Robust Multiple Model Estimation With Jensen-Shannon Divergence

Kai Zhou, Karthik Mahesh Varadarajan, Michael Zillich and Markus Vincze
Automation and Control Institute, Vienna University of Technology
{zhou, kv, zillich, vincze}@acin.tuwien.ac.at

Abstract

In order to estimate multiple structures without prior knowledge of the noise scale, this paper utilizes Jensen-Shannon Divergence (JSD), which is a similarity measurement method, to represent the relations between pairwise data conceptually. This conceptual representation encompasses the geometrical relations between pairwise data as well as the information about whether pairwise data coexist in one model’s inlier set or not. Tests on datasets comprised of noisy inlier and a large percentage of outliers demonstrate that the proposed solution can efficiently estimate multiple models without prior information. Superior performance in terms of synthetic experiments and pragmatic tests is also demonstrated to validate the proposed approach.

1. Introduction

Multiple models estimation is a fundamental component in computer vision for salient data selection, feature extraction and data parameterization. Conventional approaches such as the RANSAC [1] family show limitations when dealing with data containing multiple models, high percentage of outliers or sample selection bias, commonly encountered in computer vision applications. Several extensions of RANSAC are capable of tackling the problem of multi-model estimation [8][14], but they require the number of models to be specified by the user beforehand and therefore are less robust in real applications. These RANSAC based algorithms, which more focus on hypothesis evaluation, generate hypothetical models, then the selection of the best hypothesis is solved as a verification process which calculates every hypothesis’s residual distribution (the normalized histogram that is generated with the distances from all the data points to this hypothesis).

A change of perspective to multi-model estimation has been presented in [13], where they researched on the residual distribution for each point instead of studying on the residual distribution per hypothesis since the modes (e.g. numbers of the peaks) in the former distribution reflect the numbers of the models. This method more concentrates on the inlier selection, i.e. cluster/label the data points to obtain the inlier sets of models directly instead of evaluating hypotheses, and penetratively discovers the number of models through analysis of data w.r.t the hypotheses. Since the direct analysis of data points is not plausible in practice (requires to generate a large number of hypotheses), the relations between/among data elements have been investigated for identifying inlier indirectly.

Recently more general conceptual representations of relations between/among data, which fuse the studies in both hypothesis evaluation and data selection, have been proposed for solving the multi-model estimation task [12][9][4][10][11][3]. The preference matrix [9], Mercer kernel matrix [4][12], or energy function [3] have been utilized to provide the global representations of relations between data points. However, all of them either need a user to specify the inlier scale [9][3], or introduce new user-dependent parameters (step $h$ in computing the ordered residual kernel [4], weighting factors $\alpha, \beta$ in the to-be-optimized objective function [12]). [10][11] estimate the multiple models with relations between data points by means of a series of procedures. Their scale estimation substep focuses on scoring hypothesis to select good hypotheses for estimate inlier scale and model fitting substep identifies inliers using the estimated inlier scale. The sequential process manner makes the following steps highly depend on the previous computational results, leading to decreasing robustness and flexibility of the algorithms.

We propose a conceptual representation which encompasses the geometrical relations between pairwise data as well as the information about whether pairwise data belong to the same model or not. The key technique to represent the relations between pairwise data points is Jensen-Shannon Divergence (JSD) [2], which
is frequently used for similarity measurement in various contexts. With the JSD-based modeling of relations between pairwise data points, we elaborate a general inlier selection mechanism for discovering the potential multiple models in the noisy data. Fig. 2 summarizes the proposed mechanism. Input data contains 100 points per line (All the synthetic data for 2D line fitting tests in this paper are corrupted with Gaussian noise of standard deviation \( \sigma = 0.01 \)) and 600 gross outliers (outlier rate of 90%). Color ranges from red to blue, passes through orange, yellow and cyan, illustrates the order of data point selection.

2. Relation Between Pairwise Data

This section describes how to model the relations between pairwise data points using JSD. Let the model to be estimated be determined by \( p \) parameters. Given the input data \( \mathcal{L} = \{x_i\}_{i=1,...,N} \) and model hypotheses \( \mathcal{H} = \{\theta_j\}_{j=1,...,M} \) which are generated by randomly sampling from \( \mathcal{L} \), we calculate the distance vector \( d_i \) of all the hypotheses to each data point \( x_i \) in \( \mathcal{L} \), \( d_i = \{d_{i1},\ldots,d_{iM}\} \), where \( d_{im} \) is the distance of point \( x_i \) to hypothesis \( \theta_m \). Note in particular that we do not truncate the distances therefore no inlier scale specification or estimation is required.

2.1 Theoretical Interpretation

Our approach is inspired by several pioneering work which also uses the relations between pairwise data points to cluster/label data points since in practice it is almost impossible to find a representation function that can be used to evaluate the data points directly. [9] use Jaccard distance to represent the relations between pairwise data, these pairwise relations are naturally utilized as linkage measurement for a connectivity based clustering algorithm. [4] uses the ordered residual kernel to measure the similarities between pairwise data points. Their especially tailored kernel satisfies Mercer condition, thus it can be considered as the inner product of two data points. Then the kernel trick can be built upon it for using the statistic learning methods. However, when these two methods construct the relations between data points, either the inlier noise scale [9] or the step \( h \) for generating the difference of intersection between residuals [4] is required and crucial to the performance of the algorithm. On the contrary, the proposed JSD-based method does not require any user input parameters.

Jensen-Shannon divergence, which is based on Kullback-Leibler divergence (KLD) [5], is widely used in probability theory and statistics for measuring the similarity between two distributions. The superiority of JSD over KLD on handling zero values in distributions and removal of nuisance in the use of KLD arising from its asymmetry has been demonstrated in [6].

Since the row vectors in the generalized distance matrix \( G \) have already been normalized (sum to 1), the Jensen-Shannon divergence (JSD) between \( g_i \) and \( g_j \) can be directly calculated by

\[
G = \tilde{D} \cdot S = \{g_1;\ldots;g_N\} = \left\{ \frac{d_{\lambda_1}}{s_{\lambda_1}};\ldots;\frac{d_{\lambda_N}}{s_{\lambda_N}} \right\}
\]
that every point $p_i$ is followed by the point which has the minimum JSD $\min JSD(i)$ in the set $\{JSD(i, i + 1), JSD(i, i + 2), \ldots, JSD(i, N)\}$. Algorithm 1 illustrates the inlier discovery scheme.

If the initial point is the endpoint (inlier that exists at the border of the structure) of the ‘true’ model, the search scheme could generate the perfect order of the inliers and outliers. However, this condition can not be satisfied in practice since this prior information may not available or in multi-model case the endpoint of one model may not the start point of another one. Therefore, various alternating inlier and outlier clusters will be generated using this search scheme (fig. 2 color varies from red (inliers) to yellow (inliers of another model) through orange (outliers)). Since the JSD between outliers are significantly larger than the JSD between inliers, we can monitor the accumulative mean value of $\min JSD$ to detect the outliers. The accumulative mean value of $\min JSD$ is defined as,

$$\text{meanMin}(i) = \frac{\sum_{j=1}^{i} \min JSD(j)}{i}$$

(4)

Within the selected inliers, any robust fitting method such as LMedS [7] can be utilized to estimate the parameters of the structures since most of the gross outliers are omitted from the data set.

4. Experiments

We start to test our JSD-based inlier discovery algorithm with synthetic data for both single model and multi-model estimation, and then demonstrate the application of our methods in a plane estimation task with RGB-D point cloud data. All the results of synthetic data processing are obtained on an Intel Quad Core 2.4

\begin{algorithm}
\caption{Inlier discover scheme}
1: \textbf{input:} GDM $G = \{g_1, \ldots, g_N\} = \{g^1, \ldots, g^M\}$
2: \textbf{while} Point index $i < N$ \textbf{do}
3: \hspace{1em} $\forall g_{j}, j \in (i + 1, N]$, calculate $JSD(i, j)$.
4: \hspace{1em} find $j_{\text{min}}$ make $JSD(i, j_{\text{min}}) \leq \forall JSD(i, j)$.
5: \hspace{1em} store $JSD(i, j_{\text{min}})$ into $\min JSD(i)$
6: \hspace{1em} swap $g_{i+1}$ and $g_{j_{\text{min}}}$.
7: \hspace{1em} sort $G$ as $\{g^{\mu_1}, g^{\mu_2}, \ldots, g^{\mu_M}\}$ that satisfies $g^{\mu_1} \leq g^{\mu_2} \leq \cdots \leq g^{\mu_M}$
8: \hspace{1em} $i = i + 1$
9: \textbf{end while}
10: \textbf{for} $l = 2, \ldots, N$ \textbf{do}
11: \hspace{1em} calculate $\text{meanMin}(l)$ using eq. 4
12: \hspace{1em} if $\text{meanMin}(l-1) \geq \text{meanMin}(l)$ \textbf{then}
13: \hspace{2em} label point $l$ as inlier
14: \hspace{1em} \textbf{end if}
15: \textbf{end for}
\end{algorithm}
GHz CPU, 4 GB RAM and the real image task is operated on an Intel Mobile Core i7 2720QM CPU, 8 GB RAM laptop connected to a Microsoft Kinect.

Fig. 3 depicts the inlier selection processes and results. Both two tested cases contain 100 inliers per line and with 90% outliers rate. Fig. 4 displays the plane estimation results of RGB-D point cloud data. All the point clouds are down-sampled to 1500 points and their orders of being searched are demonstrated using color varying, along with the side views. Animations that show the line/plane estimation processes using the proposed method can be seen at our website \(^1\).

5. Conclusion

We have presented a new method for multiple structure estimation. We use the Jensen-Shannon Divergence to represent the relations between pairwise data points, followed by an inlier discovery scheme that monitors the change of the minimum JSD value for deleting outliers. Finally, to recover multiple structures from the labeled inliers, we perform robust fitting with LMedS. Our experimental results on a large number of synthetic and real data validate the proposed approach.

References