder by correntropy," *J. Franklin Inst.*, vol. 355, no. 4, pp. 1945–1966, Mar. 2018.

[30] X. Luo, J. Deng, J. Liu, W. Wang, X. Ban, and J.-H. Wang, "A quantized kernel least mean square scheme with entropy-guided learning for intelligent data analysis," *China Commun.*, vol. 14, no. 7, pp. 1–10, Jul. 2017.

[31] X. Luo, D. Zhang, L. T. Yang, J. Liu, X. Chang, and H. Ning, "A kernel machine-based secure data sensing and fusion scheme in wireless sensor networks for the cyber-physical systems," *Future Gener. Comput. Syst.*, vol. 61, pp. 85–96, Aug. 2016.

HAIJIANG ZHU received the Ph.D. degree in pattern recognition and intelligent system from the National Laboratory of Pattern Recognition, Institute of Automation, Chinese Academy of Sciences, Beijing, China, in 2004. From 2006 to 2007, he was a Visiting Scholar with the Faculty of Engineering, Iwate University, Japan. He is currently a Professor with the College of Information Science and Technology, Beijing University of Chemical Technology. He is interested in image processing and computer vision.

XIN WEN received the B.E. degree from the College of Information Science and Technology, Beijing University of Chemical Technology, Beijing, China, in 2014, where he is currently pursuing the M.S. degree with the Automation Department. He is interested in image processing and computer vision.

FAN ZHANG (S'07–M'10–SM'17) received the B.E. degree in communication engineering from the Civil Aviation University of China, Tianjin, China, in 2002, the M.S. degree in signal and information processing from Beihang University, Beijing, China, in 2005, and the Ph.D. degree in signal and information processing from the Institute of Electronics, Chinese Academy of Science, Beijing, China, in 2008.

He is currently a Professor of electronic and information engineering with the Beijing University of Chemical Technology, Beijing, China. His research interests are synthetic aperture radar signal processing, high performance computing and artificial intelligence. He has been a Reviewer for the IEEE Transactions on Geoscience and Remote Sensing, the IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, the IEEE Geoscience and Remote Sensing Letters, and the *International Journal of Antennas and Propagation*.

XUEJING WANG received the Ph.D. degree from the Changchun Institute of Optics, Fine Mechanics and Physics, Chinese Academy of Science, in 2002. She is currently an Assistant Professor with the Center for Information Technology, Beijing University of Chemical Technology. She is interested in image processing and pattern recognition.

GUANGHUI WANG (M'10–SM'17) received the Ph.D. degree in computer vision from the University of Waterloo, Waterloo, ON, Canada, in 2014. He is currently an Assistant Professor with the University of Kansas, Lawrence, KS, USA. He is also with the Institute of Automation, Chinese Academy of Sciences, China, as an Adjunct Professor. He has authored one book *Guide to Three Dimensional Structure and Motion Factorization*, (Springer-Verlag). He has authored or co-authored over 90 papers in peer-reviewed journals and conferences. His research interests include computer vision, structure from motion, object detection and tracking, artificial intelligence, and robot localization and navigation. He has served as an Associate Editor and on the editorial board of two journals, as an area chair or a TPC member of over 20 conferences, and as a reviewer of over 20 journals.

• • •