An Optimized BaySAC Algorithm for Efficient Fitting of Primitives in Point Clouds

Zhizhong Kang, Liqiang Zhang, Baoqian Wang, Zhen Li, and Fengman Jia

Abstract—Fitting primitives is of great importance for remote sensing applications, such as 3-D modeling and as-built surveys. This letter presents a method for fitting primitives that fuses the Bayesian sample consensus (BaySAC) algorithm with a statistical testing of candidate model parameters for unorganized 3-D point clouds. Instead of randomly choosing initial data sets, as in the random sample consensus (RANSAC), we implement a conditional sampling method, which is the BaySAC, to always select the minimum number of data required with the highest inlier probabilities. As the primitive parameters calculated by the different inlier sets should be convergent, this letter presents a statistical testing algorithm for the histogram of the candidate model parameter to compute the prior probability of each data point. Moreover, the probability update is implemented using the simplified Bayes formula. The proposed approach is tested with the data sets of planes, tori, and curved surfaces. The results show that the proposed optimized BaySAC can achieve high computational efficiency (five times higher than the efficiency of the RANSAC for fitting a subset of 12,500 points) and high fitting accuracy (on average, 20% higher than the accuracy of the RANSAC). Moreover, the strategy of prior probability determination is proven to be model-free and, thus, highly applicable.

Index Terms—Bayes sample consensus (BaySAC), fitting of primitives, hypothesis testing, point cloud, prior probability determination, random sample consensus (RANSAC).

I. INTRODUCTION

FITTING primitives is vital for extracting semantic information from the unorganized point clouds [1], [2] that are acquired by laser scanning or image matching. The majority of the existing techniques focus on the robust estimation of the primitive parameters (e.g., the 3-D Hough transform [3] and region growing [4]). A well-regarded technique for the segmentation and robust model fitting of point clouds is the statistical framework of the random sample consensus (RANSAC) [5]. Bartoli [6] proposed a multiple-hypotheses version of the RANSAC, which modifies the original algorithm by maximizing the likelihood fitting with a plane and allowing for a piecewise segmentation of overlapping planes. Henn et al. [7] applied the robust estimation method of the maximum likelihood estimation sample consensus [8] to derive the best-fitting roof models in a model-driven manner. Schnabel et al. [9] improved the efficiency of the RANSAC with local point selection and a simplified score function.

However, the original RANSAC algorithm assumes constant prior probabilities for data points and randomly chooses initial data sets, which likely leads to more iterations and expensive computation costs when the hypothesis set is contaminated with many outliers. To improve computational efficiency, this letter proposes a conditional sampling method based on the Bayesian sampling consensus (BaySAC) [10], 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. Instead of using specific characteristic information about a primitive, we optimize the BaySAC algorithm by presenting a novel algorithm of a statistical testing of candidate model parameters to compute the prior probability of each data point, which is thus predictably model-free. Moreover, the probability update is implemented using the simplified Bayes formula. The proposed method could work not only with laser scanning data but also with any type of dense 3-D data (e.g., data from image matching).

The main contribution of this letter is given in Section II, which introduces the optimized BaySAC method. Section III discusses our test results, after which we offer our conclusions and suggestions for further research in Section IV.

II. OPTIMIZED BAYSAC BASED ON STATISTICAL TESTING OF CANDIDATE MODEL PARAMETERS

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

Compared with the RANSAC, the key properties of the BaySAC are to determine and update the prior probabilities of the data points.

A. Prior Probability Estimation Based on Testing the Candidate Model Parameters

As the mathematical models of the primitives to be fitted are determinate, their parameters should be convergent when calculated by consecutive inlier sets. Therefore, we present a statistical testing algorithm for the candidate parameter set to compute the prior probability of each data point, which is
likely highly applicable and model-free. Kraus and Pfeifer [13] proposed a statistical histogram of the residuals of the data points deviating from the fitted terrain surface to determine the weight of every point for fitting the terrain surface. Our statistical testing process is implemented using a histogram that illustrates the distribution of the discrete hypothesis model parameter sets that are computed during different iterations and the degree of convergence of each considered candidate parameter set describing how other sets converge to it. This degree of convergence of a bin in the histogram is calculated as the number of parameter sets in that bin divided by the total number of parameter sets.

In this letter, two common geometric primitives, i.e., planes and tori, and free-curved surfaces are taken into consideration. Fig. 1 depicts the parameterization of the primitives.

**a) Standard geometric primitives.** To evaluate the degree of convergence of the two geometric primitives in Fig. 1, for each primitive, a suitable geometric measure is considered, as defined in the following. Instead of separately considering all the parameters for each parameterization, some parameters are combined such that in each case, a 2-D histogram can be used to evaluate the success of a certain choice of parameter values. To visualize the statistical process (e.g., a planar primitive), we illustrate it as a 2-D histogram [see Fig. 2(a)]. Therefore, the three parameters for the normal vector and the three parameters for the origin are degraded into two, i.e., the angle between the normal vector \( n \) of a plane and the horizontal plane (the horizontal axis), and the distance from the origin to the plane (the vertical axis). The upright axis represents the degree of convergence of each grid cell. In Fig. 2(a), the high parameter degree of convergence reflects a high probability of finding the correct planar parameters.

**b) Free surfaces.** Free surfaces are represented using the quadric parametric surface model [12]. As the model contains 27 parameters (for their definitions, see [12]), it is hard to visualize the different hypothesis model parameter sets in the geometric space. Therefore, we construct a vector with 27 dimensions for each set of the 27 parameters, and we compute the Euclidian distances between the different vectors to describe the deviation between the hypothesis parameter sets of a free surface computed during different iterations. As in Fig. 2(b), the statistical process is simply illustrated as a 1-D histogram, with the horizontal axis representing the distribution of the mean value \( m_i \) (\( i = 1, \ldots, n \)) of the 27 parameters and the vertical axis denoting the degree of convergence of the parameters of each cell whose width is two times that of the convergent range \( \Delta \).

During each iteration, the histogram of the candidate parameter sets is updated using the newly calculated hypothesis
during a test and

\[ P_t(i \in I) = \begin{cases} 1 - \frac{D_i}{m} & (D_i < m) \\ 0 & (D_i \geq m) \end{cases} \]

where \( P_t \) denotes the prior probability of point \( i \), \( D_i \) is the distance between point \( i \) and the fitted primitive, and \( m \) represents the predefined threshold for the outlier identification, which is set as five times the point precision.

### B. Probability Updating

The probability updating principle of the BaySAC [10] is as follows:

\[ P_t(i \in I) = \begin{cases} \frac{P_{t-1}(i \in I) P(H_t | i \in I)}{P(H_t)}, & i \in H_t \\ \frac{P_{t-1}(i \in I)}{P(H_t)}, & i \notin H_t \end{cases} \]  

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

Equation (3) describes a memorable form of Bayes’ theorem, which states that posterior \( P_t(i \in I) \) is proportional to prior \( P_{t-1}(i \in I) \times \text{Likelihood} \).

Here, “Likelihood” shows that the posterior is a function of the prior, and the symbol \( \propto \) denotes the proportional relationship of the two events.

As the probability that the hypothesis model is the best model that can be used to evaluate the probability that the corresponding hypothesis set is fully correct, we describe the likelihood as

\[ \text{Likelihood} \approx \frac{k}{D} \]

where \( k \) is the number of points that is consistent with the model during a test and \( D \) is the total number of data points. Therefore, we can have the simplified formula for updating the probability as follows:

\[ P_t(i \in I) = \begin{cases} \frac{k}{D} P_{t-1}(i \in I), & i \in H_t \\ P_{t-1}(i \in I), & i \notin H_t \end{cases} \]  

### C. Algorithm Process

We demonstrated the algorithm process on a small example on the fitting of a line.

Ten candidate points for fitting a line are shown in Fig. 3, which contains three outliers, i.e., points 7, 9, and 10.

The process starts with the RANSAC strategy, which randomly chooses an initial data set of two points from the ten candidate points. Meanwhile, the proposed statistical testing of the candidate line-parameter set is iteratively implemented using the newly calculated hypothesis parameter set. We compute the degree of convergence of a candidate parameter set in terms of the angles between the direction vectors of the lines fitted during different iterations. After 14 iterations, the highest degree of convergence reaches 50\%, which exceeds the predefined threshold of 30\%, and thus, its corresponding line parameters are used to determine the prior probabilities of the ten points using (1) (see Table I). We can see that the computed prior probabilities of the three outliers are 0.

Afterwards, the hypothesis testing process is implemented with the BaySAC strategy. We choose points 1 and 4 with the two highest probabilities to calculate the line parameters with which all points are then tested. The number of inlier points is five. The probabilities of points 1 and 4 are then updated according to (5), which completes the first iteration of the BaySAC. During the second iteration, points 3 and 8 are selected, with which the line model is computed and seven inlier points are picked out, which means that we have found the best model that is consistent with all the inliers. The probabilities of points 3 and 8 are also updated according to (5). In the third iteration, points 5 and 8 are chosen, and the hypothesis process repeats. However, as the best model that is found in the second iteration is consistent with all the inliers, no hypothesis line model that is calculated during the succeeding iterations can exceed it. As a result, the hypothesis testing process is completed when the number of sampling iterations reaches the predefined threshold.
III. EXPERIMENTAL RESULTS

The performance of the optimized BaySAC algorithm with the proposed strategies for prior probability determination and that of the RANSAC framework were compared in terms of the computational efficiency and fitting accuracy using real data sets. The real data sets comprise the point subsets that are selected from the terrestrial laser scanner data acquired by the RIEGL LMS-Z620 and VZ-400 laser scanners. The laser point subsets of the plane and torus are extracted from an industrial site, whereas the free-surface data were collected from a subway tunnel.

A. Computational Efficiency

Hypothesis testing is an iterative process; therefore, the computational efficiency of the proposed strategies were evaluated in terms of the number of iterations spent. The iterations of the RANSAC are shown as the dot-shaped icons in Fig. 4. The horizontal and vertical axes in Fig. 4 represent the numbers of the fitted laser point subsets and their consumed iterations, respectively. For example, the dot (10 000, 384) means that 384 iterations were made using the RANSAC to fit a plane to a subset of 10 000 points representing a plane [see Fig. 4(a)]. As described in Section II, the iterations of the BaySAC statistical testing (BaySAC-ST) comprise two parts, i.e., the random and the BaySAC. They are depicted by the square-shaped and diamond-shaped icons, respectively. The random part consists of the iterations that are consumed when determining the prior probability using the statistical testing of the candidate model parameter sets. As shown in Fig. 4, as the number of laser points in the fitted subsets increases, the rate of the spent iterations of the RANSAC increases. When using the BaySAC-ST, determining the prior probability using the statistical testing of the candidate model parameter sets consumes less than 50 iterations in most cases. As presented in Section II, after determining the prior probability of every data point, the strategy for the hypothesis testing changes from the RANSAC to the BaySAC. Fig. 4 then shows a remarkable decrease in the number of consumed iterations compared with the RANSAC.

B. Fitting Accuracy

To visualize the difference in the fitting accuracy, Fig. 5 illustrates the side views of the fitting results of the RANSAC and the optimized BaySAC on a subset of tunnel points containing a free-surface primitive, where the diamond-shaped points denote the original laser points, and the gray curves represent the fitted primitives (side view). The right section of Fig. 5(a) (a zoom-in view of the region that is highlighted with a rectangle in the left part) shows that the fitted surface that is acquired by the RANSAC fits the lower point segment well; however, it obviously deviates from the upper segment. The right section of Fig. 5(b) illustrates a more optimized fit, with the inliers evenly distributed around the fitted surface, which is achieved using the proposed optimized BaySAC.

To further test the fitting accuracy, we selected ten inliers as checkpoints for each primitive. The point-to-surface distances were then calculated and are shown in Table II. The results indicate that the optimized BaySAC algorithm achieved higher overall fitting accuracy.

TABLE II

<table>
<thead>
<tr>
<th>Primitive</th>
<th>Mean deviation (m)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Plane</td>
<td>0.0017</td>
</tr>
<tr>
<td>Torus</td>
<td>0.0016</td>
</tr>
<tr>
<td>Free surface</td>
<td>0.0014</td>
</tr>
</tbody>
</table>
IV. CONCLUSION

An optimized BaySAC algorithm for fitting the primitives from terrestrial point clouds has been proposed in this letter, which developed a strategy for determining the prior probability from the statistical characteristics of the deterministic mathematical model for hypothesis testing and simplified the inlier probability updating formula. The results show that the optimized BaySAC can sharply decrease the consumed iterations and can realize high computational efficiency. Moreover, the fitting results show that the optimized BaySAC algorithm achieves higher fitting accuracy on all experimental data sets. The robustness of the optimized BaySAC is ensured by the proposed strategy for prior probability determination, which utilizes the statistical characteristics of the determinate model in the hypothesis testing and thus is model-free. Future work will exploit other properties for reducing RANSAC iterations, such as the proximity of points, and will optimize the proposed algorithm to solve other hypothesis testing problems of the determinate model, e.g., the point cloud registration and the epipolar geometry estimation.

REFERENCES
