Research Article

Robust Homography Estimation Based on Nonlinear Least Squares Optimization

Wei Mou,[1] Han Wang,[1] and Gerald Seet[2]

[1] *School of Electrical & Electronics Engineering, Nanyang Technological University, Singapore*
[2] *School of Mechanical & Aerospace Engineering, Nanyang Technological University, Singapore*

Correspondence should be addressed to Wei Mou; mouwei@ntu.edu.sg

The homography between image pairs is normally estimated by minimizing a suitable cost function given 2D keypoint correspondences. The correspondences are typically established using descriptor distance of keypoints. However, the correspondences are often incorrect due to ambiguous descriptors which can introduce errors into following homography computing step. There have been numerous attempts to filter out these erroneous correspondences, but it is unlikely to always achieve perfect matching. To deal with this problem, we propose a nonlinear least squares optimization approach to compute homography such that false matches have no or little effect on computed homography. Unlike normal homography computation algorithms, our method formulates not only the keypoints' geometric relationship but also their descriptor similarity into cost function. Moreover, the cost function is parametrized in such a way that incorrect correspondences can be simultaneously identified while the homography is computed. Experiments show that the proposed approach can perform well even with the presence of a large number of outliers.

1. Introduction

Estimating homography given keypoint correspondences of image pairs has received much attention due to its extensive applications, for example, panoramas generation [1], motion estimation [2], camera calibration [3], and augmented reality [4].

The homography estimation given an image pair can be decomposed into two stages. In the first stage, keypoints in two images are detected and their local image descriptors are extracted. The keypoint matches are established by comparing the corresponding descriptors. After that, the false matches, due to ambiguous descriptors, are detected and removed using robust methods. In the second stage, a cost function based on remaining matches is defined and the homography is computed by minimizing the cost function.

Much attention has been paid to the first stage. The descriptor ambiguity and reliability can be improved by using more distinctive descriptors such as PCA-SIFT [5]. However, it cannot completely resolve ambiguities especially in the case of repetitive patterns or occlusions.

Hypothesize-and-verification framework such as random sample consensus (RANSAC) [6] is the most popular method to remove inevitable outliers. It randomly and repeatedly selects and verifies a small set of matches to find inliers that are consistent with some global geometric constraints.

More sophisticated variations have been proposed to improve the standard RANSAC algorithm. Some approaches such as PROSAC [7] and ARRSAC [8] focus on improving the reliability of hypothesis. To achieve this, these methods use image appearance to select keypoint correspondences with high confidence and, consequentially, speed up the search for consistent matches. Some other approaches such as MSAC [9], MLESAC [9], and MAPSAC [10] use more sophisticated measures such as likelihood or posterior instead of point consensus to better verify hypothesis. These approaches have succeeded in greatly reducing the error rate.

In order to estimate homography that can best describe the data, usually all inliers are then fed into the optimization process. Given point correspondences, a cost function is formulated based on the difference between the measured and estimated image coordinates. Homography is obtained

by minimizing this cost function using Direct Linear Transformation (DLT) or by iterative optimization methods such as Gauss-Newton or Levenberg-Marquardt. However, after the initial selection stage, it cannot be guaranteed that the data is noise-free before minimizing the cost function as there is no robust rule to define outliers. Conventional least squares optimization algorithms are in general not robust to outliers. The reason is that Gaussian distribution of noise is the basic assumption for solving least squares problem and Gaussian distribution is sensitive to outliers due to its narrow tailed nature. Error increases quadratically in least squares problem. Hence, the incorrect correspondences can easily lead to divergence of homography estimation. For example, a false point matching causes big geometric difference on the measured and estimated image coordinates. In order to compensate for this error, the homography has to deviate from its true value during the optimization process. Moreover, in some applications such as with ego motion estimation or circular panorama generation, such errors accumulate over time. Hence, even an extremely low error rate will eventually result in significant errors. As a result, given point correspondences, a robust cost function that can accurately calculate homography in the presence of outliers is needed. A desirable property for such an algorithm is that the incorrect correspondences have no or little contribution to the optimization process. In other words, during the homography calculation process, the algorithm should be able to identify the undetected mismatches as well.

Some robust cost functions such as Huber [11] have been proposed to reduce errors introduced by outliers. Unlike normal least square in which the error has a quadratic influence on the cost function, the Huber function makes cost increase linearly if the error exceeds a certain threshold. This means that it weighs large error less. However, the Huber cost function is not enough to deal with outliers, because the influence of outliers is reduced rather than removed.

Serradell et al.'s method [12] is able to solve correspondences and homography simultaneously. They combine geometric prior and appearance prior to achieve this. More specifically, the search space of the homography is constrained such as to limit the range of rotations and scales. Several homography hypotheses are sampled in this search space. Gaussian Mixture Models that best fit these samples are formed as the geometric prior. Appearance prior is based on descriptor similarity distances. The homography estimate and its covariance are iteratively updated by a Kalman filter that uses the best correspondences as measurements until the covariance becomes negligible. The potential mismatches are detected as those that are least likely to reduce the covariances of the Kalman filter. In such a way, the influence of mismatches is removed instead of reduced which makes it robust to high numbers of outliers. Their method can also be categorized into hypothesis-and-verify approach.

In this paper, we propose a robust homography estimation approach that is different from the hypothesis-and-verify approach. It allows the homography calculation and incorrect correspondences identification simultaneously within a Bayesian framework. The contributions of this paper are as follows.

(i) No initial samples or hypothesis of Homography are required.

(ii) We formulate a new cost function which integrates keypoint consensus and descriptor similarity.

(iii) We introduce a new set of parameters that represent the confidence for each correspondence, by solving which using nonlinear optimization, both of the homography and false keypoint correspondences can be determined simultaneously.

With these improvements, our method can achieve robustness even when a rather high percentage of correspondences are incorrect. The influence of outliers for homography computing is much reduced, even removed, and thus can output a more accurate homography that satisfies most pixels' geometric relationship between two images. Experimental results show the effectiveness and efficiency of the proposed approach in both synthetic images and real-life images.

2. Cost Function Formulation

We first give some notations and briefly review the conventional formulation of homography estimation before introducing our extension.

Given an image pair, $\mathbf{x}_i^j$ represents the ith point in jth image and its inhomogeneous image coordinate is denoted as (x_i^j, y_i^j). The data set of extracted keypoint correspondences is $\mathbf{M}$. Each pair of corresponding points $\mathbf{x}_k^1$ and $\mathbf{x}_i^2$ defines a single point $\mathbf{m}_i$ in a measurement space $\mathfrak{R}^4$, formed by joining the coordinates in each image. The ith keypoint correspondence of image pair is represented by vector $\mathbf{m}_i = (x_k^1, y_k^1, x_l^2, y_l^2)^T$. The homography that transforms the first image to the second image is $\mathbf{H}$. According to Bayesian framework, $\mathbf{H}$, is estimated as

$$\Pr(\mathbf{H} \mid \mathbf{M}) = \frac{\Pr(\mathbf{M} \mid \mathbf{H})\Pr(\mathbf{H})}{\Pr(\mathbf{M})}. \tag{1}$$

There is no prior knowledge about matchings and homography, we assume them to be uniformly distributed which makes $\Pr(\mathbf{H})$ and $\Pr(\mathbf{M})$ trivial to the optimization problem. Also, because feature matching pairs are determined using descriptor similarity only, we assume them to be conditionally independent from each other. Hence, (1) is changed into

$$\Pr(\mathbf{H} \mid \mathbf{M}) \propto \Pr(\mathbf{M} \mid \mathbf{H}) \propto \prod_{i=1\cdots n} \Pr(\mathbf{m}_i \mid \mathbf{H}), \tag{2}$$

where n is the number of correspondences. Let $f(\mathbf{x}_k^1, \mathbf{H})$ be the estimation of point $\mathbf{x}_k^1$ projected on image 2 from image 1 by $\mathbf{H}$. The geometric error vector between the estimated and measured image coordinates introduced by matching $\mathbf{m}_i$ is defined as

$$\mathbf{e}_i = \mathbf{x}_l^2 - f(\mathbf{x}_k^1, \mathbf{H}). \tag{3}$$

Without loss of generality, the noise in the two images is assumed to be Gaussian on each image coordinate with zero

mean and covariance matrix Σ. Equation (2) can be written as

$$
\begin{aligned}
\Pr\left(\mathbf{M} \mid \mathbf{H}\right) &= \prod_{i=1:n}\left(\frac{1}{\sqrt{2\pi\,|\Sigma|}}\right)\exp\left(-\frac{1}{2}\mathbf{e}_i^T\Sigma^{-1}\mathbf{e}_i\right) \\
&= \left(\frac{1}{\sqrt{2\pi\,|\Sigma|}}\right)^n\prod_{i=1:n}\exp\left(-\frac{1}{2}\mathbf{e}_i^T\Sigma^{-1}\mathbf{e}_i\right),
\end{aligned}
\tag{4}
$$

where n is the number of correspondences and $|\Sigma|$ is the determinant of covariance. Take negative log likelihood of all the correspondences; (4) becomes

$$
-\log\Pr\left(\mathbf{M} \mid \mathbf{H}\right) = \eta + \frac{1}{2}\sum_{i=1:n}\mathbf{e}_i^T\Sigma^{-1}\mathbf{e}_i,
\tag{5}
$$

where η is a constant number. In order to optimally estimate $\mathbf{H}$, from (1) to (5), a MAP estimate $\mathbf{H}^*$ is made such that

$$
\begin{aligned}
\mathbf{H}^* &= \arg\max_{\mathbf{H}}\Pr\left(\mathbf{H}\mid\mathbf{M}\right) \\
&= \arg\min_{\mathbf{H}} -\log\Pr\left(\mathbf{M}\mid\mathbf{H}\right) \\
&= \arg\min_{\mathbf{H}} \underbrace{\sum_{i=1:n}\mathbf{e}_i^T\Sigma^{-1}\mathbf{e}_i}_{\text{cost function } F(x)},
\end{aligned}
\tag{6}
$$

where minimizing cost function $F(x)$ is a nonlinear least squares problem and it can be solved using iterative methods such as Gauss-Newton or Levenberg-Marquardt. However, in this cost function, there is no mechanism to deal with outliers. Thus, false correspondences can easily cause the least squares optimization to converge to a wrong estimation.

It is desirable to reduce or ignore the influence of incorrect correspondences during optimization. Our main approach to achieve this is the following: as the optimization proceeds, we dynamically assign a corresponding reliability to each point correspondence. In other words, the contribution of each correspondence to the optimization problem changes while the optimization proceeds. This leads to an extended optimization problem: not only parameters of homography need to be optimized but also the reliability of feature correspondence. Referring to cost function, Σ in (6) represents the uncertainty of the measured image coordinates which is a good measure to control the portion of contribution of the corresponding point correspondence. For example, high uncertainty means the noise in the measured image coordinate is high and thus the error introduced by the corresponding keypoint match will contribute little to the optimization problem. However, in (6) we can observe that Σ is constant for all correspondences. In order to make it change dynamically, we introduce a new set of parameters $\mathbf{C}$ to be optimized which describes the covariance matrix Σ_i for each measurement.

Moreover, the similarity of corresponding keypoints from two views is also included into cost function. The collection of similarity score is denoted as $\mathbf{S}$ and $\mathbf{s}_i$ is the similarity score of correspondence $\mathbf{m}_i$. The descriptor similarity is integrated into optimization problem in such a way that it helps to

determine the values in $\mathbf{C}$ which describe the confidences of correspondences and $\mathbf{c}_i$ represents the confidence for match $\mathbf{m}_i$. With the idea of our approach being introduced, (1) can be extended as

$$
\begin{aligned}
\Pr\left(\mathbf{H}\mid\mathbf{M},\mathbf{C},\mathbf{S}\right) &\propto \Pr\left(\mathbf{M},\mathbf{C}\mid\mathbf{H},\mathbf{S}\right) \\
&\propto \Pr\left(\mathbf{M}\mid\mathbf{C},\mathbf{H}\right)\Pr\left(\mathbf{C}\mid\mathbf{S}\right) \\
&\propto \prod_{i=1\cdots n}\underbrace{\Pr\left(\mathbf{m}_i\mid\mathbf{c}_i,\mathbf{H}\right)}_{\text{likelihood}}\underbrace{\Pr\left(\mathbf{c}_i\mid\mathbf{s}_i\right)}_{\text{prior}}.
\end{aligned}
\tag{7}
$$

The first term is the likelihood of matching given the homography and uncertainty of measurements, while the second term is the prior probability of $\mathbf{c}_i$ and the uncertainty of which is based on $\mathbf{s}_i$.

With Gaussian assumption, new cost function to be minimized can be obtained by taking the negative log likelihood of (7):

$$
F'(x) = \sum_{i=1:n}\left(\underbrace{\mathbf{e}_i^T\Sigma_i^{-1}\mathbf{e}_i}_{\delta_i} + \underbrace{\boldsymbol{\varepsilon}_i^T\Psi_i^{-1}\boldsymbol{\varepsilon}_i}_{\rho_i}\right),
\tag{8}
$$

with Ψ_i corresponding to covariance of prior distribution of $\mathbf{c}_i$. $\boldsymbol{\varepsilon}_i$ is the difference between the current $\mathbf{c}_i$ and its expected value $\bar{\mathbf{c}}_i$. We set $\bar{\mathbf{c}}_i$ to 1 for each correspondence.

Next, we elaborate on how does $\mathbf{c}_i$ describe Σ_i. To simplify the problem, we assume that the error distribution of measurements along x and y directions are independent from each other and their standard deviations are the same. Thus, $\mathbf{c}_i$ is defined as a scalar and it is enough to describe Σ_i. As a result, δ_i in (8) becomes

$$
\delta_i = \mathbf{e}_i^T\begin{bmatrix}\mathbf{c}_i^{-2} & 0 \\ 0 & \mathbf{c}_i^{-2}\end{bmatrix}^{-1}\mathbf{e}_i = \left(\mathbf{c}_i\cdot\mathbf{e}_i\right)^T\mathbf{I}_{2\times2}\left(\mathbf{c}_i\cdot\mathbf{e}_i\right),
\tag{9}
$$

where $\mathbf{c}_i$ is defined as the inverse of the standard deviation of Σ_i. However, in (9) we can see that it can also be interpreted as the weight of $\mathbf{e}_i$ which is the geometric error introduced by correspondence m_i. If m_i is an outlier, the corresponding $\mathbf{e}_i$ will contribute little to optimization problem by assigning a small weight to it and will be ignored if $\mathbf{c}_i = 0$, because in this case, $\mathbf{e}_i$ will not be added to the global error term. In this interpretation, if we scale $\mathbf{c}_i$ to $[0, 1]$, $\mathbf{c}_i$ not only makes optimization process robust to outliers but also works as outlier and inlier selector as false $\mathbf{m}_i$ will make $\mathbf{c}_i$ close to 0, while correct $\mathbf{m}_i$ makes it close to 1.

Similarly as δ_i, we can write ρ_i in (8) as

$$
\rho_i = \left(\mathbf{s}_i\cdot\boldsymbol{\varepsilon}_i\right)^T\mathbf{I}_{2\times2}\left(\mathbf{s}_i\cdot\boldsymbol{\varepsilon}_i\right).
\tag{10}
$$

The local feature descriptor pair of detected keypoints in $\mathbf{m}_i$ is $(\boldsymbol{\varphi}_k^1, \boldsymbol{\varphi}_l^2)$ in which $\boldsymbol{\varphi}_k^j$ is the descriptor vector of kth keypoint in image j. Let $\text{dist}(\boldsymbol{\varphi}_k^1, \boldsymbol{\varphi}_l^2)$ represent the Euclidean distance between descriptors $\boldsymbol{\varphi}_k^1$ and $\boldsymbol{\varphi}_l^2$. $\mathbf{s}_i$ is defined as

$$
\mathbf{s}_i = \frac{\lambda}{1 + \text{dist}\left(\boldsymbol{\varphi}_k^1, \boldsymbol{\varphi}_l^2\right)}\cdot\frac{\text{distNN}^{\text{2nd}}_{\boldsymbol{\varphi}_k^1}}{\text{distNN}^{\text{1st}}_{\boldsymbol{\varphi}_k^1}},
\tag{11}
$$

where distNN^{1st} and distNN^{2nd} denote Euclidean distance between φ_k^1 and its first and second nearest neighbour in descriptor space. λ is a scale factor that we set to 0.03 manually. The ratio between first and second nearest neighbour can represent the distinctiveness of detected keypoints and thus is used to reduce the errors that introduced repeated patterns.

Similar as c_i, high s_i means more contribution of ε_i to optimization problem as is shown in (10). The higher s_i is, the more difficult for ε_i to change which means the less likely that c_i deviates from its expected value $\bar{c}_i$ which is 1. In other words, the higher the s_i is, the more likely that m_i is classified as an inlier.

The reason why we choose to work with descriptor similarity instead of appearance difference is twofold: first, the descriptors and distances to nearest neighbours have been computed before the optimization and thus saving computational time. Second, in this way, S will stay constant during optimization process, no matter how H and C change, because descriptors have been extracted beforehand. Hence, no more new parameters are needed to be optimized.

After introducing all variables of our cost function as in (8), we can have a better intuitive understanding about it. δ_i is responsible for minimizing the geometric error of homography transformation. e_i in it represents how inconsistent the correspondence m_i with estimated homography H. In order to minimize it, high e_i which normally means outliers, will push c_i towards zero. Although it has the nice property of ignoring the errors introduced by outliers, even for low e_i which is nonzero, nothing prevents c_i reducing its value. This means that effects of all m_i are reduced even ignored during optimization which will result in wrong estimation of homography. Fortunately, ρ_i helps to avoid this behaviour. It serves as a penalty whenever c_i decreases. This is achieved by setting all initial value of c_i and its expected value $\bar{c}_i$ to 1. As a result, decreasing c_i corresponds to increasing prior error ε_i. Moreover, we introduce s_i to reflect similarity and distinctiveness of detected keypoints in m_i and the value of it determines the amount of penalty added to the cost function.

The least squares optimization can be further robustified. Due to Gaussian assumption, the error vectors e_i and ε_i in (8) have quadratic influence on cost function $F'(x)$. Even a single outlier would have major negative effect. In order to be a more outlier robust, the Huber cost function is used and makes cost increase linearly if error exceeds certain threshold; thus (8) becomes

$$F'(x) = \begin{cases} F'(x), & \text{if } \sqrt{F'(x)} < b \\ 2b\sqrt{F'(x)} - b^2, & \text{otherwise.} \end{cases} \quad (12)$$

There are even more robust cost functions such as Blake-Zisserman, corrupted Gaussian, or Cauchy. Compared to them, the reason why Huber function is preferred is that it does not introduce new local minima due to its convexity [13].

For a given optimization problem, there are often several ways to parametrize it. For each c_i, one parameter is enough to parametrize it since it is a weight factor for the contribution of δ_i to optimization problem. Although H has only 8 DOF, we parametrize it using 9 parameters. As discussed in [13], it is not necessary or advisable to use 8-parametrization

by removing scale factor because when using the minimal parameterization, it is more likely for the optimization process stuck in the local minimum. Also, in our problem, the number of parameters in C is usually much more than 9. A minimal parameterization of H has little help on efficiency of our approach. Hence, the total number of parameters need to be solved during optimization is 9+n where n is the number of keypoint correspondences.

3. Experiments and Evaluations

In this section we evaluate our approach using both synthetic and real-life pictures. Synthetics images are generated by transforming the original image I_s to target image I_t through a homography H generated randomly. We apply FAST [14] keypoint detector on I_s and I_t. BRIEF [15] descriptors are extracted from detected keypoints. Keypoint correspondences M are determined simply by finding the nearest neighbours in descriptor space.

3.1. Synthetic Noise. To evaluate the performance of our approach, besides the original outliers, more randomly generated false correspondences are added to M. The homography estimation results of our approach are shown in Table 1. The accuracy of estimated homography is evaluated using root mean-squared error (RMSE). From the results we can see that even the number of outliers is over ten times more than inliers; our homography estimation still remains feasible compared to the approach of minimizing the conventional cost function (RMSE w/o C).

Our purpose to introduce S into cost function is to integrate descriptor similarity into optimization process and penalize the behaviour of decreasing C. The lower the s_i is, the easier it is to decrease c_i and thus the easier the m_i is classified as an outlier. The column "RMSE w/o S" in Table 1 shows that without S, the algorithm still has some robustness against outliers even when the error rate is higher than 0.5; however, the RMSE increases significantly especially when the error rate goes higher.

When the optimization process finishes, a keypoint correspondence is classified as outlier if its c_i is below 0.5, since we initially set c_i to 1 and its value tends to move towards 0. False inliers deviate homography estimation more severely than false outliers because using a wrong correspondence during optimization process is much worse than not using a correct correspondence. The proposed approach can keep the number of false inliers very low compared to the total number of correspondences. Because many outliers are generated randomly, some of them are accidentally close to the truth as shown in Figure 1. Such false matchings are acceptable to our algorithm. Because the error it introduced to the whole optimization problem is considerably small and any effort to reject such outliers would introduce new parameters to be tuned and thus complicate the problem.

It is worth noticing that more outliers do not necessarily mean more error on homography estimation. For example, in Table 1, the RMSE of the test with 309 outliers is lower than many of those with fewer outliers. The reason for this

TABLE 1: Homography estimation results on random generated false matches. TP and FP mean true positive and false positive number, while TN and FN represent true negative and false negative number. The column "RMSE w/o $\mathbf{S}$" represents the RMSE of approach without using descriptor similarity $\mathbf{S}$, while "RMSE w/o $\mathbf{C}$" represents the RMSE of standard optimization approach whose cost function is introduced in (6).

True matches	False matches	TP	FP	TN	FN	RMSE	RMSE w/o $\mathbf{S}$	RMSE w/o $\mathbf{C}$
42	0	42	0	0	0	0.166	0.165	0.184
42	51	42	1	50	0	0.573	0.462	13.992
42	103	42	1	102	0	0.597	0.699	21.420
42	154	42	1	153	0	0.222	0.608	27.589
42	206	42	1	205	0	0.633	5.816	30.845
42	257	42	1	256	0	0.622	17.311	34.856
42	309	42	1	308	0	0.228	28.479	37.672
42	360	41	1	359	1	0.652	36.286	38.408
42	412	42	1	411	0	0.594	37.278	40.348
42	463	41	1	462	1	0.821	39.889	41.875
42	515	42	1	514	0	0.825	38.737	42.467
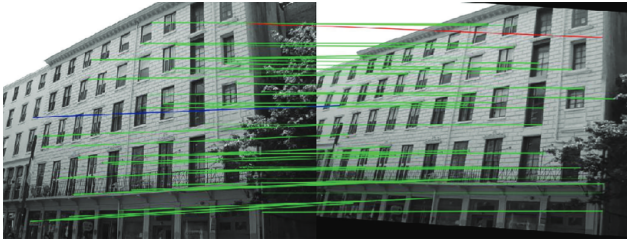
FIGURE 1: Keypoint correspondences between original and transformed images with 567 outliers. The green lines indicate correctly detected inliers, while blue means false negative and red means false positive detections. The correctly detected outliers are ignored for the sake of clearance of the figure.

is that almost all of the $\mathbf{c}_i$ lies in the area that is either very close to 1 or very close to 0 as seen in Figure 2(b). This means the effects of most outliers are removed instead of reduced, while, in the case of 206 outliers as shown in Figure 2(a), $\mathbf{c}_i$ of some correspondences lies close to 0.5 which means that the effect of some inliers is not enforced properly. Hence, even tests with fewer total outliers may still have higher RMSE.

The best case is to assign $\mathbf{c}_i$ of detected outliers to 0. However, this cannot be done naturally in our algorithm. In order to minimize (8), both $\mathbf{H}$ and $\mathbf{C}$ need to be optimized. As their Jacobians are needed to solve the optimization problem using an iterative method, the values of $\mathbf{C}$ during the optimization process have to be continuous instead of discrete. Hence, this inaccuracy is inevitable by just solving cost function. However, it is very easy to refine the homography because our algorithm naturally assigns each matching a confidence $\mathbf{c}_i$ which is based on both the geometric error and descriptor similarity. We simply select the best m inliers to refine the homography using DLT; then the inaccuracy caused by continuous $\mathbf{C}$ can be removed.

We compare our approach with RANSAC and LMedS. Homography is estimated using the same data set as in Table 1 and the result is shown in Figure 3(a).

As expected, the RMSE of LMedS rise when error rate is over 50%. With our robustified problem formulation, the RMSE of our approach keeps low and almost constant with the increasing number of outliers. Our approach shows similar robustness compared to RANSAC and the same robustness can be easily achieved by using the refinement described above. We use optimized OpenCV implementations for both RANSAC and LMedS. We repeated tests 5 times for 11 noise levels and the run time is averaged for each level. The result is shown in Figure 3(b). LMedS approach keeps low and constant runtime at around 4 ms. The runtime for our approach started at 3.3 ms with 42 inliers and 0 outliers ended with 46.08 ms with 567 outliers. The RANSAC approach started at 0.81 ms and ended with 18249 ms. The runtime of RANSAC increases much faster than our approach because the required number of sampling times increases significantly with the portion of outliers [13].

The problem size of our approach increases with the number of keypoint correspondences. Besides 9-parameter homography, for each new correspondence $\mathbf{m}_i$, a new parameter $\mathbf{c}_i$ that needs to be optimized is added to the problem formulation. Jacobian $\mathbf{J}$ and Hessian $(\mathbf{J}^T\mathbf{J})$ are required to minimize the cost function and their size also increases with the number of parameters. When solving $n \times n$ equations, it takes $O(n^3)$ for a dense system [16]. Thus, the size of the problem is not the bottle neck as solving it is much more crucial.

By exploiting the sparseness of the optimization problem, computational benefits can be gained through avoiding storing and operating on zero elements [17]. Hence, the sparser the structure, the more computational benefits we can gain.

Fortunately, when solving our minimization problems, the Jacobian and Hessian matrix in our problem have a sparse block structure as shown in Figure 4. Only the black entries are nonzero. We can divide the Jacobian matrix into 4 zones. Top-left zone is the Jacobian for geometric error $\mathbf{e}$ with respect to homography $\mathbf{H}$. The geometric error for each correspondence $\mathbf{e}$ is a $2d$ vector and the number of
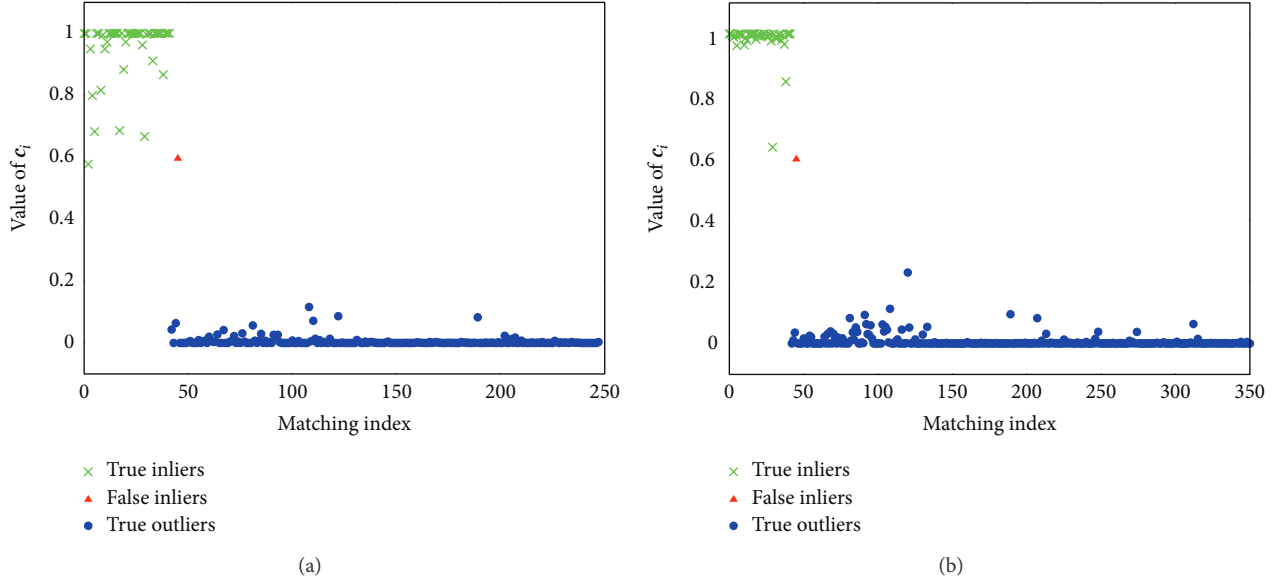
(a)

(b)

FIGURE 2: Distribution of **C** which represents the contribution of each matching to the optimization problem. Also, it is an indicator of outliers as we classify $\mathbf{m}_i$ as outlier if $\mathbf{c}_i < 0.5$. Without loss of generality, inliers are put in the first 42 indexes. (a) Test with 206 outliers in which 1 false inlier appeared. (b) Test with 309 outliers in which 1 false inlier appeared.
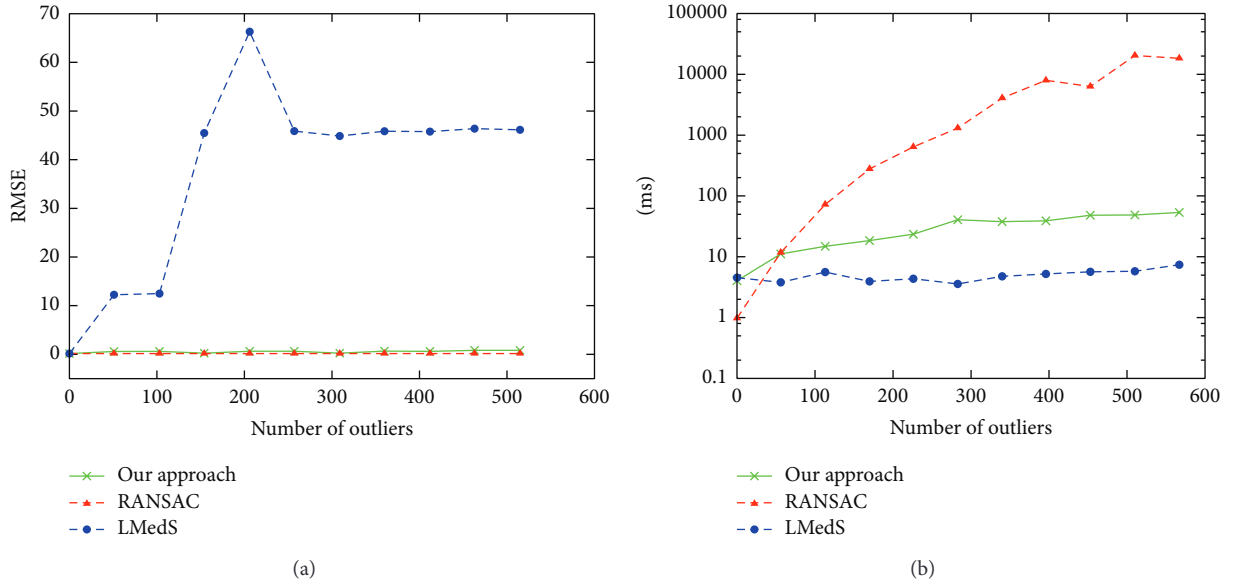

(a)

(b)

FIGURE 3: Performance comparison with RANSAC and LMedS. (a) Root mean squared error comparison and (b) runtime comparison.

parameters of homography is 9. The top-right zone is the partial derivatives of **e** with respect to parameters C whose size equals the number of correspondences. Because $\mathbf{c}_i$ only influences $\mathbf{e}_i$, the only nonzero entries in this area are $\partial\mathbf{e}_i/\partial\mathbf{c}_i$. The entries in bottom-left zone are partial derivatives of $\boldsymbol{\varepsilon}$ with respect to **H** and they are all zero. Because we only use descriptor similarity whose value is independent on homography estimation to prevent that **H** is involved into calculation of $\boldsymbol{\varepsilon}$. The bottom-right zone is diagonal and the only nonzero entries are the partial derivatives $\partial\boldsymbol{\varepsilon}_i/\partial\mathbf{c}_i$.

It is easy to calculate that the density of Jacobian matrix is $5/(9 + n)$, while for Hessian it is $(63 + 19n)/(9 + n)^2$. Hence the more the number of correspondences is, the sparser the matrix would be which means more computational benefits can be obtained as we only operate on nonzero blocks.

The convergence property of proposed approach is depicted in Figure 5. For homography estimation, compared to the standard cost function which reports geometric error only, our approach introduces additional parameters contributing to the overall error. As can be seen from the figure,
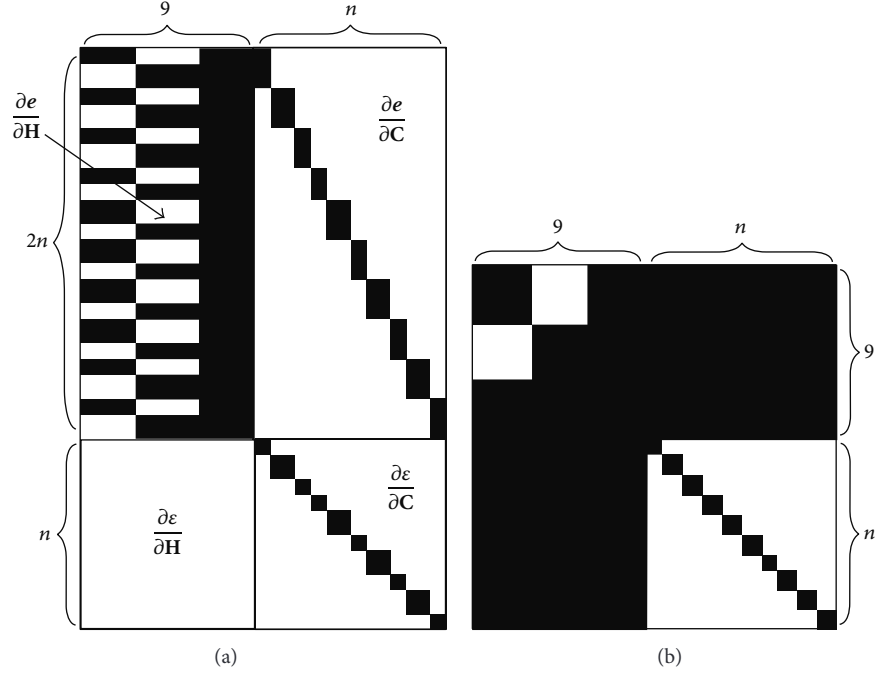
FIGURE 4: Sparse structure example of Jacobian $\mathbf{J}$ (a) and Hessian $\mathbf{J}^T\mathbf{J}$ (b) of the proposed optimization problem. The black entries are nonzero. n is the number of keypoint correspondences. In this specific example, n equals 10.
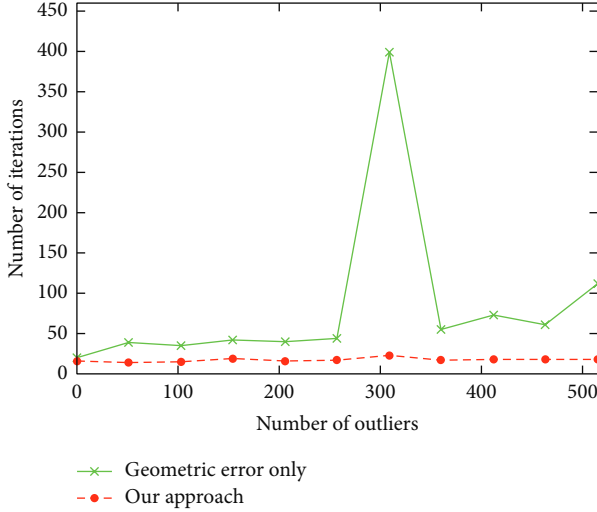

FIGURE 5: Number of iterations needed before optimization process converges. For our approach, the number of iterations is between 14 and 23.

our approach needs about 15 iterations to converge, while the standard cost function for homography estimation needs significantly more iterations to be optimized. This indicates that compared to the standard cost function, our cost function has fewer local minimals and the newly introduced parameters offer faster convergence during optimization process.

3.2. Experiments on Real-Life Images. Compared to synthetic images, homography estimation is more complex for real-life images due to illumination variation or camera distortions and so on. The robustness and efficiency of our approach are tested using real-life images and the results are shown in Table 2. The input and reference image pairs are under different viewing conditions. To evaluate the homography estimation, we transform the input image through the estimated homography and the differences between reference images and transformed images are shown in the third and fourth columns in Table 2. We can see that the result is similar as the one in the synthesized experiment. Our approach can achieve similar robustness as RANSAC while saving much in computational time.

4. Conclusion

In this paper, we propose a novel homography estimation and outlier detection approach which is essentially different from conventional hypothesis-and-verify approaches such as RANSAC. We formulate the homography estimation and outlier detection problem together into a single nonlinear least squares problem. The new cost function combines both geometric error and descriptor similarity, by minimizing which of the homography and outliers can be determined simultaneously.

Experiment results demonstrate that our approach achieves similar robustness as RANSAC under different viewing conditions. Due to the sparse structure of Jacobian and

TABLE 2: Experiment results on real pictures. The first pair of images is under different blur levels and the second pair with different orientations and zoom. The third pair is for different view angles and repeated patterns.

Input	Reference	Our approach	RANSAC	Time (ours)	Time (RANSAC)
				16.79 ms	190.94 ms
				13.08 ms	4525.1 ms
				18 ms	57 ms

Hessian of proposed cost function, our algorithm remains efficient even with the presence of large amount of outliers.

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

References

[1] M. Brown and D. G. Lowe, "Automatic panoramic image stitching using invariant features," *International Journal of Computer Vision*, vol. 74, no. 1, pp. 59–73, 2007.

[2] D. Nister, "Preemptive ransac for live structure and motion estimation," in *Proceedings of the 9th IEEE International Conference on Computer Vision*, pp. 199–206, 2003.

[3] Z. Zhang, "A flexible new technique for camera calibration," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, no. 11, pp. 1330–1334, 2000.

[4] D. Wagner, G. Reitmayr, A. Mulloni, T. Drummond, and D. Schmalstieg, "Pose tracking from natural features on mobile phones," in *Proceedings of the 7th IEEE/ACM International Symposium on Mixed and Augmented Reality (ISMAR '08)*, pp. 125–134, 2008.

[5] Y. Ke and R. Sukthankar, "Pca-sift: a more distinctive representation for local image descriptors," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR '04)*, pp. II-506–II-513, 2004.

[6] M. Fischler and R. Bolles, "Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography," *Communications of the ACM*, vol. 24, no. 6, pp. 381–395, 1981.

[7] O. Chum and J. Matas, "Matching with prosac—progressive sample consensus," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR '05)*, pp. 220–226, 2005.

[8] R. Raguram, J. M. Frahm, and M. Pollefeys, "A comparative analysis of ransac techniques leading to adaptive real-time random sample consensus," in *European Conference on Computer Vision*, vol. 5303, pp. 500–513, 2008.

[9] P. Torr and A. Zisserman, "Mlesac: a new robust estimator with application to estimating image geometry," *Computer Vision and Image Understanding*, vol. 78, no. 1, pp. 138–156, 2000.

[10] P. Torr, "Bayesian model estimation and selection for epipolar geometry and generic manifold fitting," *International Journal of Computer Vision*, vol. 50, no. 1, pp. 35–61, 2002.

[11] P. J. Huber, *Robust Statistics*, John Wiley & Sons, 1981.

[12] E. Serradell, M. Özuysal, V. Lepetit, P. Fua, and F. Moreno-Noguer, "Combining geometric and appearance priors for robust homography estimation," in *European Conference on Computer Vision*, vol. 6313, pp. 58–72, 2010.

[13] R. Hartley and A. Zisserman, *Multiple View Geometry in Computer Vision*, Cambridge University Press, 2nd edition, 2004.

[14] E. Rosten and T. Drummond, "Machine learning for high-speed corner detection," in *European Conference on Computer Vision*, vol. 3951, pp. 430–443, 2006.

[15] M. Calonder, V. Lepetit, M. Ozuysal, T. Trzcinski, C. Strecha, and P. Fua, "BRIEF: computing a local binary descriptor very fast," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 34, pp. 1281–1298, 2012.

[16] B. Triggs, P. Mclauchlan, R. Hartley, and A. Fitzgibbon, "Bundle adjustment—a modern synthesis," in *Vision Algorithms: Theory and Practice*, Lecture Notes in Computer Science, pp. 298–375, Springer, 2000.

[17] M. I. A. Lourakis and A. A. Argyros, "Sba: a software package for generic sparse bundle adjustment," *ACM Transactions on Mathematical Software*, vol. 36, pp. 1–30, 2009.