A Robust Image Matching Method based on Optimized BaySAC

Zhizhong Kang, Fengman Jia, and Liqiang Zhang

Abstract
This paper proposes a robust image-matching method, which integrates SIFT with the optimized Bayes SAMpling Consensus (BaySAC). As the point correspondences are likely contaminated by outliers, we present a novel robust estimation method involving an efficient BaySAC for eliminating falsely accepted correspondences. The key points of the proposed hypothesis testing algorithm are determining and updating the prior probabilities of pseudo-correspondences. First, we propose a strategy for prior probability determination in terms of the statistical characteristics of a deterministic mathematical model for hypothesis testing. Moreover, the inlier probability updating is simplified based on a memorable form of Bayes’ Theorem. The proposed approach is validated on a variety of image pairs. The results indicate that when compared with the performance of RANdom SAMpling Consensus (RANSAC) and the original BaySAC, the proposed optimized BaySAC consumes less computation and obtains higher matching accuracy when the hypothesis set is contaminated with more outliers.

Introduction
Image-matching refers to the process of identifying the point correspondences between two or multiple images in the same scene or project. This process is an important topic in the digital photogrammetry and computer vision fields, including image registration (e.g., Han et al., 2012; Wu et al., 2012; Wang et al., 2013), object recognition (Yang and Cohen, 1999; Wakahara et al., 2001; Lin et al., 2012; Wang et al., 2014), object tracking (e.g., Donoser and Bischof, 2006; Goedeme et al., 2004), three-dimensional visualization (Zhang et al., 2012) and camera calibration. As external environmental noise introduces many challenges to image-matching, it is necessary to find a robust, accurate and fast image-matching method. The relation between two or more images of the same scene taken from different views can be described using the epipolar geometry. The fundamental matrix (FM) is the representation of the epipolar geometry in computer vision. Calculating the fundamental matrix with high accuracy is a necessary step in computer vision applications.

The fundamental matrix can be computed from the pixel coordinates of the corresponding points in the uncalibrated images. To approximate the fundamental matrix, the first step is to select corresponding points, which is also the key part of image-matching. Generally, the image matching methods can be divided into two categories: intensity-based and feature-based methods. The intensity-based matching methods can provide sub-pixel or better accuracy; however, these methods are sensitive to intensity change and geometry deformations. Therefore, stable and automatic feature-based algorithms are widely used in image-matching applications. There are many famous feature-based methods that have been successfully applied, including Harris, SIFT, Principal Components Analysis SIFT (PCA-SIFT), and Speeded Up Robust Features (SURF). The Harris algorithm (Harris and Stephens, 1988) proposed in 1988 is based on the eigenvalues of the second moment matrix. However, Harris corners are not scale-invariant. The SIFT algorithm (Lowe, 1999) was first introduced by David Lowe, and its remarkable advantages are scale-invariant and high speed. Ke and Sukthankar (2004) proposed an improved SIFT algorithm called PCA-SIFT which applied Principal Components Analysis (PCA) to the normalized gradient patch instead of using SIFT’s smoothed weighted histograms. The paper revealed that PCA-SIFT could produce increased accuracy and faster matching. SURF was proposed by Bay et al. (2006). In the context of scale space extrema detection, SURF filters the integral images using a box filter as an approximation of second-order Gaussian partial derivatives. SURF was demonstrated to approximate or even outperform previously proposed schemes with respect to repeatability, distinctiveness, and robustness, yet it could be computed much faster. Juan and Gwun (2009) presented a comparative study of SIFT, SURF and PCA-SIFT. Their study revealed that except for its computation efficiency, SIFT had better performance in the situation of rotation, scale, blur changes, and affine transformations than SURF and PCA-SIFT. Researchers have proposed many other distribution-based descriptors for image-matching using histograms to represent different features (Mikolajczyk and Schmid, 2005; Morel and Yu, 2009; Guo et al., 2010). Making use of the robust color invariance against varying imaging conditions, some algorithms with color-based descriptors have been presented (Diplaros et al., 2006; Abdel-Hakim and Farag, 2006; Stokman and Gevers, 2007; Verma et al., 2011). Some filter-based descriptors were proposed in the methods of Moreno et al. (2009) and Gómez and Romero (2011). Geometric constraints, such as distance invariance (Kang et al., 2009), distribution pattern (Dí et al., 2011) and triangulation (Wu et al., 2011), have also been applied in SIFT-based image-matching. After correspondences are selected, many proposed methods can be used for estimating the fundamental matrix. These methods are roughly divided into three classes: linear methods, iterative methods, and robust methods. Linear methods can largely reduce the computing time but are

Zhizhong Kang and Fengman Jia are with the Department of Remote Sensing and Geo-Information Engineering, School of Land Science and Technology, China University of Geosciences, Xueyuan Road 29, Haidian District, Beijing 100083, China (zzkang@cugb.edu.cn).

Liqiang Zhang is with the Research Center for Remote Sensing and GIS, School of Geography, Beijing Normal University, 100875 Beijing, China.
sensitive to error introduced during correspondence selection. The eight-point algorithm (Hartley, 1997) is the most widely used linear method. Iterative methods estimate the optimized fundamental matrix by minimizing a certain physical quantity (such as the Euclidean distance between a point and its corresponding epipolar line) during iterations (Brandt and Heikkonen, 2000). Iterative methods have higher accuracy, but more time-consuming. Robust methods can effectively remove outliers by weighting for residuals of corresponding points, e.g., M-Estimators (Huber, 2009). A well-regarded technique for robust estimation is the ransac statistical framework (Torr and Murray, 1997). As has been proven to be capable of dealing with more than 50 percent of outliers (Torr and Murray, 1997; Matas and Chum, 2004), RANSAC has been increasingly utilized in image-matching, and some other improved algorithms based on RANSAC have been proposed (e.g., Bhattacharya and Gavrila, 2012). Torr and Zisserman (2000) applied the robust estimation method of the Randomized Maximum Likelihood Estimation $S$ample Consensus (RANSAC) to derive best-fitting roof models in a model-driven manner. Torr and Davidson (2003) presented Synthesis of Importance Sampling and Random Sample Consensus (IMPSAC), which uses a hierarchical resampling algorithm. Randomized RANSAC (R-RANSAC) (Matas and Chum, 2004) increases speed of the model parameter estimation using randomization of the hypothesis evaluation step, as a large number of erroneous model parameters obtained from contaminated samples are evaluated in RANSAC. An optimized randomized RANSAC (Chum, 2008) has been proposed. It is based on the estimation of two probabilities characterizing the problem and critically influencing the design of the optimal strategy. Cheng and Lai (2009) described a consensus sampling technique for fast and robust model fitting involving improving the consensus sampling technique, model evaluation, and robust standard deviation estimation. Papalazarou et al. (2013) proposed a method for complex curves estimation, which is built on Sparse-Plus-Dense RANSAC, to select multiple model instances in two-dimensional and three-dimensional data.

Such original and improved RANSAC methods have been employed in several computer vision applications, such as fundamental matrix estimation and primitive fitting. However, these algorithms assume all data points have the same prior probabilities. Therefore, the hypothesis set is chosen randomly in each time, which likely leads to more iterations when the hypothesis set is contaminated with more outliers. To reduce the computational cost, in this paper we introduce a new robust method of BaySAC estimation to derive best-fitting roof models in a model-driven manner. The BaySAC algorithm by presenting a strategy of statistical testing of candidate fundamental matrices to compute the prior probability of each pseudo-correspondence. After the hypothesis testing, we update the inlier probability based on a memorable form of Bayes’ Theorem, which describes the likelihood between the prior and posterior probabilities of a pseudo-correspondence by considering the probability of the hypothesis set that the pseudo-correspondence belongs to, is fully correct.

The main contribution of this paper is detailed in the next Section, which introduces the optimized BaySAC method for robust correct match selection, followed by a discussion of our experimental results. Last, we offer our conclusions and suggestions for further research.

**Robust Image Matching Method based on the Optimized BaySAC**

The robust matching strategy employed in this paper uses a “SIFT + normalized eight-point algorithm + robust estimation” overall in the scheme. SIFT is the way to detect correspondences in image pairs and the normalized eight-point algorithm is used to calculate the fundamental matrix. We focus on the robust estimation algorithm for the purpose of eliminating falsely accepted correspondences.

Our robust matching strategy starts with the correspondences selection process. The most widely used SIFT algorithm is the method we choose in this step.

### Introduction of the SIFT Algorithm

Our optimized algorithm focuses on the purpose of eliminating falsely accepted correspondences, so the highly robust SIFT algorithm is chosen to detect correspondences. The SIFT algorithm is implemented as the following steps to generate a set of distinctive invariant features (points):

1. Scale-space extrema detection: employing a difference-of-Gaussian function to identify potential points of interest that are invariant to scale and orientation.
2. Key-point localization: fitting an analytical model (mostly in the form of a parabola) at each candidate location to determine the location and scale.
3. Orientation assignment: assigning one or more orientations to each key-point location based on local image gradient directions.
4. Key-point descriptor: measuring the local image gradients at the selected scale in the region around each key-point, and
5. Correspondences determination: a match is identified if the ratio of the distance from the closest neighbor of a key-point to that of the second closest is less than a predefined threshold. The k-dimensional tree (k-d tree) (Bentley, 1975) is used to speed up the search.

Although the SIFT is a widely-accepted robust method, in practice, it is impossible to correctly match all 3D correspondences with the above mentioned strategy. Most likely, some incorrect corresponding points are also inadvertently accepted. Therefore, a robust estimation method is used to detect and remove these false correspondences.

To implement a robust estimation process, we first need to determine the mathematical model that describes the raw data. The fundamental matrix is employed as it is a commonly used model to express the epipolar geometry between a stereo pair.

### The Hypothesis Model - Fundamental Matrix

The fundamental matrix is a basic tool in the analysis of scenes taken with two uncalibrated cameras. In this paper, we compute the fundamental matrix using the normalized eight-point algorithm (Hartley, 1997).

The eight-point algorithm is an algorithm used in computer vision to estimate the fundamental matrix related to a stereo camera pair from a set of corresponding image points. It calculates the fundamental matrix from eight or more points and it has the advantage of simplicity of implementation.

The fundamental matrix $(F)$ is defined by Equation 1:

\[
 u'Fu = 0 \quad (1)
\]

where $u = (u \ v \ 1)$, and $u' = (u' \ v' \ 1)$ represent a pair of corresponding points, and $(f_{11}, f_{12}, f_{13}, f_{21}, f_{22}, f_{23}, f_{31}, f_{32}, f_{33})$ is the vector of the fundamental matrix $F$. The coefficients of Equation 1 can be easily rewritten in terms of the known $u$ and $u'$ as follows:

\[
 uu'f_{11} + uu'f_{12} + uf_{13} + vv'f_{21} + vv'f_{22} + vf_{23} + uf_{31} + v'f_{32} + f_{33} = 0 \quad (2)
\]
The coefficients can be represented as a vector \((uu', uv', \ u, \ vv', \ v', 1)\), and then we can obtain a set of linear equations as follows:

\[
Af = 0 \quad (3)
\]

where \(A\) is the coefficient matrix, and \(f\) is the vector form of the fundamental matrix.

To avoid trivial solutions, it is possible to find a solution of Equation 3 with at least eight points matches under the constraint if \(f = 0\). However, the eight-point algorithm is extremely susceptible to noise, and the poor performance of the eight-point algorithm can most likely be traced to the bad condition number of the coefficient matrix when it is calculated by original image coordinates (Hartley, 1997). Thus, Hartley proposed a normalized eight-point algorithm, which involves as following: (a) Translate the points so that their centroid is at the origin; and (b) Scale the points isotropically so that the average distance from the origin is equal to \(\sqrt{2}\).

This process precedes the algorithm with a translation and scaling normalization of the coordinates of the matched points, which can improve the quality of the coefficient matrix and reduce the influence of noise.

Efficient BaySAC Match Selection

The original RANSAC-based algorithm tries many hypothesis sets contaminated by outliers because of the assumption of constant prior probabilities of data points, which likely leads to more iterations and expensive computation cost when the hypothesis set is contaminated with many outliers.

To improve the computational efficiency, this paper proposes a conditional sampling method based on BaySAC, which always selects the minimum number of data required with the highest inlier probabilities as a hypothesis set and thus reduces the number of iterations needed to find a good model. As the two key innovations for BaySAC, the methods to determine and update the prior probabilities of the data points will be introduced in detail in the following sections.

In the BaySAC algorithm, the hypothesis set that is most likely to be correct is chosen, and this is determined in terms of the prior inlier probabilities of the data points.

Prior Probability Estimation

Before the BaySAC process is started, the prior inlier possibilities need to be determined. The geometric constraints derived from the image overlap and the triangulation constructed using pseudo-corresponding point pairs, can be used to determine the prior probabilities of the pseudo-corresponding point pair. However, such a strategy is typically either platform-oriented (such as aerial-borne and satellite-based) or complex and computationally expensive. Since the fundamental matrices calculated by the consecutive inlier sets are expected to converge in the direction of the optimal solution, we propose a statistical testing strategy to determine the prior inlier probability based on this assumption, instead of the random prior probability using by the original BaySAC (Bottrell et al., 2009). This strategy is likely highly applicable and model-free.

Our presented strategy is based on a histogram to dynamically evaluate the convergence of the hypothesis FMs during the hypothesis testing process. Different convergent cluster of the hypothesis FMs will be presented in the histogram. We select the oldest parameter set as the reference point for each convergent cluster of parameter sets. The more the hypothesis FMs converge to a cluster, the more possible is it that the reference FM of the cluster is correct. Therefore, we present the convergence degree of a cluster to evaluate the correct possibility. The convergence degree is a percentage which describes the number of FMs converge to a FMs cluster. It is calculated by dividing the number of the hypothesis FMs in the cluster by the total number of hypothesis FMs.

To determine the convergence of a newly computed FM to a FMs cluster, we need to evaluate the deviation between the new FM and the reference FM of the cluster. As the fundamental matrix contains nine elements, we construct a vector with nine dimensions for each set of nine elements and compute the Euclidian distances between the different vectors to describe the deviation between the new FM and the reference FM of the cluster. If the deviation is smaller than the convergent range \(\Delta\), the number of FMs in the cluster is increased by one and the convergence degree is updated accordingly. If the new hypothesis FM doesn’t converge to any existing cluster, it will be regarded as a new cluster in the histogram.

Only for the purpose of visualization, the mean value of FM elements is used to construct the histogram of hypothesis FMs (Figure 1). The one-dimensional histogram illustrates the statistical process in which the vertical axis denotes the convergence degree of each cluster, and the horizontal axis represents the distribution of the FMs clusters. The mean value \(m(i = 1,...,n)\) of the nine elements of the reference FM \(i\) determines the location of the cluster \(i\); \(n\) is the number of the FMs clusters for which the degrees of convergence are computed. The width of a cluster is two times of the convergent range \(\Delta\).

Figure 1. The histogram of hypothesis FMs for statistical testing.
(i.e., eight pairs of corresponding points) to compute the hypothesis FM. The calculated hypothesis FM is then verified by other correspondence pairs.

In the hypothesis testing process of BaySAC, at each iteration the inlier probabilities of the current hypothesis set are updated in terms of the result of the verification of the current hypothesis model.

**Probability Updating**

The probability updating principle of BaySAC (Botterill et al., 2009) is formulated as follows:

\[
P_t(i \in I) = \frac{P_{t-1}(i \in I)P(H_t \subset I | i \in I)}{P(H_t \subset I)} , \quad i \in H_t
\]

where \( I \) is the set of all inliers, and \( H_t \) is the hypothesis set of \( n \) data points used at iteration \( t \) of the hypothesis testing process. \( P_{t-1}(i \in I) \) and \( P(H_t \subset I) \) denote the inlier probability for data point \( i \) at iteration \( t-1 \) and \( t \), respectively. \( P(H_t \subset I) \) is the probability of the presence of outliers in the hypothesis set, and \( P(H_t \subset I | i \in I) \) is the probability of the presence of outliers in the hypothesis set under the condition that point \( i \) is an inlier.

Equation 6 describes a memorable form of Bayes’ Theorem that states the posterior \( P(I|\inlier) \) is proportional to the prior \( P(I|\not\inlier) \) times the likelihood. For details about the deduction of Equation 6, please refer to (Lee, 2012):

\[
P(I|\inlier) \propto P(I|\not\inlier) \times \text{Likelihood} \quad (6)
\]

where **Likelihood** describes that the posterior is a function of the prior. The symbol \( \propto \) denotes the proportional relationship of the two events.

As **Likelihood** is a value describing the change relationship between the posterior and prior of each point in the hypothesis set taking the result of each hypothesis testing into consideration, we take \( P(I|\not\inlier) \) as the evaluation of this change relationship, which is the estimation of Likelihood. \( P(I|\not\inlier) \) is the probability that describes how possible the hypothesis model is the best one. It is calculated as \( k/D \), where \( k \) is the number of points in the consensus set after a hypothesis testing, and \( D \) is the total number of data points. Then the **Likelihood** is described by using Equation 7:

\[
\text{Likelihood} \approx k/D .
\]

Thus, we can obtain the simplified formula for updating the probability as follows:

\[
P_t(i \in I) = \begin{cases} 
\frac{k}{D} P_{t-1}(i \in I) & i \in H_t \\
0 & i \not\in H_t 
\end{cases}
\]

\[
(8)
\]

**Algorithm Process**

Figure 2 shows the flow chart of the proposed hypothesis testing process which proceeds as follows:

1. It first starts with the **RANSAC** strategy which chooses initial datasets randomly. In addition, during each iteration, the FM histogram is updated by using the newly calculated hypothesis FM. If the new FM does not converge to any existing candidate FM, it will be regarded as a new candidate in the histogram. Otherwise, we increase the convergence degree (CD) of the existing candidate FM.

2. When the CD of a candidate FM reaches a predefined threshold \( CD_r \), this FM is employed to determine the prior probabilities of data points. The hypothesis testing process is then implemented with the BaySAC strategy.

3. Select \( n \) data points with the highest inlier probabilities as the hypothesis set and compute the hypothesis FM by using this set.

4. Test the hypothesis FM with all the correspondence pairs and determine the inliers and outliers based on a certain threshold.

5. Update the inlier probabilities of each data point using the simplified Bayes’ rule.

6. Repeat the hypothesis testing process with new inlier probabilities until the sampling iteration number reaches the predefined threshold \( T \), which is computed as follows:

\[
T = \frac{\log(1-p)}{\log(1-w^n)}
\]

where \( p \) denotes the probability that the algorithm in some iterations selects only inliers from the input data set when it chooses the \( n \) points from which the model parameters are estimated. Let \( w \) be the probability of choosing an inlier each time a single point is selected, which is computed as the number of inliers in data divided by the number of data points.

7. Select the model consistent with most data points as the best model.

8. Use all inlier points to compute the optimal FM through the least-squares adjustment.
We demonstrate the algorithm process by using a small example regarding the computation of the fundamental matrix between two images. Twenty candidate correspondences for computing the FM are shown in Figure 3a, which contains three outliers, i.e., points numbered 18, 19, and 20 (highlighted in the box).

Figure 3. An example of the optimized BaySAC: (a) Candidate correspondences, and (b) Hypothesis FMs histogram.
of the hypothesis set are then updated according to Equation 8. In the next iteration, the hypothesis set is still chosen, with which the FM is computed, and 17 inlier correspondences are picked out, which indicates that we have found the best model consistent with all inliers. The probabilities of the current hypothesis set are also updated according to Equation 8. In the next iteration, the hypothesis set is still chosen, and the hypothesis process repeats. However, as the best model found in the third iteration is consistent with all inliers, no hypothesis FM calculated during the succeeding iterations can exceed it. As a result, the hypothesis testing process completes when the number of sampling iterations reaches the predefined threshold \( T \) calculated by using Equation 9.

In this example, \( p \) is set as 0.99 and \( n \) is 8; \( w \) is updated during each iteration. In the first iteration it equals 0.4, which is computed as the number (8) of the points in the hypothesis set, divided by the total number (20) of data points. According to Equation 9, \( T \) is also varying and is calculated as 7,025 in the first iteration.

Although the determination and evaluation of the probabilities also takes some time, this proposed strategy can remarkably reduce the number of iterations for finding a good model, and thus reduce the computational costs. For applications where the number of iterations is bounded, the probability of failing to find a model is reduced, which makes the computed results more reliable.

### Experimental Results

The performances of the original BaySAC and the BaySAC algorithm with the proposed strategies of the prior probability determination and the RANSAC framework were compared using real datasets in terms of computational efficiency and matching accuracy.

The real datasets (Figure 4) comprise image pairs acquired on different platforms (pedestrian, vehicle-based, and satellite-based) and with different types of stereo vision (lateral and along the optical axis). Table 2 describes the details of the experimental image pairs.

Table 3 lists the acronyms of four different evaluated hypothesis testing strategies. For performance comparison, the parallax constraints derived from the image overlap were used to determine the prior probabilities of pseudo-corresponding point pairs for the datasets of lateral stereo vision, which is noted as BaySAC-PC. The BaySAC (Original BaySAC using random prior probability) and BaySAC-ST (BaySAC using statistical testing to determine prior probability) strategies were tested on all datasets. All strategies were implemented in C++ and all of the tests were performed on a PC with an Intel i7-2720QM CPU processor running at 2.2 GHz with 4 GB RAM.

### Table 1. Probabilities Update Results

<table>
<thead>
<tr>
<th>ID</th>
<th>Prior probabilities (Iter.1)</th>
<th>Prior probabilities (Iter.2)</th>
<th>Prior probabilities (Iter.3)</th>
<th>Prior probabilities (Iter.4)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.9914 0.9914 0.685801 0.566525</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>0.9998 0.612439 0.612439 0.500336</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>0.9993 0.9993 0.744775 0.619731</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>0.9729 0.9729 0.547699 0.547699</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>0.9998 0.612439 0.612439 0.500336</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>0.9998 0.612439 0.612439 0.612439</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>0.9804 0.9804 0.603687 0.603687</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>0.9998 0.612439 0.612439 0.612439</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>0.9896 0.9896 0.672364 0.554402</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>0.9998 0.612439 0.612439 0.612439</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>0.9596 0.9596 0.9596 0.81355</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>0.9998 0.612439 0.612439 0.612439</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>0.9998 0.612439 0.612439 0.612439</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>0.9789 0.9789 0.592489 0.592489</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>0.9869 0.9869 0.652209 0.536217</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>16</td>
<td>0.9998 0.612439 0.612439 0.612439</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>17</td>
<td>0.9957 0.9957 0.717901 0.595485</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>18</td>
<td>0 0 0 0</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Figure 4. The experimental image pairs: (a) Image pair I, (b) Image pair II, (c) Image pair III, (d) Image pair IV, (e) Image pair V, and (f) Image pair VI.
the candidate FM is employed to determine the prior probabilities of data points by using Equation 4. Figure 5 shows the distributions of candidate fundamental matrices and their convergence degrees. We can see that there is always a peak of convergence degree that has a corresponding candidate FM that is predictably correct.

**Image Matching Accuracy**

The comparison of the results of the image-matching performance of the four robust estimation methods on six datasets is listed in Table 5. The image-matching performance is evaluated in terms of the number of the CPs matched by SIFT divided by the number of the actually correct CPs (C/M).

Image pairs I through III were acquired moving forward along the optical axis with different image resolutions (i.e., 3,456 × 2,304 pixels, 1,920 × 1,080 pixels and 4,288 × 2,848 pixels). The numbers of the correspondences extracted using the SIFT algorithm (with the threshold 0.8) are 538, 102, and 469, respectively. The RANSAC, original BaySAC, and our optimized BaySAC-ST algorithms were implemented to eliminate falsely accepted correspondences. As shown in Table 5, the BaySAC-ST algorithm reserved the most correct correspondences.

<table>
<thead>
<tr>
<th>Image pair</th>
<th>Robust estimation strategy</th>
<th>Matched CPs</th>
<th>Correct CPs</th>
<th>C/M</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td>RANSAC</td>
<td>536</td>
<td>461</td>
<td>86.0%</td>
</tr>
<tr>
<td></td>
<td>BaySAC</td>
<td>536</td>
<td>488</td>
<td>91.0%</td>
</tr>
<tr>
<td></td>
<td>BaySAC-ST</td>
<td>536</td>
<td>409</td>
<td>93.1%</td>
</tr>
<tr>
<td>II</td>
<td>RANSAC</td>
<td>131</td>
<td>62</td>
<td>47.3%</td>
</tr>
<tr>
<td></td>
<td>BaySAC</td>
<td>131</td>
<td>68</td>
<td>51.9%</td>
</tr>
<tr>
<td>III</td>
<td>RANSAC</td>
<td>469</td>
<td>243</td>
<td>51.8%</td>
</tr>
<tr>
<td></td>
<td>BaySAC</td>
<td>469</td>
<td>271</td>
<td>57.8%</td>
</tr>
<tr>
<td></td>
<td>BaySAC-ST</td>
<td>469</td>
<td>288</td>
<td>61.4%</td>
</tr>
<tr>
<td>IV</td>
<td>RANSAC</td>
<td>833</td>
<td>618</td>
<td>74.2%</td>
</tr>
<tr>
<td></td>
<td>BaySAC</td>
<td>833</td>
<td>664</td>
<td>79.7%</td>
</tr>
<tr>
<td></td>
<td>BaySAC-PC</td>
<td>833</td>
<td>699</td>
<td>83.9%</td>
</tr>
<tr>
<td></td>
<td>BaySAC-ST</td>
<td>833</td>
<td>692</td>
<td>83.1%</td>
</tr>
<tr>
<td>V</td>
<td>RANSAC</td>
<td>114</td>
<td>24</td>
<td>21.1%</td>
</tr>
<tr>
<td></td>
<td>BaySAC</td>
<td>114</td>
<td>33</td>
<td>28.9%</td>
</tr>
<tr>
<td></td>
<td>BaySAC-PC</td>
<td>114</td>
<td>16</td>
<td>14.0%</td>
</tr>
<tr>
<td></td>
<td>BaySAC-ST</td>
<td>114</td>
<td>40</td>
<td>35.1%</td>
</tr>
<tr>
<td>VI</td>
<td>RANSAC</td>
<td>434</td>
<td>358</td>
<td>82.5%</td>
</tr>
<tr>
<td></td>
<td>BaySAC</td>
<td>434</td>
<td>368</td>
<td>84.8%</td>
</tr>
<tr>
<td></td>
<td>BaySAC-PC</td>
<td>434</td>
<td>405</td>
<td>93.3%</td>
</tr>
<tr>
<td></td>
<td>BaySAC-ST</td>
<td>434</td>
<td>373</td>
<td>85.9%</td>
</tr>
</tbody>
</table>

**Table 5. Numerical Comparison of Computational Efficiencies**

<table>
<thead>
<tr>
<th>Image pair</th>
<th>RANSAC</th>
<th>BaySAC</th>
<th>BaySAC-ST-Random</th>
<th>BaySAC-ST</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td>0.0447s</td>
<td>0.0219s</td>
<td>0.0317s</td>
<td>0.0479s</td>
</tr>
<tr>
<td>II</td>
<td>0.1684s</td>
<td>0.1184s</td>
<td>0.0127s</td>
<td>0.0892s</td>
</tr>
<tr>
<td>III</td>
<td>23.941s</td>
<td>16.3505s</td>
<td>0.8257s</td>
<td>7.722s</td>
</tr>
<tr>
<td>IV</td>
<td>0.8636s</td>
<td>0.6887s</td>
<td>0.0959s</td>
<td>0.4969s</td>
</tr>
<tr>
<td>V</td>
<td>38.898s</td>
<td>26.191s</td>
<td>1.473s</td>
<td>13.943s</td>
</tr>
<tr>
<td>VI</td>
<td>0.1124s</td>
<td>0.0394s</td>
<td>0.0117s</td>
<td>0.0412s</td>
</tr>
</tbody>
</table>

The original BaySAC achieved better performances than RANSAC. Although there was a large difference in the images pose between the two images of image pair III, our proposed strategy identified 17 correct correspondences more than BaySAC and 45 correct correspondences more than RANSAC.

Image pairs IV through VI were captured with different lateral overlaps and image resolutions (i.e., 4,000 × 2,672 pixels, 4,288 × 2,848 pixels and 512 × 15,358/512 × 16,474 pixels). There were 833, 117, and 429 correspondences extracted using the SIFT algorithm. The robust estimation strategies RANSAC, BaySAC, BaySAC-PC, and BaySAC-ST were utilized for match selection. Table 5 shows that BaySAC-ST also achieved better performances than RANSAC and original BaySAC. The results also reveal that BaySAC-PC gained better matching accuracies than those of BaySAC-ST when there were approximately constant parallaxes between correspondences (for instance image pairs IV and VI). However, Figure 6 shows that there are multiple peaks in the histogram of the parallax between correspondences of the image pair V, which indicates that the parallax becomes variable in the close-ranged image pair. As a result, BaySAC-PC reserved only 16 correct correspondences, whereas the number of correct correspondences selected by BaySAC-ST was 40.

Figure 7 shows the enlarged views of the matching results. In Figure 7c, the BaySAC-ST eliminates the false correspondence (highlighted in the lower left boxes in 7a and 7b) accepted by the three other strategies. Moreover, BaySAC-ST selects more correct correspondences within the left window than those of other three strategies.

The unsatisfactory performances of BaySAC-PC are due to the strategy of prior probability determination, which is based on the assumption that the parallax between correspondences should be roughly constant.

**Computational Efficiency**

To verify the computational performances of the three different strategies, we performed a total of ten tests. In Figure 8, the horizontal and vertical axes represent the sequence number of implemented hypothesis testing (1 to 10) and their computational time, respectively. The computational time of RANSAC, BaySAC and BaySAC-PC are indicated by square, diamond and triangle, respectively (Figure 8). As previously described, the computational time of BaySAC-ST comprise two parts, i.e., the random (BaySAC-ST-Random) and BaySAC. These random parts are depicted by asterisk, and the total runtime is represented by cross, respectively. The random part consists of the consumed time when determining the prior probability using the statistical testing of the candidate model parameters sets.

Table 6 lists the numerical comparison of computational efficiencies. The computational costs of BaySAC-ST are overall smaller than those of RANSAC and BaySAC. Table 6 also shows that the two largest differences of computational costs among BaySAC-ST, RANSAC, and original BaySAC occur in the results of image pairs III and V. When performing a comparison with Table 5, we can find that the outlier ratios of the above two image pairs are the highest, which reveals that the larger the outlier ratio is, the higher efficiency achieved by BaySAC-ST compared with RANSAC and BaySAC. For instance, concerning image pair V, the average runtime of RANSAC and BaySAC are 38.988 sec and 26.191 sec, whereas the runtime of BaySAC-ST is only 13.943 sec. Figure 8 also illustrates that BaySAC-PC achieved the highest computational efficiencies in the results of image pairs IV and VI. However, BaySAC-ST resulted in the highest cost in the results of image pair V because of the varying parallax between correspondences in this close-ranged image pair.
Figure 5. The distributions of candidate fundamental matrices: (a) Image pair I, (b) Image pair II, (c) Image pair III, (d) Image pair IV, (e) Image pair V, and (f) Image pair VI.

Figure 6. Histogram of parallax distribution for image pair V.

**Conclusions**

In this paper, we have proposed a conditional sampling algorithm for image-matching based on the BaySAC framework. The hypothesis process always selects eight correspondences with the highest inlier probabilities as a hypothesis set. To implement BaySAC in the match selection process, we derived a strategy for prior probability determination from the statistical characteristic of the deterministic mathematical model for hypothesis testing. The inlier probability updating was simplified based on a memorable form of Bayes’ Theorem.

The proposed algorithm was implemented by using six image pairs and evaluated in terms of the computational efficiency and matching accuracy. The results reveal that the presented statistical testing strategy can always detect a peak of convergence degree, whose corresponding candidate FM is predictably correct and used to compute the prior probabilities of the pseudo-correspondences. When there are only a few outliers, computational time of this statistical process occupies a main part in the total runtime of BaySAC-ST, while it becomes negligible when the data point set is contaminated by a large proportion of outliers.
The matching results indicate that the BaySAC-ST algorithm can overall achieved better performances both for matching accuracy and computational efficiency than RANSAC and original BaySAC. The more outliers contain the pseudo-correspondences, the higher computational efficiency of our proposed algorithm gains compared with RANSAC and original BaySAC. The results also indicate that the proposed statistical testing strategy can determine sound prior inlier probability free of the changes of the platform and stereo vision with which the images are captured.

The fundamental matrix is a linear transform for images under the constraint of epipolar geometry. The eight-point algorithm is a linear method to calculate the fundamental matrix. Thus if the transform between images is nonlinear (e.g., panoramic image pair), a new method should be considered in our future work. Moreover, since the presented hypothesis testing algorithm has good applicability in determinate model estimation, future work will also optimize the proposed algorithm to solve other hypothesis testing problems of the determinate model, for example, primitive fitting.

Acknowledgments
The comments of anonymous reviewers for improvement of this paper are gratefully acknowledged. This research was supported by the Natural Science Foundation of China under Grant No. 41171358, the Fundamental Research Funds for the Central Universities under Grant No. 2652012103, and the National High Technology Research and Development Program of China (863 Program) with the serial number 2012AA12A204.

References


Figure 8. The computational efficiencies: (a) Image pair I, (b) Image pair II, (c) Image pair III, (d) Image pair IV, (e) Image pair V, and (f) Image pair VI.


(Received 05 October 2013; accepted 13 June 2014; final version 30 June 2014)