Augmented Maximum Correntropy Criterion for Robust Geometric Perception

Jiayuan Li, Qingwu Hu, Xinyi Liu, and Yongjun Zhang, *Member, IEEE*

***Abstract*—Maximum correntropy criterion (MCC) is a robust and powerful technique to handle heavy-tailed nonGaussian noise, which has many applications in the fields of vision, signal processing, machine learning, etc. In this article, we introduce several contributions to the MCC and propose an augmented MCC (AMCC), which raises the robustness of classic MCC variants for robust fitting to an unprecedented level. Our first contribution is to present an accurate bandwidth estimation algorithm based on the probability density function (PDF) matching, which solves the instability problem of the Silverman's rule. Our second contribution is to introduce the idea of graduated nonconvexity (GNC) and a worst-rejection strategy into MCC, which compensates for the sensitivity of MCC to high outlier ratios. Our third contribution is to provide a definition of local distribution measure to evaluate the quality of inliers, which makes the MCC no longer limited to random outliers but is generally suitable for both random and clustered outliers. Our fourth contribution is to show the generalizability of the proposed AMCC by providing eight application examples in geometry perception and performing comprehensive evaluations on five of them. Our experiments demonstrate that 1) AMCC is empirically robust to 80%−90% of random outliers across applications, which is much better than Cauchy M-estimation, MCC, and GNC-GM; 2) AMCC achieves excellent performance in clustered outliers, whose success rate is 60%−70% percentage points higher than the second-ranked method at 80% of outliers; 3) AMCC can run in real-time, which is 10−100 times faster than RANSAC-type methods in low-dimensional estimation problems with high outlier ratios. This gap will increase exponentially with the model dimension.**

***Index Terms*—Geometry perception, maximum correntropy criterion (MCC), outliers, robust estimation, robust fitting.**

I. Introduction

ROBUST geometric fitting, a technique of simultaneously estimating geometric models (e.g., line/circle, affine, rigid, camera poses, and pose graph) and filtering outliers from contaminated observations (e.g., image features, 3-D correspondences, and relative poses), is an important and fundamental issue in robotics and computer vision. It has a wide spread of applications such as image matching [1], [2], point cloud registration (PCR) [3], [4], satellite pose estimation [5], [6], structure from motion (SfM) [7], simultaneous localization and mapping (SLAM) [8], [9], to name a few. Generally, robust geometric fitting methods can be divided into two groups, i.e., global solvers and nonglobal approaches. The very high computational complexity of global solvers makes them rarely used in practice. Therefore, this article focuses on nonglobal solvers that are much faster and sufficiently accurate, among which the random sample consensus (RANSAC) family [10], M-estimation [11], [12], and maximum correntropy criterion (MCC) family [13] are the most popular methods.

RANSAC [10] approximately solves the consensus maximization problem through a hypothesis and verification framework. It alternates between random sample fitting (hypothesis) and model consensus calculation (verification) until the stop criterion of the iterative program is reached. The model corresponding to the largest consensus set is accepted as the final solution. RANSAC is an intuitive, flexible, and highly robust method. However, RANSAC and its variants still have some limitations: 1) RANSAC-type methods are approximate solvers [14]; 2) Its computational complexity grows exponentially with the outlier ratio; 3) RANSAC-family can hardly be applied to high-dimensional fitting problems.

M-estimation [11], [12] is a concept from statistics. Typical M-estimations include Huber, Cauchy, Welsch, Biweight, and Geman-McClure (GM), whose idea is to penalize outliers by a robust cost. M-estimations have several advantages over RANSAC-type methods: 1) M-estimations are globally optimal if good initializations are available; 2) Its computational complexity is independent of the outlier ratio; 3) M-estimations are suitable for both low- and high-dimensional problems. However, traditional M-estimations are only suitable for low outlier ratio problems ($<50\%$) [14]. With the efforts of scholars, some M-estimation variants (e.g., graduated nonconvexity (GNC) [15], [16]) can handle 70% of random outliers, but they are still far from the RANSAC.

MCC is also an M-estimation in essence [13], [17], whose central idea is to maximize the correntropy between two given variables according to the information potential and has been widely used in signal processing and machine learning. It inherits all the advantages of M-estimation. Moreover, since MCC analyzes the residual errors from a probabilistic perspective and adaptively estimates the kernel bandwidth to control the shape of the kernel function, it is applicable to any noise environment [18]. Although MCC is more robust than conventional M-estimations, it is less stable and suffers from failures even at low outlier ratios.

Received 28 June 2024; accepted 19 September 2024. Date of publication 21 October 2024; date of current version 19 November 2024. This work was supported by the National Natural Science Foundation of China (NSFC) under Grant 42030102 and Grant 42271444. This article was recommended for publication by Associate Editor and Editor S. Behnke upon evaluation of the reviewers' comments. *(Corresponding authors: Xinyi Liu; Yongjun Zhang.)*

The authors are with the School of Remote Sensing and Information Engineering, Wuhan University, Wuhan 430072, China (e-mail: liuxy0319@whu.edu.cn; zhangyj@whu.edu.cn).

Data is available online at https://github.com/LJY-WHU/AMCC.

Digital Object Identifier 10.1109/TRO.2024.3484608

Current RANSAC-type methods, M-estimations, and MCC-family cannot handle clustered outliers well (see Fig. 3 for examples of clustered outliers). Clustered outliers are important for robotics and computer vision. For example, in SLAM or SfM, dynamic objects (vehicles, pedestrians, etc.) weaken the accuracy of existing methods, since dynamic correspondences (extracted on dynamic objects) and static ones have different motion models. Once the number of dynamic correspondences dominates, motion tracking is easy to fail. Actually, dynamic objects are much smaller compared with the static scene and dynamic correspondences are spatially clustered.

This article aims to propose a robust algorithm with properties: 1) real-time efficiency; 2) local optimal solution; 3) high robustness close to RANSAC; 4) good generalization to different geometric problems; and 5) robust to both random and clustered outliers. To achieve these goals, we propose a general and scalable estimator, named augmented maximum correntropy criterion (AMCC), which raises the robustness of classic MCC variants to an unprecedented level. Extensive experiments on simulated and real data demonstrate that AMCC significantly outperforms current state-of-the-arts in terms of robustness or efficiency, especially in cases with clustered outliers. In summary, our contributions are as follows.

1) We propose a fast, accurate, and robust algorithm, i.e., AMCC, for geometric fitting;
2) We present a bandwidth estimator via probability density function (PDF) matching to solve the instability of the Silverman's rule [19];
3) We introduce the idea of GNC and a worst-rejection strategy into MCC to improve its robustness;
4) We propose a novel concept of local distribution measure (LDM) based on local and global PDF to measure the inlier quality.

Novelty with respect to GNC methods: Using GNC to improve the robustness of M-estimation is not a new idea [15], [16], [20]. Compared with M-estimation, MCC is suitable for any noise environment [18]. One insight is that we can further improve the robustness of MCC based on the close relationship between MCC and M-estimation. It is the first time that the GNC is used in the framework of MCC.

II. Related Work

Here, two types of relevant methods are reviewed, i.e., RANSAC methods and M-like estimations.

A. RANSAC-Family

Combinatorial formulations are popular for outlier removal in low-dimensional geometric fitting problems (e.g., line/circle fitting and affine/rigid estimation) [21], among which consensus maximization that seeks to find a solution maximizing the inlier set is one of the most widely used formulations as follows:

$$\begin{aligned} &\arg\max_{\boldsymbol{\theta}} |\mathcal{I}| \\ &\text{s.t. } r_i \leq \varepsilon \quad \forall i \in \mathcal{I} \end{aligned} \tag{1}$$

where $\boldsymbol{\theta}$ is a geometric model, r_i is the residual error of the ith observation, ε is an inlier threshold, $\mathcal{I}$ represents an inlier set, and $|\mathcal{I}|$ denotes the size of set $\mathcal{I}$.

RANSAC is the most intuitive way to solve the problem (1), whose basic idea is to iteratively search for the best solution with a given confidence using a hypothesize-and-verify framework, i.e., random sample fitting (hypothesis) and model consensus calculation (verification). There are many variants of RANSAC, which improve one or several steps of the traditional RANSAC. For example, NAPSAC [22], PROSAC [23], and progressive NAPSAC [24] improve random sampling with the help of local neighbors or similarity scores. LORANSAC [25], FLORANSAC [26], and Graph-cut RANSAC [27] introduce a local optimization stage into RANSAC for further refinement based on the observation that models estimated by minimal sample sets are inconsistent with the one estimated by all inliers. R-RANSAC [28] presents a $T_{d,d}$ test to improve the efficiency of RANSAC, where the $T_{d,d}$ is passed if all d data points out of d randomly selected are consistent with the hypothesized model. R-RANSAC-SPRT [29] and optimal randomized RANSAC [30] introduce sequential probability ratio tests (SPRT) to further accelerate the iterative procedure. Contrario RANSAC [31], MAGSAC [32], and MAGSAC++ [33] present soft threshold methods to reduce the dependence on the inlier/outlier decision parameter. USAC [34], USACv20 [35], and VSAC [36] integrate several improvements into their frameworks. In summary, the advantages of RANSAC-type methods are conceptually simple, flexible, and highly robust; and their disadvantages are nonoptimal, time-consuming, and only suitable for low-dimensional problems.

B. M-Like Estimations

M-estimation is the preferred robust technique for high-dimensional geometric perception problems [e.g., bundle adjustment and pose graph optimization (PGO)], whose basic idea is to use a robust cost to penalize outliers by assigning small weights. It is generally formulated as follows:

$$\arg\min_{\boldsymbol{\theta}} \sum_{i=1}^{n} \rho\left(r_i\right) \tag{2}$$

where $\rho(\cdot)$ is a redescending robust cost function. Unlike the combinatorial formulation of RANSAC, M-estimation is a continuous optimization problem [21]. This essentially determines that M-estimation is optimal (local) while RANSAC methods are approximate solvers. M-estimations (e.g., Huber, Cauchy, Welsch, Biweight, and GM) are the generalizations of the maximum-likelihood estimator. S-estimations [37] improve the scale estimator of M-estimations via minimizing the dispersion of residual errors. MM-estimations [38] combine the advantages of M-estimations and S-estimations, i.e., they use S-estimations for scale regression and apply M-estimations for fixed-scale model regression. The common limitations of these traditional M-like estimations are that they are only suitable for problems with low outlier ratios ($<50\%$) and are prone to getting stuck in local minima.

Recently, scholars try to solve the bottleneck problem of the M-like estimations mentioned above. For instance, PbM-estimator [40] and generalized pbM-estimator [41] reformulate the M-estimation as a projection-based optimization problem based on the errors-in-variables model. Weighted q-norm estimator [42] combines a l_q-norm $(0 < q < 1)$ cost and an M-estimator for geometric fitting. GNC [15], [16] introduces a surrogate cost with a control parameter to optimize the M-estimation problem. It starts from a convex surrogate and changes the control parameter to increase the amount of non-convexity until convergence. Barron [43] proposed an adaptive robust cost. Although many efforts have been made, the robustness of these methods is still far from RANSAC.

III. MCC Solver Revisit

A. Background

Correntropy proposed in the information theoretic learning is often used to process nonGaussian noise and impulsive noise [18], [44], [45], [46]. It is a local similarity measure between two random variables A and B as follows:

$$V_\sigma(A,B) = \mathbb{E}\left[k_\sigma\left(A-B\right)\right] \tag{3}$$

where V_σ represents the correntropy, $\mathbb{E}$ denotes the expectation operator, and $k_\sigma(\cdot)$ is a kernel function with a bandwidth σ that satisfies the Mercer's theorem [47], i.e., $k_\sigma(\cdot)$ should be a symmetric positive-definite function. The Gaussian kernel $G_\sigma(\cdot)$ is the most popular kernel function.

In practice, the expectation $\mathbb{E}[k_\sigma(A-B)]$ can only be approximated by the sample mean, since the true joint PDF is unknown and only a finite number of samples $\{(A_i,B_i)\}_1^n$ are available. Thus, the correntropy becomes

$$\begin{aligned} V_\sigma(A,B) &\approx \frac{1}{n}\sum_{i=1}^{n} k_\sigma\left(A_i-B_i\right) \\ &\approx \frac{1}{n}\sum_{i=1}^{n} k_\sigma\left(r_i\right) \end{aligned} \tag{4}$$

where $r_i = A_i - B_i$ is a residual.

Further, Liu et al. [13], [17] proposed a robust criterion called MCC for an adaptive model with parameters $\boldsymbol{\theta}$ based on the sample-based correntropy, i.e., (4), as follows:

$$\begin{aligned} \boldsymbol{\theta}^* &= \arg\max_{\boldsymbol{\theta}} \frac{1}{n}\sum_{i=1}^{n} k_\sigma\left(A_i-B_i\right) \\ &= \arg\max_{\boldsymbol{\theta}} \frac{1}{n}\sum_{i=1}^{n} k_\sigma\left(r_i\left(\boldsymbol{\theta}\right)\right) \end{aligned} \tag{5}$$

where A represents the model output, e.g., $A = \{kx_1 + m, \ldots, kx_n + m\}$ for line fitting (LF) (see example 1 for details), $A = \mathbf{R}$ for single rotation averaging (SRA) (Example 4), and $A = \{\pi(\mathbf{P}_1\tilde{\boldsymbol{X}}), \ldots, \pi(\mathbf{P}_n\tilde{\boldsymbol{X}})\}$ for multiview triangulation (MT) (Example 7); B is the desired response, e.g., $B = \{y_1, \ldots, y_n\}$ for LF, $B = \{\tilde{\mathbf{R}}_1, \ldots, \tilde{\mathbf{R}}_n\}$ for SRA, and $B = \{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n\}$ for MT; $\boldsymbol{\theta}^*$ is the (local) optimal solution. Essentially, the input of MCC is a set of errors $\{r_i(\boldsymbol{\theta})\}_1^n$.

Many MCC variants have been proposed to improve the original MCC. For example, Chen et al. [46] extended MCC by considering a nonzero mean Gaussian kernel; GMCC [44] introduces a generalized Gaussian density function as the kernel; KMCC-L0 [48] proposes a sparse MCC cost with a constraint of approximated l_0-norm. Thanks to the excellent performance of MCC, it has a wide range of applications in signal processing and vision tasks, such as filtering [45], [49], [50], face recognition [18], classification [51], [52], principal component analysis [53], matrix completion [54], and learning [55]. Recently, MCC has also been introduced into geometric perception problems. For instance, Liang et al. [56] introduced MCC into the constrained least-square ellipse fitting problem; Hu et al. [57] further improved the MCC-based ellipse fitting by using a Laplacian kernel. Wu et al. [58] used MCC as the similarity measure of the iterative closest point for PCR; Huang et al. [59] proposed a motion averaging framework based on Laplacian kernel-based MCC. The main drawbacks of MCC in geometric fitting are 1) the sensitivity to high outlier ratios; 2) the instability of kernel bandwidth estimation; and 3) failures under clustered outliers.

B. Solver

Generally, the optimization of the above MCC problem involves two subproblems.

1) *Bandwidth σ update*: Calculate σ with fixed residual errors $\boldsymbol{r}^{(t-1)} = \{r_i^{(t-1)}\}_1^n$ (t is an iteration counter and $\boldsymbol{r}^{(t-1)}$ is a residual vector that consists of the residuals of the n observations in the $(t-1)$th iteration.) based on the Silverman's rule [13], [19], which is an empirical criterion that estimates bandwidth based on the standard deviation and size of samples as follows:

$$\sigma^{(t)} = 1.06 \times \min\left(\text{std}(\boldsymbol{r}^{(t-1)}), \frac{R}{1.34}\right) \times n^{-0.2} \tag{6}$$

where $\text{std}(\cdot)$ denotes a standard deviation operator and R is the interquartile range of the errors.

2) *Model $\boldsymbol{\theta}$ update*: Optimize model parameters $\boldsymbol{\theta}^{(t)}$ based on (5) with a known bandwidth $\sigma^{(t)}$ as follows:

$$\boldsymbol{\theta}^{(t)} = \arg\max_{\boldsymbol{\theta}} \frac{1}{n}\sum_{i=1}^{n} k_{\sigma^{(t)}}\left(r_i\left(\boldsymbol{\theta}\right)\right). \tag{7}$$

Then, the residual errors are updated according to $\boldsymbol{\theta}^{(t)}$, obtaining $\boldsymbol{r}^{(t)}$.

These two subproblems are iterated alternately until convergence to obtain a local optimal solution $\boldsymbol{\theta}^*$.

C. Application Examples

MCC has a wide range of applications in signal processing and machine learning, e.g., face recognition, signal filtering, etc. However, we are more interested in its applications in geometric perception. Here, we provide eight commonly used geometric fitting problems in vision, including LF, circle fitting (CF), image feature matching (IFM), SRA, perspective-n-points (PnP), PCR, MT, and PGO.

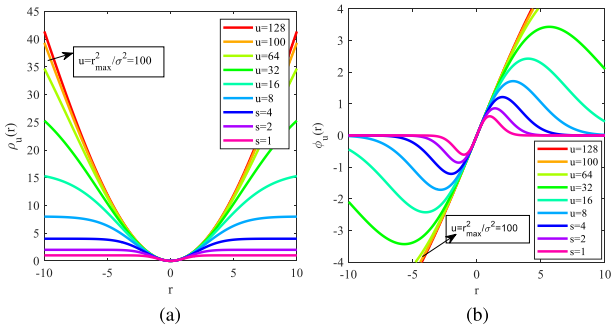


Fig. 1. Cost and influence functions of our surrogate cost for MCC, where $\sigma = 1$ and largest residual $r_{\max}$=10. (a) Cost function. (b) Influence function.

Example 1 (LF): Given n 2-D outlier-contaminated points $\{(x_i, y_i) \in \mathbb{R}^2\}_1^n$, the goal of LF is to find the optimal straight line $\boldsymbol{\theta} = (k, m)$, where k is the slope and m is the intercept term. The residual function is: $r_i(\boldsymbol{\theta}) = y_i - (kx_i + m)$. Fig. 2(a) plots an instance of LF.

Example 2 (CF): Given n 2-D outlier-contaminated points $\{(x_i, y_i) \in \mathbb{R}^2\}_1^n$, the goal is to seek the optimal parameters $\boldsymbol{\theta} = (x_0, y_0, \tilde{R})$ that fit a circle, where (x_0, y_0) is the circle center and $\tilde{R}$ is the radius. The residual function is chosen as: $r_i(\boldsymbol{\theta}) = \sqrt{|(x_i - x_0)^2 + (y_i - y_0)^2 - \tilde{R}^2|}$. Fig. 2(b) plots an instance of CF.

Example 3 (IFM): Given n 2-D-2-D outlier-contaminated image feature correspondences $\{(\boldsymbol{x}_i \in \mathbb{R}^2, \boldsymbol{y}_i \in \mathbb{R}^2)\}_1^n$ (e.g., matched by scale invariant feature transform [60] or radiation-variation insensitive feature transform (RIFT) [61]), the goal is to estimate the best warping function $f : \mathbb{R}^2 \mapsto \mathbb{R}^2$ that aligns these two point sets. Here, we choose the affine transformation as the warping function and its corresponding residual function is: $r_i(\boldsymbol{\theta}) = \|\mathbf{A}\boldsymbol{x}_i + \boldsymbol{t} - \boldsymbol{y}_i\|_2$, where $\boldsymbol{\theta} = (\mathbf{A}, \boldsymbol{t})$, $\mathbf{A} \in \mathbb{R}^{2\times 2}$ is an affine matrix, and $\boldsymbol{t} \in \mathbb{R}^2$ is a translation vector. Fig. 2(c) plots an instance of IFM.

Example 4 (SRA): Given n outlier-contaminated observations of an unknown 3-D rotation $\{\tilde{\mathbf{R}}_i \in SO(3)\}_1^n$, the goal of SRA is to estimate the optimal average rotation $\boldsymbol{\theta} = \mathbf{R} \in SO(3)$. The residual function is the chordal distance between $\mathbf{R}$ and $\tilde{\mathbf{R}}_i$: $r_i(\boldsymbol{\theta}) = \|\mathbf{R} - \tilde{\mathbf{R}}_i\|_F$, where $\|\cdot\|_F$ is the Frobenius norm. Fig. 2(d) plots an instance of SRA.

Example 5 (PnP): Given n 3-D-2-D outlier-contaminated correspondences $\{(\boldsymbol{X}_i \in \mathbb{R}^3, \boldsymbol{x}_i \in \mathbb{R}^2)\}_1^n$, where $\{\boldsymbol{X}_i\}_i^n$ are 3-D object points and $\{\boldsymbol{y}_i\}_i^n$ are their corresponding projections on an image, the goal is to estimate the best function $f : \mathbb{R}^3 \mapsto \mathbb{R}^2$ that maps the 3-D object onto an image. PnP is also known as the image orientation. The residual function is chosen as the reprojection distance as follows: $r_i(\boldsymbol{\theta}) = \left\| \frac{\mathbf{P}_{(1:2)}[\boldsymbol{X}_i \ \ 1]^T}{\mathbf{P}_{(3)}[\boldsymbol{X}_i \ \ 1]^T} - \boldsymbol{x}_i \right\|_2$, where $\mathbf{P} = \mathbf{K}[\mathbf{R} \ \boldsymbol{t}] \in \mathbb{R}^{3\times 4}$ is the camera projection matrix ($\mathbf{P}_{(j)}$ is the jth row of $\mathbf{P}$), $\mathbf{K} \in \mathbb{R}^{3\times 3}$ is a precalibrated camera internal parameter matrix, $\mathbf{R} \in SO(3)$ is a rotation matrix, $\boldsymbol{t} \in \mathbb{R}^3$ is a translation vector, and $\boldsymbol{\theta} = (\mathbf{R}, \boldsymbol{t})$. Fig. 2(e) plots an instance of PnP.

Example 6 (PCR): Given n 3-D-3-D outlier-contaminated point correspondences $\{(\boldsymbol{X}_i \in \mathbb{R}^3, \boldsymbol{Y}_i \in \mathbb{R}^3)\}_1^n$ extracted from point clouds $\mathcal{P}$ and $\mathcal{Q}$ (e.g., fast point feature histograms (FPFH) [62] and deep-learned features [63], [64]), the goal is to estimate the best mapping function $f : \mathbb{R}^3 \mapsto \mathbb{R}^3$ that aligns point clouds $\mathcal{P}$ and $\mathcal{Q}$. Here, 6-DOF rigid transformation $\boldsymbol{\theta} = (\mathbf{R}, \boldsymbol{t}) \in SO(3) \times \mathbb{R}^3$ is chosen as the model and the residual function is: $r_i(\boldsymbol{\theta}) = \|\mathbf{R}\boldsymbol{X}_i + \boldsymbol{t} - \boldsymbol{Y}_i\|_2$, where $\mathbf{R} \in SO(3)$ is a rotation matrix and $\boldsymbol{t} \in \mathbb{R}^3$ is a translation vector. Fig. 2(f) plots an instance of PCR.

Example 7 (MT): Given n 2-D outlier-contaminated image observations $\{\boldsymbol{x}_i \in \mathbb{R}^2\}_1^n$ obtained from n images of the same object point $\boldsymbol{X} \in \mathbb{R}^3$ and the camera projection matrices $\{\mathbf{P}_i \in \mathbb{R}^{3\times 4}\}_1^n$ of these n images, the goal of MT is to estimate the optimal object point $\boldsymbol{\theta} = \boldsymbol{X}$ [65], [66]. The residual function is: $r_i(\boldsymbol{\theta}) = ||\pi(\mathbf{P}_i\tilde{\boldsymbol{X}}) - \boldsymbol{x}_i||_2$, where π is the reprojection function and $\tilde{\boldsymbol{X}} = [\boldsymbol{X}, 1]^T$. In triangulation, object point $\boldsymbol{X}$ must always be in front of the cameras. Thus, it must satisfy the following constraints: $\{\mathbf{P}_{i:(3)}\tilde{\boldsymbol{X}} > 0\}_1^n$, where $\mathbf{P}_{i:(3)}$ is the third row of $\mathbf{P}_i$. Fig. 2(g) plots an instance of MT.

Example 8 (PGO): Given n pairwise relative pose observations $\{(\mathbf{R}_{ij} \in SO(3), \boldsymbol{t}_{ij} \in \mathbb{R}^3)\}_1^n$, the goal is to recover a set of poses $\{(\mathbf{R}_i \in SO(3), \boldsymbol{t}_i \in \mathbb{R}^3)\}_1^m$ from them, where symbols $\mathbf{R}$ and $\boldsymbol{t}$ denote rotation and translation, respectively. PGO is essentially a directed graph problem, where $\mathbf{R}_{ij}$ is an edge and $\mathbf{R}_i$ and $\mathbf{R}_j$ are the nodes of the edge. The residual function is chosen as follows [67]:

$$r_i(\boldsymbol{\theta}) = \sqrt{\left\|\mathrm{Log}(\mathbf{R}_{ij}^T\mathbf{R}_i^T\mathbf{R}_j)\right\|_{\boldsymbol{\Lambda}_{ij}^{\mathbf{R}}}^2 + \left\|\mathbf{R}_{ij}^T(\boldsymbol{t}_{ij} - \mathbf{R}_i^T(\boldsymbol{t}_i - \boldsymbol{t}_j))\right\|_{\boldsymbol{\Lambda}_{ij}^{\boldsymbol{t}}}^2} \tag{8}$$

where $\boldsymbol{\Lambda}_{ij}^{\boldsymbol{R}}$ and $\boldsymbol{\Lambda}_{ij}^{\boldsymbol{T}}$ are the information matrices of rotation and translation, respectively. $\|\cdot\|_{\boldsymbol{\Lambda}}$ is the Mahalanobis norm of a matrix or a vector, e.g., $\|\boldsymbol{a}\|_{\boldsymbol{\Lambda}} = \boldsymbol{a}^{\mathrm{T}}\boldsymbol{\Lambda}\boldsymbol{a}$. $\mathrm{Log}(\cdot)$ is the logarithm map of the rotation group, which converts a rotation matrix to a 3-D vector. Fig. 2(h) plots an instance of PGO.

IV. Our Augmented MCC

As aforementioned, the limitations of MCC for robust fitting lie in twofold: the sensitivities to clustered outliers and high outlier ratios. The sensitivity to clustered outliers is a common limitation of robust estimation techniques, such as MCC, RANSAC, and M-estimation. Moreover, MCC can not deal with cases with high outlier ratios. First, kernel bandwidth is very important to the performance. MCC uses the Silverman's rule to determine the bandwidth, which is not accurate when the number of samples is not large enough. Under high outlier ratios, bandwidth estimation becomes more inaccurate. Second, the model estimation of MCC for geometric fitting is often converted to an M-like estimation for efficient optimization, which is only able to handle cases with low outlier ratios. To solve these problems, we introduce an accurate bandwidth estimation algorithm based on PDF matching, present an idea of GNC and worst-rejection strategy for MCC model estimation, and define a concept of LDM to tackle clustered outliers.

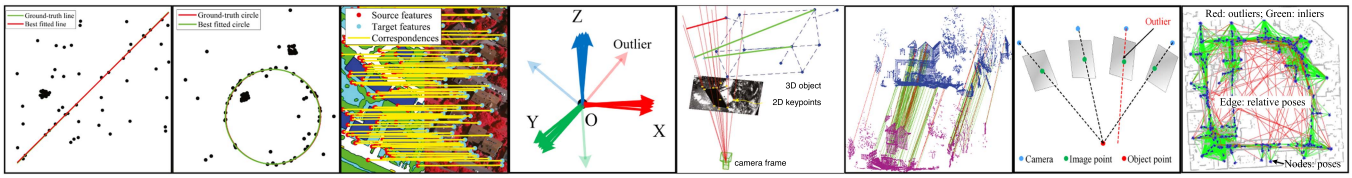


Fig. 2. Robust geometric fitting examples considered in this article that can be solved by our AMCC. (Examples 1–8). (a) LF. (b) CF. (c) IFM. (d) SRA. (e) PnP [6]. (f) PCR. (g) MT. (h) PGO [39].

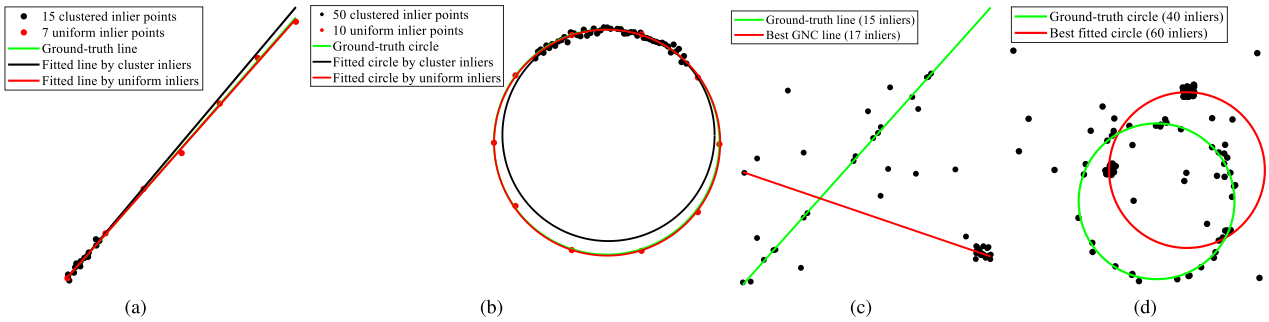


Fig. 3. Influence of inlier quality on geometry fitting. (a) and (b) Examples of LF and CF based on the least-squares without outliers, respectively. (c) and (d) Examples of LF and CF based on robust estimators with outliers, respectively. (a) LF. (b) CF. (c) Robust LF. (d) Robust CF.

TABLE I
DETAILED SETTINGS OF THE COMPARED ALGORITHMS (MNI REPRESENTS MAXIMUM NUMBER OF ITERATIONS)

Method	Parameters	Implementations
Cauchy M-estimation	Tuning constant: 2.385; $E_0 = +\infty$; MNI: 100.	MATLAB code; single thread http://rosa.unipr.it/fsda.html
MCC	$E_0 = +\infty$; MNI: 100.	MATLAB code; single thread https://ljy-rs.github.io/web
RANSAC	Sample size: 3; confidence: 0.99; MNI: 10^5.	MATLAB code; single thread https://www.peterkovesi.com/matlabfns/index.html
GNC-GM	$\mu^{(0)} = 5r_{max}^{(0)}$; GNC factor: 1.4; $E_0 = +\infty$; MNI: 100.	MATLAB code; single thread https://github.com/MIT-SPARK/GNC-and-ADAPT
MAGSAC++	Sample size: 3; confidence: 0.99; maximum threshold: 50; MNI: 10^5.	C++ code; single thread https://github.com/danini/magsac
Our AMCC	$s = 3$; $K = 20$; $M = 5$; $\tau = 1.4$; $N = 10$; $E_0 = +\infty$.	MATLAB code; single thread https://github.com/LJY-WHU/AMCC

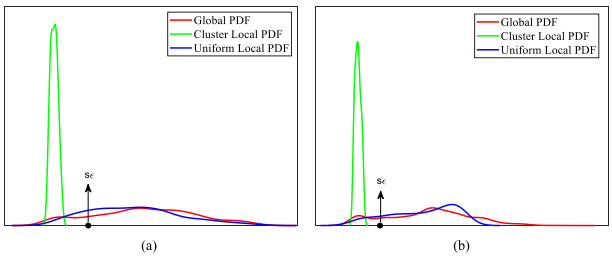


Fig. 4. Local PDFs of clustered observations and uniform observations, as well as the global PDF of all observations based on the LF and CF examples in Fig. 3. (a) PDFs of line fitting in Fig. 3(c). (b) PDFs of circle fitting in Fig. 3(d).

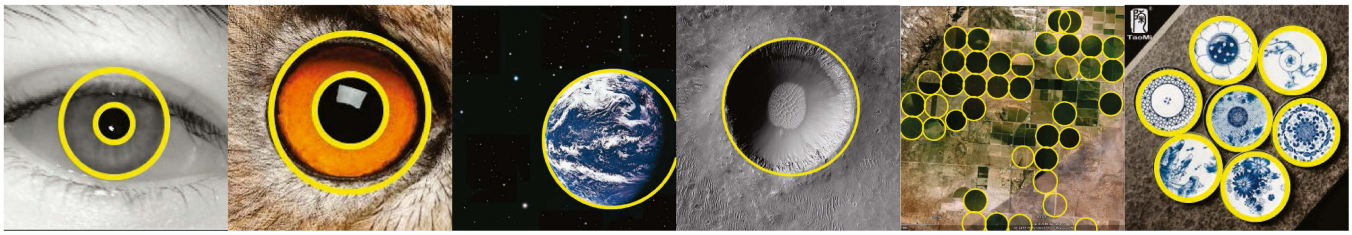
A. Bandwidth σ Estimation

As shown in [46], bandwidth and variable center can be obtained by optimizing the following problem based on the idea of PDF matching:

$$(\sigma^*, c^*) = \arg\min_{\sigma,c} \left\{ \int \left[k_\sigma\left(r-c\right)\right]^2 dr - 2\mathbb{E}\left[k_\sigma\left(r-c\right)\right] \right\} \quad (9)$$

where c is the variable center. In this article, we focus on the geometric fitting problems, where we only want to retain inliers, and the residual errors of inliers are usually distributed around zero. Therefore, a kernel function centered at zero is more in line with the expectation. We also use the Gaussian kernel in our AMCC, then, problem (9) becomes the following:

$$\begin{aligned}\sigma^* &= \arg\min_{\sigma} \left\{ \int \left[G_\sigma(r)\right]^2 dr - 2\mathbb{E}\left[G_\sigma(r)\right] \right\} \\ &= \arg\min_{\sigma} \left\{ \frac{1}{2\sqrt{\pi}\sigma} - \frac{\sqrt{2}}{\sqrt{\pi}\sigma n} \sum_{i=1}^{n} \exp\left(-\frac{r_i^2}{2\sigma^2}\right) \right\}. \end{aligned} \quad (10)$$



(a) (b) (c) (d) (e) (f)

Fig. 5. CF on six real images. (Example 2). (a) Human iris. (b) Owl iris. (c) Earth. (d) Meteor crater. (e) Irrigation system. (f) Plates.

TABLE II
SUCCESS RATE (↑) COMPARISON OF FOUR TYPICAL TASKS UNDER RANDOM OUTLIERS, WHERE LF, IFM, PNP, AND PCR REPRESENT LINE FITTING, IMAGE FEATURE MATCHING, PERSPECTIVE-N-POINTS, AND POINT CLOUD REGISTRATION, RESPECTIVELY

Task	Method	Outlier rate [%]					
		10	30	50	70	80	90
LF	Cauchy	100	100	100	16	8	3
	MCC	100	98	99	89	53	15
	RANSAC	100	100	100	100	100	99
	GNC-GM	100	100	100	98	83	21
	MAGSAC++	100	100	100	100	100	100
	Our AMCC	100	100	100	100	99	89
IFM	Cauchy	100	100	97	0	0	0
	MCC	100	100	100	87	45	5
	RANSAC	100	100	100	100	100	100
	GNC-GM	100	100	100	85	39	1
	MAGSAC++	100	100	100	100	100	100
	Our AMCC	100	100	100	99	85	42
PnP	Cauchy	100	100	90	0	0	0
	MCC	100	100	92	82	81	81
	RANSAC	100	100	100	100	100	100
	GNC-GM	100	99	97	89	68	11
	MAGSAC++	100	100	100	100	100	100
	Our AMCC	100	100	100	100	100	96
PCR	Cauchy	100	100	27	0	0	0
	MCC	100	100	91	90	89	91
	RANSAC	100	100	100	100	100	100
	GNC-GM	100	100	100	100	99	95
	MAGSAC++	100	100	100	100	100	100
	Our AMCC	100	100	100	100	100	99

TABLE III
SUCCESS RATE (↑) COMPARISON OF FOUR TYPICAL TASKS UNDER CLUSTERED OUTLIERS, WHERE LF, IFM, PNP, AND PCR REPRESENT LINE FITTING, IMAGE FEATURE MATCHING, PERSPECTIVE-N-POINTS, AND POINT CLOUD REGISTRATION, RESPECTIVELY

Task	Method	Outlier rate [%]					
		10	30	50	70	80	90
LF	Cauchy	100	82	7	7	4	2
	MCC	97	91	71	36	12	4
	RANSAC	100	100	86	59	36	13
	GNC-GM	100	85	21	25	14	10
	MAGSAC++	100	100	88	60	36	13
	Our AMCC	100	100	100	100	98	81
IFM	Cauchy	100	61	3	0	0	0
	MCC	100	94	64	34	17	2
	RANSAC	100	100	95	69	31	18
	GNC-GM	100	79	17	12	5	0
	MAGSAC++	100	99	68	42	20	8
	Our AMCC	100	100	100	98	98	89
PnP	Cauchy	98	83	37	0	0	0
	MCC	97	94	75	43	34	26
	RANSAC	100	100	92	48	30	18
	GNC-GM	97	94	86	70	53	13
	MAGSAC++	100	100	97	45	23	12
	Our AMCC	100	100	100	100	100	99
PCR	Cauchy	100	75	0	0	0	0
	MCC	99	99	64	47	40	25
	RANSAC	100	100	99	85	60	25
	GNC-GM	100	98	33	34	25	14
	MAGSAC++	100	100	99	68	37	13
	Our AMCC	100	100	100	100	100	99

To solve this problem, we let $\chi = \frac{1}{\sigma}$, obtaining

$$\chi^* = \arg\min_{\chi} \left\{ g(\chi) := \frac{\chi}{2\sqrt{\pi}} - \frac{\sqrt{2}\chi}{\sqrt{\pi}} \frac{1}{n} \sum_{i=1}^{n} \exp\left(-\frac{r_i^2}{2}\chi^2\right) \right\}. \tag{11}$$

We perform the Taylor expansion on $\exp(-\frac{r_i^2}{2}\chi^2)$, and only take its linear part.Substituting it into (11) yields a quadratic equation (see Appendix A for details) as follows:

$$\begin{aligned} &\arg\min_{\chi} g(\chi) \\ &\approx \arg\min_{\chi} \left\{ \sqrt{\tfrac{2}{\pi}} b\chi_0 \chi^2 + \left(\tfrac{1}{2\sqrt{\pi}} - \sqrt{\tfrac{2}{\pi}} b\chi_0^2 - \sqrt{\tfrac{2}{\pi}} a \right) \chi \right\} \end{aligned} \tag{12}$$

where $\chi_0 > 0$ is a known approximation of χ and

$$\begin{cases} a = \frac{1}{n} \sum_{i=1}^{n} \exp\left(-\frac{r_i^2}{2}\chi_0^2\right) \\ b = \frac{1}{n} \sum_{i=1}^{n} r_i^2 \exp\left(-\frac{r_i^2}{2}\chi_0^2\right) \end{cases}. \tag{13}$$

Obviously, $b \geq 0$. In practice, there must be noise or outliers in the observations, and it is impossible for all residuals to be 0. Therefore, we have $\sqrt{\frac{2}{\pi}} b\chi_0 > 0$ and the quadratic function has a unique solution as follows:

$$\chi^* = -\frac{\frac{1}{2\sqrt{2}} - b\chi_0^2 - a}{2b\chi_0}. \tag{14}$$

Since a good approximation χ_0 is generally unavailable, we use an iterative approach starting from $\frac{1}{\text{std}(\boldsymbol{r}^0)}$ to find the optimal χ^*.

B. Model θ Estimation

Let us start with two tools, i.e., the Black–Rangarajan duality [68] and GNC [15], which are used for outlier process and nonconvex optimization, respectively. The model estimation stage of our AMCC is built on top of them.

Tool 1 (Black–Rangarajan Duality [68]): Given a robust cost $\rho(\cdot)$, if there exists a function $\phi(z) \doteq \rho(\sqrt{z/s})$, where z is a intermediate variable and $s > 0$ is a scale factor, and its first and second partial derivatives $\phi'(z)$ and $\phi''(z)$ satisfy $\lim_{z\to 0} \phi'(z) = 1$, $\lim_{z\to\infty} \phi'(z) = 0$, and $\phi''(z) < 0$, then the robust estimation problem $\boldsymbol{\theta}^* = \arg\min_{\boldsymbol{\theta}} \frac{1}{n} \sum_{i=1}^{n} \rho(r_i)$ is

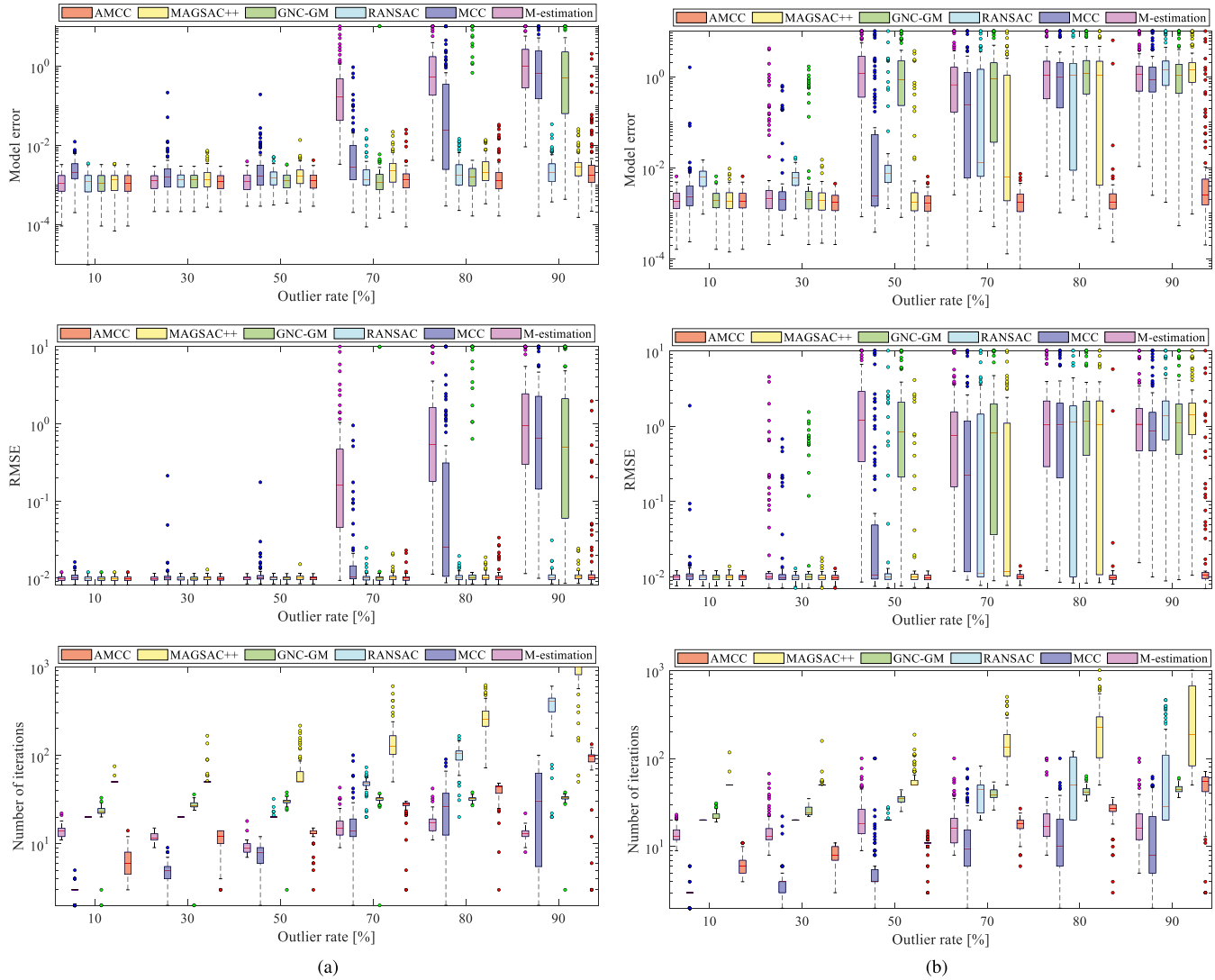


Fig. 6. LF results on simulated data, i.e., first row—model error, middle row—RMSE, and last row—number of iterations. (Example 1). (a) Random outliers. (b) Clustered outliers.

equivalent to the following outlier process:

$$\boldsymbol{\theta}^* = \underset{\boldsymbol{\theta}, w_i \in [0,1]}{\arg\min} \sum_{i=1}^{n} s w_i r_i^2 + \Psi_\rho(w_i) \tag{15}$$

where $w_i = \phi'(z_i) \in [0, 1]$ is the weight of an observation, and $\Psi_\rho(w)$ is an outlier process function. As suggested in [68], a straightforward mechanism for selecting this function is $\Psi_\rho(w) = \phi(z) - zw$.

Tool 2 (GNC [15]): Rather than directly optimizing a nonconvex cost $\rho(\cdot)$, GNC introduces a surrogate cost $\rho_\mu(\cdot)$ with a control parameter μ for optimization. This surrogate cost $\rho_\mu(\cdot)$ must satisfy two conditions: 1) There are some values of μ such that $\rho_\mu(\cdot)$ is a convex function, i.e., $\exists u$, $\rho''_\mu(\cdot) \geq 0$; 2) the original nonconvex cost $\rho(\cdot)$ must be a special case of $\rho_\mu(\cdot)$, i.e., $\rho(\cdot) \in \{\rho_\mu(\cdot)\}_u$. Generally, when μ goes to infinity or 1, $\rho_\mu(\cdot)$ recovers $\rho(\cdot)$. In the optimization, GNC starts from a convex surrogate to solve the problem globally, and then changes μ gradually to increase the amount of nonconvexity until the original cost is recovered. Since the GNC starts from a convex function and uses the current solution as the initialization of the subsequent iteration, it can largely avoid getting stuck in local optima.

Now, let us turn to the model estimation problem. Once the bandwidth is obtained, the model $\boldsymbol{\theta}$ optimization is as follows:

$$\begin{aligned}\boldsymbol{\theta}^* &= \underset{\boldsymbol{\theta}}{\arg\max} \frac{1}{n} \sum_{i=1}^{n} \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{r_i^2}{2\sigma^2}\right) \\ &= \underset{\boldsymbol{\theta}}{\arg\min} \frac{1}{n} \sum_{i=1}^{n} -\frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{r_i^2}{2\sigma^2}\right). \end{aligned} \tag{16}$$

Note that $\sigma > 0$ here is a known value. Since multiplying (16) by a constant greater than zero or adding a constant has no effect on the optimization results, we multiply it by $\sqrt{2\pi}\sigma^3$ and add a

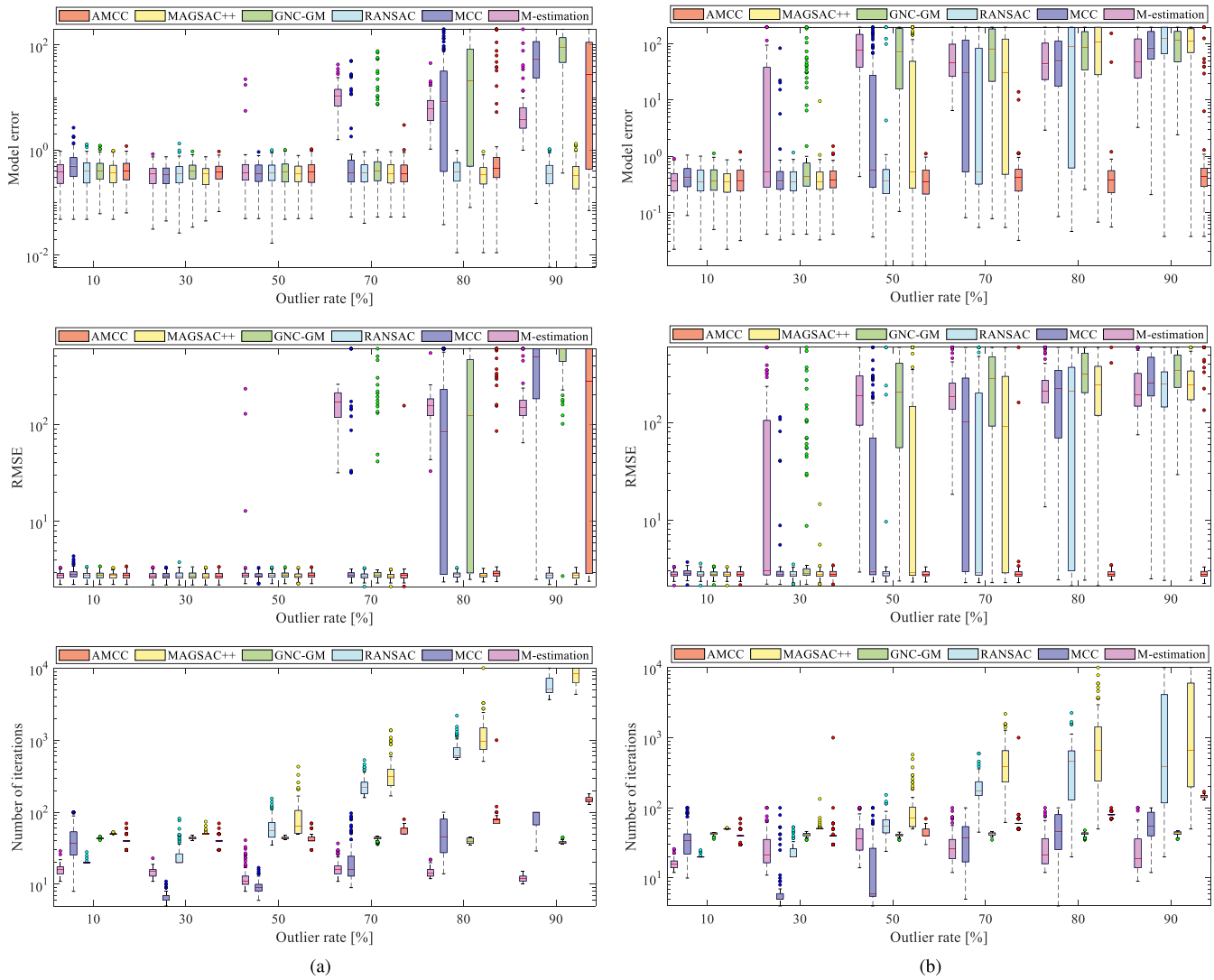


Fig. 7. IFM results on simulated data, i.e., first row—model error, middle row—RMSE, and last row—number of iterations. (Example 3). (a) Random outliers. (b) Clustered outliers.

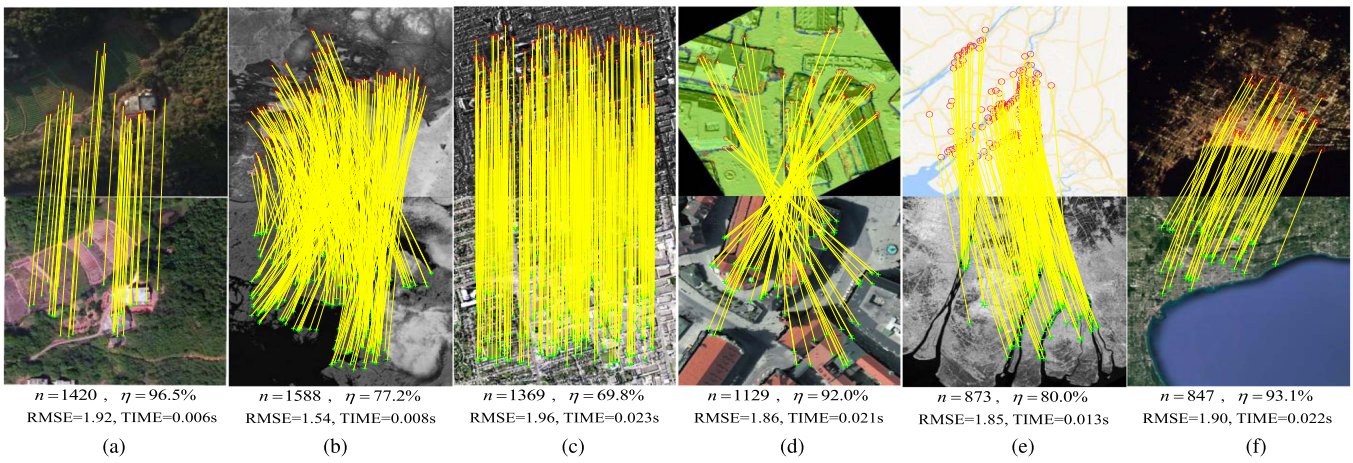
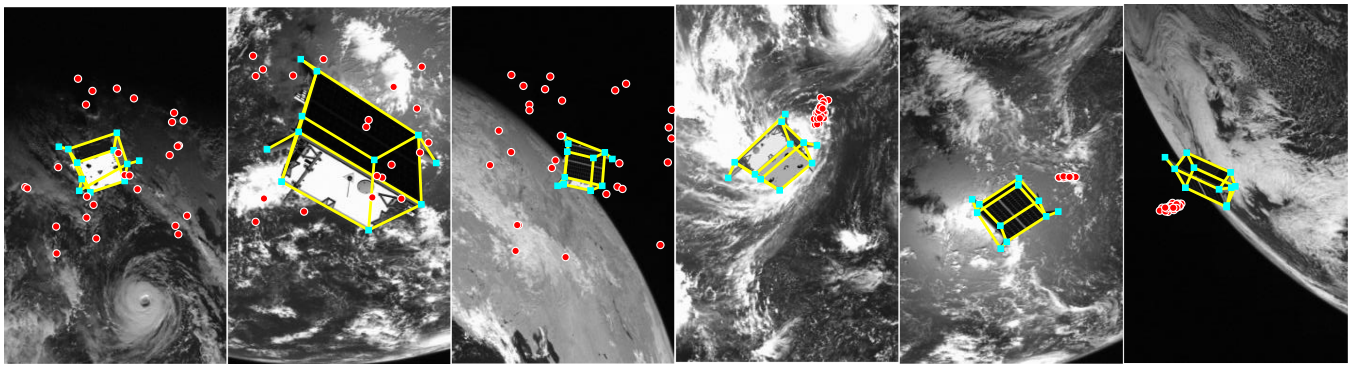


Fig. 8. IFM on a multimodal dataset with six types of images. Each line represents a correct correspondence. (Example 3). (a) Optical-optical. (b) Infrared-optical. (c) SAR-optical. (d) Depth-optical. (e) Map-optical. (f) Night-day.



27 outliers, $e_{model} = 0.05$ RMSE=1.38, TIME=0.11s (a)
20 outliers, $e_{model} = 0.005$ RMSE=2.92, TIME=0.14s (b)
28 outliers, $e_{model} = 0.04$ RMSE=1.07, TIME=0.16s (c)
29 outliers, $e_{model} = 0.02$ RMSE=2.01, TIME=0.15s (d)
15 outliers, $e_{model} = 0.05$ RMSE=1.79, TIME=0.16s (e)
24 outliers, $e_{model} = 0.03$ RMSE=2.32, TIME=0.21s (f)

Fig. 9. Satellite pose estimation on the SPEED dataset. The first three instances suffer from random outliers while the last three ones contain clustered outliers. Cyan squares represent inliers and red dots are outliers. (Example 5). (a) Random outliers. (b) Random outliers. (c) Random outliers. (d) Clustered outliers. (e) Clustered outliers. (f) Clustered outliers.

constant σ^2, giving as follows:

$$\boldsymbol{\theta}^* = \arg\min_{\boldsymbol{\theta}} \frac{1}{n}\sum_{i=1}^{n} \sigma^2 \left((1 - \exp\left(-\frac{r_i^2}{2\sigma^2}\right)\right). \tag{17}$$

This becomes an M-like estimation, where $\rho(r) = \sigma^2((1-\exp(-\frac{r^2}{2\sigma^2}))$. Furthermore, one can verify $\rho(r)$ satisfies $\lim_{|r|\to\infty}\rho'(r) = 0$, where $\rho'(r)$ is the first derivative of $\rho(r)$ with respect to r. Thus, it is a typical M-estimation. To improve the robustness of (17), we introduce a surrogate function of $\rho(r)$ based on the theory of GNC as follows:

$$\rho_\mu(r) = \mu\sigma^2 \left(1 - \exp\left(-\frac{r^2}{2\mu\sigma^2}\right)\right). \tag{18}$$

The function $\rho_\mu(r)$ is such that: 1) It is convex when $\mu \geq \frac{r^2}{\sigma^2}$ according to $\rho''_\mu(r) = (1 - \frac{r^2}{\mu\sigma^2})\exp(-\frac{r^2}{2\mu\sigma^2})$. When $\mu \to \infty$, $\lim_{\mu\to\infty}\rho'_\mu(r) = \lim_{\mu\to\infty} r\exp(-\frac{r^2}{2\mu\sigma^2}) = r$, and $\lim_{\mu\to\infty}\rho''_\mu(r) = 1$. This is exactly the same as the first and second derivatives of the least squares cost function. Hence, $\rho_\mu(\cdot)$ acts as the typical least-squares cost. 2) It becomes the original $\rho(r)$ when $\mu = 1$ [see Fig. 1(a)].

Although putting $\rho_\mu(r)$ into (17), it is still hard to solve due to its nonconvex nature. Fortunately, we can transfer the robust estimator to an outlier process based on the Black–Rangarajan duality. Let

$$\phi(z) \doteq \rho_\mu(\sqrt{2z}) = \mu\sigma^2\left(1 - \exp\left(-\frac{z}{\mu\sigma^2}\right)\right) \quad \text{s.t.} \quad z \geq 0 \tag{19}$$

we can verify that $\lim_{z\to 0}\phi'(z) = 1$, $\lim_{z\to+\infty}\phi'(z) = 0$, and $\phi''(z) < 0$. Therefore, problem (17) becomes (see Appendix B for details) the following:

$$\boldsymbol{\theta}^* = \arg\min_{\boldsymbol{\theta}, w_i\in[0,1]} \sum_{i=1}^{n} \frac{1}{2} w_i r_i^2 + \mu\sigma^2 w_i (\ln w_i - 1). \tag{20}$$

This problem can be efficiently solved by alternating optimization (see Appendix C for details). Specifically, at each inner iteration, the model parameters $\boldsymbol{\theta}$ are first optimized with fixed weights $\{w_i\}_1^n$, $\{w_i\}_1^n$ are then optimized with fixed $\boldsymbol{\theta}$, and finally, the control parameter μ is decreased by a factor τ.

Remark 1 (Robust to high outlier rate): Influence function $\psi(r) = \frac{\partial\rho(r)}{\partial r}$ can intuitively reflect the robustness of a cost function, which shows the impact of an observation on the energy cost. The influence function of our surrogate $\rho_\mu(r)$ is displayed in Fig. 1(b). As can be seen, when μ is small, large residuals have no effect on the energy cost, so it has strong robustness. However, once the initialization is not good, true inliers may obtain large residuals and be eliminated as outliers, causing the function to fall into local minima. On the contrary, when μ is big, large residuals still have impacts on the energy cost, so its robustness is limited. Fortunately, since its influence interval (the residual range that has a large impact on the energy) is large, it is not sensitive to the initializations and can effectively avoid local minima. Obviously, they are complementary. As μ decreases, the robustness keeps increasing, and the model estimation accuracy becomes more precise. Although the nonconvexity also keeps increasing, the result of the previous iteration can provide a good initialization for the next iteration, thereby reducing the possibility of getting stuck into local minima.

Worst-Rejection Strategy: Although the GNC strategy improves robustness, the true outlier rate of the observations remains the same in each iteration. Intuitively, if we can discard some true outliers during the iterative process, the true outlier rate will become lower, which can also offset the increase in nonconvexity to some extent and make the problem simpler. In our method, we believe that observations with the largest residuals are most likely to be true outliers. Hence, we sort residuals in a decreasing manner and reject the first M observations via a bool vector $\boldsymbol{h}$ in each iteration t.

Remark 2 (Differences from existing GNC methods): Using the Black–Rangarajan duality and GNC to improve the robustness of M-estimation has been investigated in [15], [16], and [20]. Differently, the research object of this article is not M-estimation but MCC. Although MCC is essentially an

M-estimation, it analyzes the residual errors from a probabilistic perspective. Therefore, MCC is suitable for any noise environment [18]. Meanwhile, the bandwidth of the kernel in MCC has a clear physical meaning, which controls the shape of the kernel function and can be estimated adaptively. One of the main insights is that we can use the close relationship between MCC and M-estimation to improve the robustness of MCC. To the best of the authors' knowledge, the proposed AMCC is the first MCC variant that can stably deal with high outlier ratios and clustered outliers.

C. Local Distribution Measure

Classical robust estimation techniques usually only consider the number of inliers, while ignoring the quality of inliers. For example, RANSAC-family and maximum consensus algorithms always use the number of inliers as the cost function. So, one wonders if the quality of each inlier is the same and equally important to the real model we are looking for.

Let us first look at two examples in Fig. 3(a) and (b), i.e, LF and CF based on the least-squares. The observations in these two examples suffer only from noise and are free of outliers. As can be seen, although the number of clustered inlier observations is much more than that of uniformly distributed inlier observations, the fitting accuracy of clustered inliers is far less than that of uniformly distributed ones. Therefore, inliers actually have quality problems, and the most important factor that affects the quality is their distribution characteristics. If we ignore the quality of inliers, then the situations in Fig. 3(c) and (d) will occur, that is, classical robust estimation methods (e.g., GNC, RANSAC, etc.) usually find the wrong model when there are clustered outliers in the observations.

In this section, we present a metric to describe the density of each observation called LDM, whose definition is as follows:

Definition 1 (LDM): Given a set of observations $\{\boldsymbol{x}_i\}_1^n$, suppose that the global PDF satisfied by the pairwise distances between observations is $f^g(D)$, and the local PDF of the pairwise distances between observations in the local area of an observation $\boldsymbol{x}_i$ is $f_i^l(D)$. Then, the LDM C_i of $\boldsymbol{x}_i$ is defined as the ratio of the local probability to the global probability that the distance is on the interval $[0, s\varepsilon]$

$$C_i = \frac{\int_0^{s\varepsilon} f_i^l(D)dD}{\int_0^{s\varepsilon} f^g(D)dD} \tag{21}$$

where ε is an inlier threshold, s is a scale factor. $s\varepsilon$ is a small value, and the probability on $[0, s\varepsilon]$ reflects the degree of clusterness of observations.

Fig. 4 shows examples of local and global PDFs. As can be seen, the PDF of uniform observations is very close to the global PDF, so its LDM is close to 1. In contrast, the PDF of clustered observations has a high peak and a narrow bandwidth, which is far from the global PDF, and its LDM is much larger than 1. From the definition, we can see that the key of LDM is to calculate the global PDF $f^g(D)$ and local PDF $f_i^l(D)$. A straightforward way is to use the kernel density estimation (KDE) to fit its corresponding PDF. However, KDE is relatively slow and will greatly reduce the efficiency of our AMCC. Instead, motivated by the idea of histogram, we direct calculate the probability based on the discrete distances between observations as follows:

$$\begin{aligned}\int_0^{s\varepsilon} f^g(D)dD &\approx \frac{1}{n}\sum_1^n \lfloor D_i \le s\varepsilon \rfloor \\ &= \frac{2}{n(n-1)}\sum_{i=1}^{n}\sum_{j=i+1}^{n} \lfloor \|\boldsymbol{x}_i - \boldsymbol{x}_j\|_2 \le s\varepsilon \rfloor \end{aligned} \tag{22}$$

Algorithm 1: Augmented MCC (LF example).

Input: Observations $\mathcal{O} = \{\boldsymbol{x}_i = (x_i, y_i) \in \mathbb{R}^2\}_1^n$.
Output: Optimal model $\boldsymbol{\theta}^* = (k, m)$.

// $s\varepsilon$: s times of inlier threshold;
// K: number of neighbors;
// M: number of rejected points;
// τ: step-size for bandwidth update;
// N: iteration number of inner loop;
// ϵ: convergence threshold;
// E_0: initial energy cost;

Initialization: $s\varepsilon$; K; M; τ; N; ϵ; E_0; $t=1$
// Local distribution measure weights;
$\boldsymbol{w'} = \mathit{ldm\_weights}(\mathcal{O}, K, s\varepsilon)$;
// Weighted least-squares line fitting;
$\boldsymbol{\theta}^{(0)} = \mathit{linefit\_wls}(\mathcal{O}, \boldsymbol{w'})$;
// Residual calculation;
$\boldsymbol{r}^{(0)} = \mathit{residual\_function}(\mathcal{O}, \boldsymbol{\theta}^{(0)})$;
$\boldsymbol{h}^{(0)} = \mathit{ones}(n, 1)$;

while *TRUE* **do**
  $t_{\text{start}} = t - 1$;
  // Kernel bandwidth estimation;
  $\sigma^{(t)} = \sigma\_\mathit{update}(\boldsymbol{r}^{(t-1)})$;
  $\sigma_{start} = \sigma^{(t)}$;
  while $t - t_{\text{start}} \le L$ **do**
    // Weights: $\tilde{\boldsymbol{w}}^{(t)} = \boldsymbol{w}^{(t-1)} \circ \boldsymbol{w'} \circ \boldsymbol{h}^{(t-1)}$;
    $\tilde{\boldsymbol{w}}^{(t)} = \boldsymbol{w}\_\mathit{update}(\boldsymbol{r}^{(t-1)}, \sigma^{(t)}, \boldsymbol{w'}, \boldsymbol{h}^{(t-1)})$;
    $\boldsymbol{\theta}^{(t)} = \mathit{linefit\_wls}(\mathcal{O}, \tilde{\boldsymbol{w}}^{(t)})$;
    $\boldsymbol{r}^{(t)} = \mathit{residual\_function}(\mathcal{O}, \boldsymbol{\theta}^{(t)})$;
    // Worst rejection strategy;
    $\boldsymbol{h}^{(t)} = \mathit{worst\_rejection}(\boldsymbol{r}^{(t)}, M)$;
    // Energy cost update;
    $E = \tilde{\boldsymbol{w}}^{(t)} \circ \boldsymbol{r}^{(t)} \circ \boldsymbol{r}^{(t)}$;
    // Check convergence
    if $|E - E_0| < \epsilon$ **then** break
    // Kernel bandwidth update;
    $\sigma^{(t+1)} = \sigma^{(t)}/\tau$;
    $E_0 = E$; $t = t + 1$;
  end
  // Check convergence
  if $|\sigma_{start} - \sigma_0| < \epsilon$ **then return** $\boldsymbol{\theta}^{(t)}$
  $\sigma_0 = \sigma_{start}$; $t = t + 1$;
end

where $\lfloor x \rfloor$ returns 1 if condition x is true, otherwise, returns 0; $\|\cdot\|_2$ is the l_2-norm. For local probability, we first search K neighbors of each $\boldsymbol{x}_i$, and then compute the probability in the same way as (22). Then, we assign each observation a Gaussian-like weight based on the LDM as follows:

$$w_i' = \exp\left(-\frac{C_i^2}{2(\text{std}\,(\boldsymbol{C}))^2}\right) \tag{23}$$

where $\boldsymbol{C} = \{C_i\}_1^n$.

Remark 3 (Relation between LDM and data selection): Actually, LDM is also a method for selecting uniformly distributed data. The difference is that we do not select data directly, but score the data (weights range from 0 to 1), which is a soft selection method. In contrast, the commonly used method is a hard selection method, that is, giving the data a score of 1 (retain) or 0 (discard). Hard methods may lose a lot of useful information and cause bias. For example, cluster inliers may be completely discarded and some uniform inliers may not be selected.

D. Main Algorithm

Our proposed AMCC is summarized in *Algorithm 1*, where we take the LF task as an example. Given a set of observations $\mathcal{O} = \{\boldsymbol{x}_i = (x_i, y_i) \in \mathbb{R}^2\}_1^n$, AMCC first computes a LDM weight to measure the clusterness of each observation. Then, it performs the main framework of two-layer loops. In the outer loop, AMCC estimates the Gaussian kernel bandwidth. The inner loop conducts a GNC-like algorithm, which mainly consists of four steps, i.e., weight update (contains three components, including LDM weight, residual weight, and worst rejection binary weight), model update, worst rejection, and bandwidth update. It can be seen that we directly decay the kernel bandwidth σ without introducing an additional μ. The reason for this is that our σ*_update* function estimates the bandwidth of all observation residuals (including outliers), so its value is very large compared to the bandwidth corresponding to the correct observations, which imply contains the parameter μ.

V. Experiments and Evaluations

In this section, we first qualitatively and quantitatively evaluate the proposed AMCC algorithm on five examples provided in Section III-C using simulated and real data. We consider both random outliers and clustered outliers in our experiments. Then, we conduct an ablation study to demonstrate the effectiveness of our several contributions.

Baselines: We choose three well-known algorithms (i.e., Cauchy M-estimation, MCC, and RANSAC) and two state-of-the-arts (i.e., GNC-GM [16] and MAGSAC++ [33]) as baselines for comparison. For fairness, we use the official implementation code of these methods and their recommended parameter settings. Details are summarized in Table I. For our method, we set the number of neighbors $K = 20$ and scale $s = 1$ for LDM calculation, the number of rejected observations $M = 5$ for worst rejection, and the GNC factor $\tau = 1.4$.

Evaluation Metrics: We use four metrics for quantitative evaluation: 1) the model error e_{model} of an estimated $\boldsymbol{\theta}^*$ compared to the ground truth one $\boldsymbol{\theta}^\circ$, i.e., $e_{\text{model}} = \|\boldsymbol{\theta}^* - \boldsymbol{\theta}^\circ\|$, where we use l_2 norm for vectors and Frobenius norm for matrix; 2) the root mean square error (RMSE) of inlier observations; 3) the number of iterations that an algorithm costs, which reflects the runtime of a method; and 4) success rate describes the number of successful estimates out of 100 Monte Carlo runs, where an estimate is considered successful when its RMSE is less than 3 times the noise level.

Hardware: All the experiments are conducted on a laptop with i7-8550U CPU @ 1.8GHz and 8GB of RAM.

A. Line/Circle Fitting

Setup: At each Monte Carlo run, a set of 1-D points $\{x_i\}_1^{n_{\text{in}}}$ that follows a normal distribution $\mathcal{N}(0, 1)$ and a ground truth 2-D line model $\boldsymbol{\theta}^\circ = (k^\circ, m^\circ)$ are first randomly generated, where $n_{\text{in}} = (1-\eta)n$ is the number of inlier observations with a given outlier rate η; then, inliers are produced by $y_i = k^\circ x_i + m^\circ + \epsilon_i$, where ϵ_i represents Gaussian inlier noise $\epsilon_i \sim \mathcal{N}(0, \sigma_{\text{noise}}^2)$ with $\sigma_{\text{noise}} = 0.01$; outliers are 2-D random points that have no relation to the ground truth model. For *random outliers* setup, we use a Gaussian random generator $\mathcal{N}(\mathbf{0}, \mathbf{I}_2)$ to produce outliers, where $\mathbf{I}_2$ is a 2×2 identity matrix; for *clustered outliers* setup, we randomly generate $T \in \{1, 2, 3\}$ 2-D cluster centers $\{\hat{\boldsymbol{c}}_i\}_1^T$ and produce clustered outliers at each center by $\mathcal{N}(\hat{\boldsymbol{c}}_i, \sigma_{\hat{\boldsymbol{c}}_i}^2 \mathbf{I}_2)$, where $\sigma_{\hat{\boldsymbol{c}}_i} \in [\sigma_{\text{noise}}, 10\sigma_{\text{noise}}]$. In our experiments, we fix n_{in} to be 50, sweep η from 10% to 90%, and perform 100 Monte Carlo runs for each configuration.

Random outlier results: Fig. 6(a) plots the comparison metrics, from which we can make several observations. 1) M-estimation and MCC perform well only on low outlier ratios. Once the outlier ratio exceeds 50%, M-estimation fails completely. MCC is slightly better than M-estimation, but there are still many failures. Moreover, MCC may also fail in the case of low outlier ratios, e.g., 30%, which indicates that MCC is not as stable as M-estimation. 2) The idea of GNC can effectively improve the robustness of M-estimation. Traditional GM estimator can only handle up to 50% of outliers while GNC-GM can deal with 70%–80% of outliers. 3) RANSAC-type methods (RANSAC and MAGSAC++) have the highest robustness; however, the computational complexity of RANSAC increases exponentially with the increase of the outlier ratio. For example, MAGSAC++ requires an order of magnitude more iterations than our AMCC at an outlier ratio of 70%. 4) Our proposed AMCC solves the instability problem of MCC and raises the robustness of classic MCC to a new level. AMCC can stably handle 80% of outliers, and even at an outlier ratio of 90%, its success rate is still around 90%, which is better than the GNC-GM (see Table II for details).

Remark 4 (AMCC is not as robust as RANSAC): Statistical robust estimators (M-estimations and MCCs) and RANSAC family are two different types of methods with different advantages and disadvantages. With random outliers, as long as the number of iterations of RANSAC is sufficient, it can always find an approximation to the correct solution no matter how high the outlier ratio is. Hence, the robustness of any existing statistical estimators is inferior to the RANSAC family in the case of a

super-high outlier rate (random outliers). This is the bottleneck of the entire field of statistical estimator-based robust fitting, and it is not the goal of this article to propose a method that outperforms the RANSAC family in terms of robustness under random outliers. Instead, we aim to improve the outlier-resistant performance of traditional MCC from 50% to a level close to that of RANSAC, while speeding up by $1-2$ orders of magnitude compared to RANSAC.

Clustered outlier results: From Fig. 6(b), we can see that: 1) All the five compared baselines are very sensitive to clustered outliers. M-estimation, MCC, and GNC-GM can only handle less than 50% of outliers. Once the outlier ratio exceeds 50%, RANSAC-type methods become very unstable, since their costs maximize the number of inliers while ignoring the quality of inliers. 2) Our AMCC is suitable for both random outliers and clustered outliers. It still achieves a success rate of 98% at 80% of clustered outliers, which is a 62 percentage point improvement over the second-ranked method (see Table III for details).

Real CF results: We select six real images with circular objects (human iris, owl iris, Earth, meteor crater, irrigation system, and plates) to show our fitting performance. For images with a single circular object, we use image binarization, morphological operations, and edge detection techniques to extract the data points of the circle. Obviously, these points contains many outliers. For images with multiple circles, we first locate each circle object using a circle detector, then take the same approach as above to extract data points for each circle. The results are displayed in Fig. 5. As can be seen, our AMCC achieves accurate CF.

B. Image Feature Matching

Setup: We first randomly sample a set of 2-D points $\{\boldsymbol{x}_i\}_1^n$ using a normal generator $\mathcal{N}(\mathbf{0}, 100^2\mathbf{I}_2)$. Then, we generate a random ground truth model $\boldsymbol{\theta}^\circ = (\mathbf{A}^\circ, \boldsymbol{t}^\circ)$, where $\mathbf{A}^\circ = \boldsymbol{s}^\circ\mathbf{R}^\circ$, $\boldsymbol{s}^\circ \in [0.5, 2]$ is a 2×1 scale vector which represents the scales in the x-axis and y-axis, $\mathbf{R}^\circ$ is a 2-D rotation matrix whose angle belongs to $[-\pi/2, \pi/2]$, and $\boldsymbol{t}^\circ \in [-100, 100]$ is a 2×1 translation vector. Using $\boldsymbol{\theta}^\circ$, we produce observations $\{\boldsymbol{y}_i\}_1^n$ by $\boldsymbol{y}_i = \mathbf{A}^\circ\boldsymbol{x}_i + \boldsymbol{t}^\circ + \boldsymbol{\epsilon}_i$ if $\boldsymbol{y}_i$ is an inlier, where $\boldsymbol{\epsilon}_i \sim \mathcal{N}(\mathbf{0}, 2^2\mathbf{I}_2)$ is a 2×1 noise vector; or by $\boldsymbol{y}_i \sim \mathcal{N}(\mathbf{0}, 100^2\mathbf{I}_2)$ if $\boldsymbol{y}_i$ is a random outlier; or by $\boldsymbol{y}_i \sim \mathcal{N}(\hat{\boldsymbol{c}}_i, \sigma_{\hat{\boldsymbol{c}}_i}^2\mathbf{I}_2)$ if $\boldsymbol{y}_i$ is a clustered outlier. Again, we fix n_{in} to be 50 and sweep η from 10% to 90%. We also perform 100 Monte Carlo runs for each configuration.

Random outlier results: Fig. 7(a) shows the evaluation results, from which we can make several observations. 1) Although MCC is essentially an M-estimation, it has better robustness than traditional M-estimations such as the Cauchy estimate. The reason is that MCC studies the residuals from a probabilistic perspective and estimates the kernel bandwidth adaptively. 2) Our AMCC is robust to 80% of outliers, which is better than Cauchy M-estimation (robust to 50% of outliers), typical MCC (robust to 70% of outliers), and GNC-GM (robust to 70% of outliers). 3) Comparing Fig. 7(a) with Fig. 6(a), we can find that RANSAC-type methods requires almost an order of magnitude more iterations in the IFM task than that of the LF task. Actually, the computational complexity of RANSAC-type methods is exponential to the size of the minimum sampling sets, which is determined by their theoretical basis. Therefore, RANSAC-type methods can hardly be applied to high-dimensional problems, such as face recognition and PGO. 4) Our AMCC only requires no more than 100 iterations and is not closely related to the model dimension. In fact, MCC-type methods have been widely used in high-dimensional problems.

Clustered outlier results: From Fig. 7(b), we can see that our AMCC is able to handle cases with up to 90% of clustered outliers, which is even better than the cases of random outliers. In contrast, other compared baselines can only deal with up to 50% of outliers, among which Cauchy estimate suffers from many failures even at an outlier ratio of 30%. From Table III, the success rates of Cauchy M-estimation, MCC, RANSAC, GNC-GM, MAGSAC++, and our AMCC at an outlier ratio of 80% are 0%, 17%, 31%, 5%, 20%, and 98%, respectively. Our success rate is 67 percentage points higher than the second-ranked method.

Real multimodal image matching: To showcase the excellent performance of AMCC on real data, we perform IFM on a multimodal dataset consisting of six types of image pairs (i.e., optical-optical, infraredoptical, synthetic aperture radar (SAR)-optical, depth-optical, map-optical, and night-day). We use the RIFT [61] or locally normalized image feature transform [69] to establish putative feature correspondences. The ground truth affine transformation for each pair is established by least-squares fitting with ten manually picked correspondences. It is very challenging to match these multimodal images due to severe nonlinear radiation differences. Therefore, these matching pairs have a high outlier ratio from 69.8% to 96.5%. From the results in Fig. 8, we can observe that AMCC achieves good results on all six pairs. Its RMSE is lower than 2 pixels under a 3-pixels inlier threshold and it only takes about 0.02 s to deal with 1000 correspondences.

C. Perspective-n-Points

Setup: Given a camera internal parameter matrix $\mathbf{K}$, we first randomly sample a set of 3-D image space points $\{\boldsymbol{X}_i^c\}_1^n$ in a box [-10, 10] × [-10, 10] × [1, 2]. Then, we randomly generate a 3-D rotation matrix $\mathbf{R}^\circ$ whose angles belong to $[-\pi/2, \pi/2]$, and the 3-D translation vector $\boldsymbol{t}^\circ$ is the mean of $\{\boldsymbol{X}_i^c\}_1^n$. The corresponding object space point of $\boldsymbol{X}_i^c$ is obtained via transformation $\boldsymbol{X}_i = (\mathbf{R}^\circ)^{-1}(\boldsymbol{X}_i^c - \boldsymbol{t}^\circ)$. Using $\boldsymbol{\theta}^\circ = (\mathbf{R}^\circ, \boldsymbol{t}^\circ)$, 3-D object point $\boldsymbol{X}_i$ is projected onto an image to get its ground truth 2-D image correspondence $\boldsymbol{x}_i^\circ$ via $d_i[\boldsymbol{x}_i^\circ \quad 1]^T = \boldsymbol{K}\boldsymbol{X}_i^c$, where d_i is the depth. Finally, observations $\{\boldsymbol{x}_i\}_1^n$ is generated by $\boldsymbol{x}_i = \boldsymbol{x}_i^\circ + \boldsymbol{\epsilon}_i$, where $\boldsymbol{\epsilon}_i \sim \mathcal{N}(\mathbf{0}, 2^2\mathbf{I}_2)$ if $\boldsymbol{x}_i$ is an inlier, or $\boldsymbol{\epsilon}_i \sim \mathcal{N}(\mathbf{0}, 1000^2\mathbf{I}_2)$ if $\boldsymbol{x}_i$ is a random outlier, or by $\boldsymbol{x}_i \sim \mathcal{N}(\hat{\boldsymbol{c}}_i, \sigma_{\hat{\boldsymbol{c}}_i}^2\mathbf{I}_2)$ if $\boldsymbol{x}_i$ is a clustered outlier. We use the unit quaternion to represent the rotation and apply the Gauss–Newton method for PnP optimization, where the initialization for the model $\boldsymbol{\theta}$ is required. Therefore, we initialize $\boldsymbol{t}$ by generating a translation vector within $[50\% \times \boldsymbol{t}^\circ, 150\% \times \boldsymbol{t}^\circ]$ and initialize $\mathbf{R}$ by adding a random disturbance within $[-\pi/6, \pi/6]$ to the ground truth angles. For a fair comparison, the same initialization is applied for all compared methods. In this task, we set n_{in} to be 100.

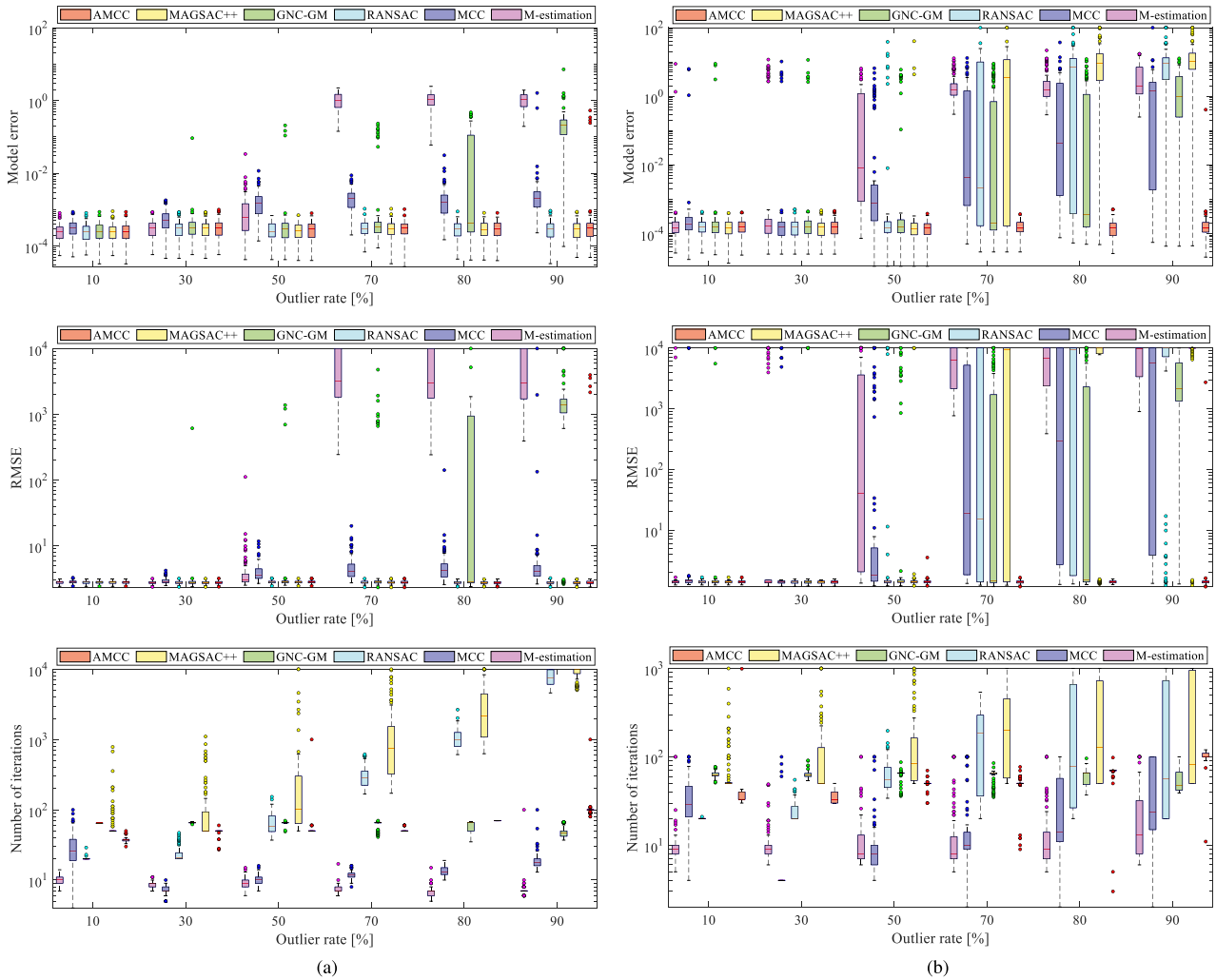


Fig. 10. PnP results on simulated data, i.e., first row – model error, middle row – RMSE, last row – number of iterations. (Example 5). (a) Random outliers. (b) Clustered outliers.

Random outlier results: Fig. 10(a) shows the evaluation results, from which we can make several observations. 1) Our AMCC is robust to 90% of outliers with appropriate initializations, which is almost comparable to RANSAC-type methods. AMCC is a local optimal method, while RANSAC-type methods only find approximate solutions. Thus, AMCC generally has a higher model fitting accuracy than RANSAC-type methods. In our experimental results, RANSAC-type methods are as accurate as our AMCC, since their results are refined by an M-like estimation and the inlier threshold ($3\sigma_{\text{noise}}$) is very precise. 2) Although MCC performs better than Cauchy M-estimation and GNC-GM at high outlier ratios, it is not stable even in cases with low outlier ratios. Our AMCC overcomes this limitation, i.e., its success rate is always 100% when $\eta \leq 80\%$ (see Table II for details). 3) Even at an outlier ratio of 90%, our AMCC requires no more than 100 iterations, which is two orders of magnitude smaller than the ones of RANSAC-type methods.

Clustered outlier results: Fig. 10(b) plots the results of clustered outliers. As can be seen, our AMCC is far superior to other compared methods. From Table III, despite a high outlier ratio of 90%, our success rate still reaches 99%, which is 73 percentage points higher than the second-ranked MCC method.

Satellite pose estimation results: We estimate the 6-DOF pose of satellites on the SPEED dataset [70], where the 11 2-D-3-D correspondences for each image are provided by the pipeline described in [5]. To show the robustness of AMCC, we spoil these observations by adding 15–30 random or clustered outliers. Fig. 9 provides six examples, where our AMCC achieves accurate localization of the satellites within 0.3 s. The model error is better than 0.05 and the RMSE of reprojection errors is smaller than 3 pixels.

D. Point Cloud Registration

Setup: We first randomly sample a set of 3-D points $\{\boldsymbol{X}_i\}_1^n$ in a box [-100, 100] × [-100, 100] × [-100, 100]. Then we generate a random ground truth 6-DOF rigid model $\boldsymbol{\theta}^\circ = (\mathbf{R}^\circ, \boldsymbol{t}^\circ)$, where the angles of $\mathbf{R}^\circ$ are within $[-\pi/2, \pi/2]$ and the translations of $\boldsymbol{t}^\circ$ are within $[-100, 100]$. Using $\boldsymbol{\theta}^\circ$, we generate 3-D correspondence observations $\{\boldsymbol{Y}_i\}_1^n$ by $\boldsymbol{Y}_i = \mathbf{R}^\circ \boldsymbol{X}_i + \boldsymbol{t}_i^\circ + \boldsymbol{\epsilon}_i$ if $\boldsymbol{Y}_i$ is an inlier, where $\boldsymbol{\epsilon}_i \sim \mathcal{N}(\mathbf{0}, 0.3^2\mathbf{I}_3)$ is a 3×1 noise

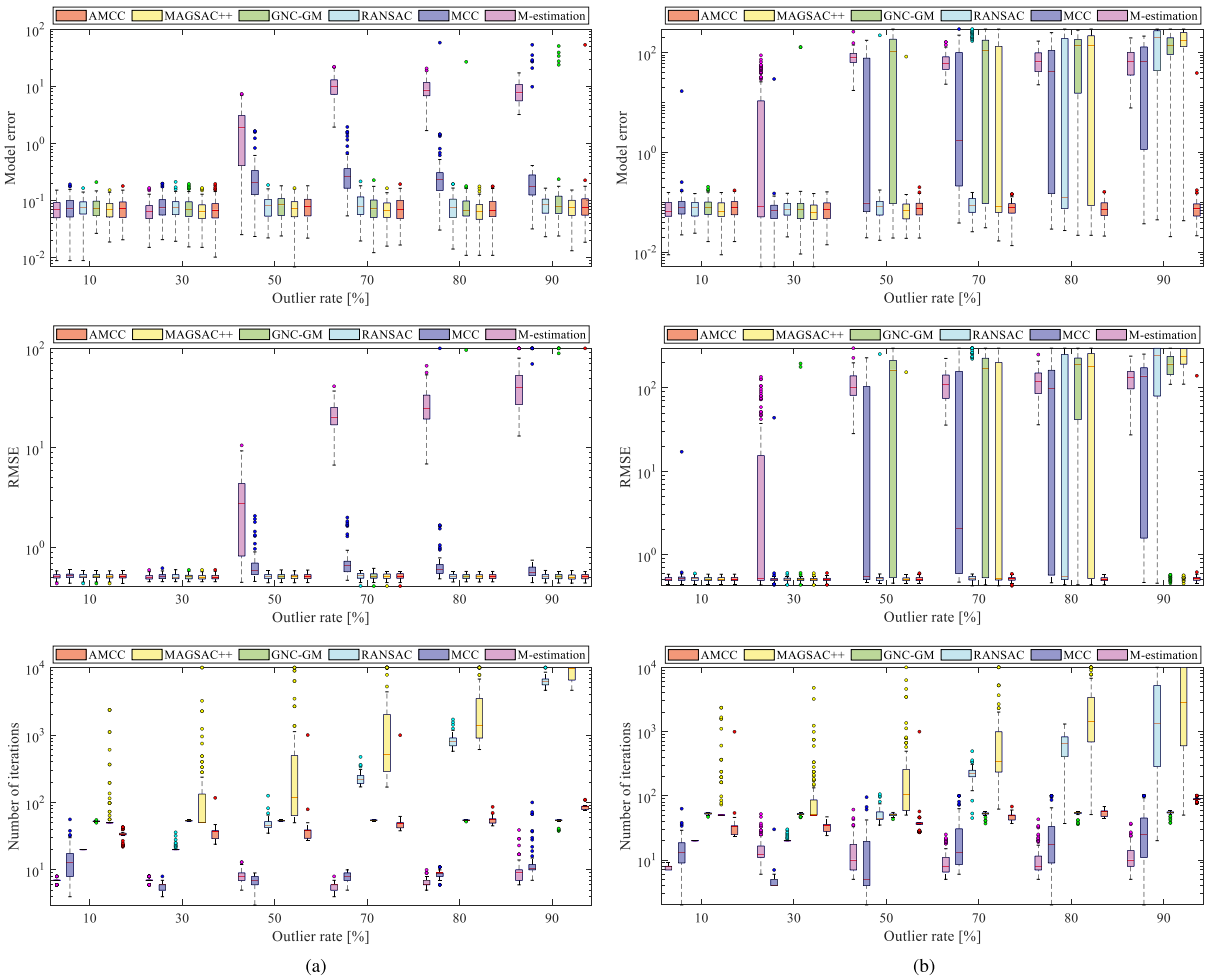


Fig. 11. Point cloud registration results on simulated data, i.e., first row – model error, middle row – RMSE, last row – number of iterations. (Example 6). (a) Random outliers. (b) Clustered outliers.

vector; or by $\boldsymbol{\epsilon}_i \sim \mathcal{U}(\Phi)$ if $\boldsymbol{Y}_i$ is a random outlier, where region $\Phi = \{\boldsymbol{Y}_i = (X_i, Y_i, Z_i) | X_i, Y_i, Z_i \in [-100, 100]\}$; or by $\boldsymbol{Y}_i \sim \mathcal{N}(\hat{\boldsymbol{c}}_i, \sigma^2_{\hat{\boldsymbol{c}}_i} \mathbf{I}_3)$ if $\boldsymbol{Y}_i$ is a clustered outlier. We fix n_{in} to be 50 and randomly sample $T \in \{1, 2, 3\}$ clusters.

Random outlier results: Fig. 11(a) shows the evaluation metrics, from which we can make several observations. 1) GNC-GM and MCC perform better on rigid fitting than other tasks (e.g., LF and IFM), while Cauchy M-estimation performs much worse. Cauchy M-estimation has many failures at an outlier ratio of 50%. 2) Our AMCC achieves almost the same robustness as that of RANSAC-type methods. AMCC achieves a success rate of 99% under 90% of outliers, which is better than GNC-GM and MCC (see Table II for details).

Clustered outlier results: Fig. 11(b) plots the results of clustered outliers. As can be seen, our AMCC is far superior to other compared methods. From Table III, the success rates of Cauchy M-estimation, MCC, RANSAC, GNC-GM, MAGSAC++, and our AMCC at an outlier ratio of 90% are 0%, 25%, 25%, 14%, 13%, and 99%, respectively. Our success rate is 74 percentage points higher than the second-ranked method.

Scan matching results: We pick six scan pairs from two open datasets with ground truth transformations, i.e., the small-scale FGR dataset [20] and large-scale ETH LIDAR dataset,[1] for point cloud registration. For each scan pair, we first apply the VoxelGrid filter to downsample the original point clouds, in which the downsampling resolutions for the FGR and ETH datasets are 0.01 m and 0.1 m, respectively. Then, we use the intrinsic shape signatures (ISS) [71] and FPFH [62] to establish 3-D correspondences. Finally, a support line voting strategy [72] is used to filter out a portion of outliers. The results are displayed in Fig. 12. As shown, AMCC only takes about 0.15 s to achieve an accurate registration with more than 80% of outliers. Its model error is better than 0.03 and its RMSE is lower than the downsampling resolution.

E. Running Time

In the previous experiments, we reported the number of iterations required by different methods. To show the efficiency of each method more intuitively, we provide the running time information of each method based on two tasks, i.e., IFM and point cloud registration. The results are displayed in Fig. 13,

[1][Online]. Available: http://www.prs.igp.ethz.ch/research/completed_projects/automatic_registration_of_point_clouds.html

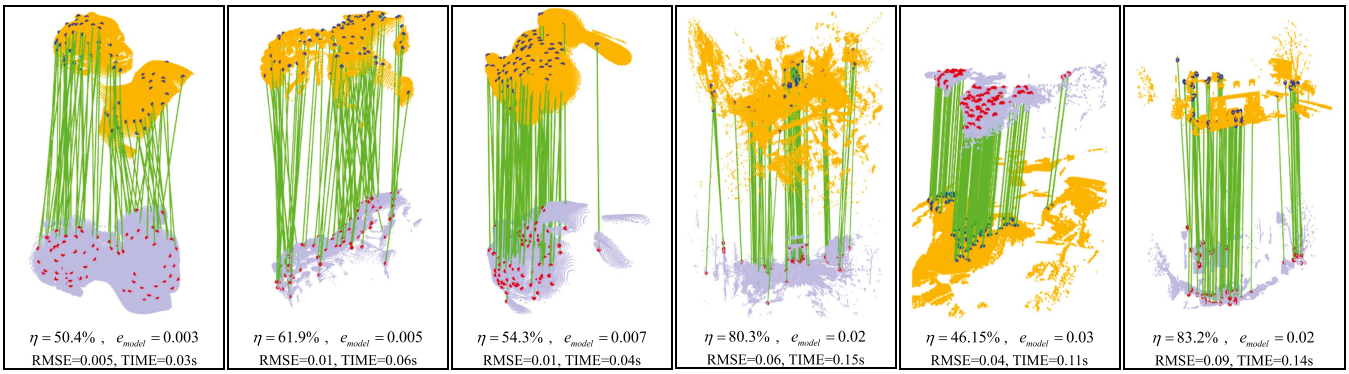


Fig. 12. Scan matching on two open datasets with ground truth transformations. (Example 6). (a) Bimba. (b) Dragon. (c) Bunny. (d) Arch. (e) Courtyard. (f) Facade.

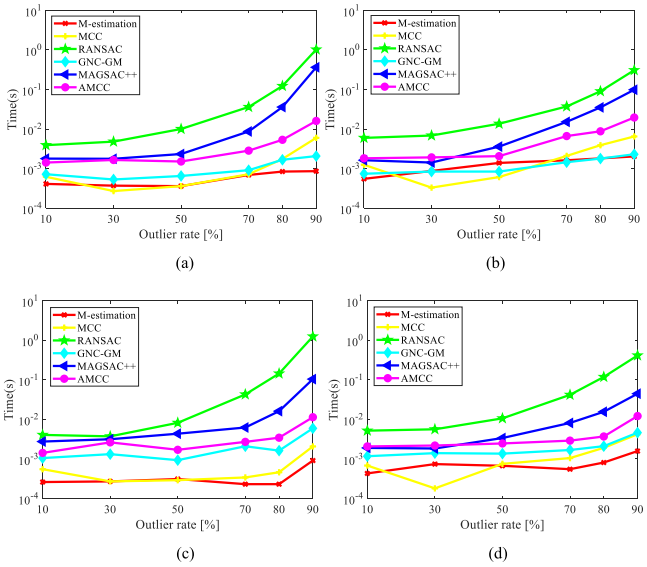


Fig. 13. Running time comparison on the tasks of IFM and point cloud registration. Note that MAGSAC++ is implemented in C++ while others are implemented in MATLAB. (a) IFM with random outliers. (b) IFM with clustered outliers. (c) PCR with random outliers. (d) PCR with clustered outliers.

from which we can make several observations. 1) RANSAC-type methods are much slower than M-estimations and MCC algorithms at high outlier rates. For example, at 80% of outliers, our AMCC is almost two orders of magnitude faster than RANSAC and one order of magnitude faster than MAGSAC++. Note that MAGSAC++ is implemented in C++ while others are implemented in MATLAB. From Figs. 7 and 11, we can see that the number of iterations of MAGSAC++ is even higher than the one of RANSAC. Hence, MAGSAC++ is actually slower than RANSAC. 2) Our AMCC can run in real-time, which is important for tasks such as SLAM. At 90% of outliers, the number of observations is 500 (a moderate size) and our AMCC only costs about 10 ms. 3) The runtime of AMCC is almost independent of the outlier rate, just like M-estimation and MCC. The slight increase in the time curve is due to the increased number of observations.

F. Ablation Study

Setup: We introduce several contributions to the traditional MCC. To demonstrate the effectiveness and necessity of each contribution, we conduct an ablation study experiment based on the LF task. The simulation process is the same as in Section V-A. We compare the proposed AMCC with classic MCC, AMCC,^{1} AMCC,^{2} AMCC,^{3} and AMCC,^{4} where the details of the last four methods are as follows.

1) AMCC^1: Replace the Silverman's rule in traditional MCC with our proposed kernel bandwidth estimation algorithm. Namely, removing the GNC idea, worst-rejection strategy, and LDM part from our AMCC.
2) AMCC^2: Removing the worst-rejection strategy and LDM part from our complete AMCC method.
3) AMCC^3: Removing the LDM part from our complete AMCC method.
4) AMCC^4: Replace our kernel bandwidth estimation algorithm with the Silverman's rule in traditional MCC.

As can be seen, the only difference between MCC and AMCC^1 (AMCC and AMCC^4) is the kernel bandwidth estimation method.

Results: Fig. 14 plots the comparison metrics, from which we can make several observations. 1) Our kernel bandwidth estimation algorithm is more stable than the one in the traditional MCC at low outlier ratios. Comparing MCC and AMCC^1 in Fig. 14(a), although MCC is more robust to high outlier ratios, AMCC^1 has a higher success rate at low outlier ratios. For example, MCC suffers from several failures at an outlier ratio of 30%. 2) The GNC idea can effectively improve the robustness and perfectly compensate for the sensitivity of our kernel bandwidth estimation method to high outlier ratios. From Fig. 14(a), AMCC^1 becomes very unreliable once the outlier ratio reaches 50%, while AMCC^2 is almost robust to 80% of outliers. The only difference between AMCC^1 and AMCC^2 is that AMCC^2 has one more GNC step than AMCC^1. 3) The worst-rejection strategy can further improve the robustness. As shown, the robustness of AMCC^3 increases to 90% while AMCC^2 is only robust to 80% of outliers. 4) The compatibility of our kernel bandwidth estimation method with GNC and worst-rejection is better than

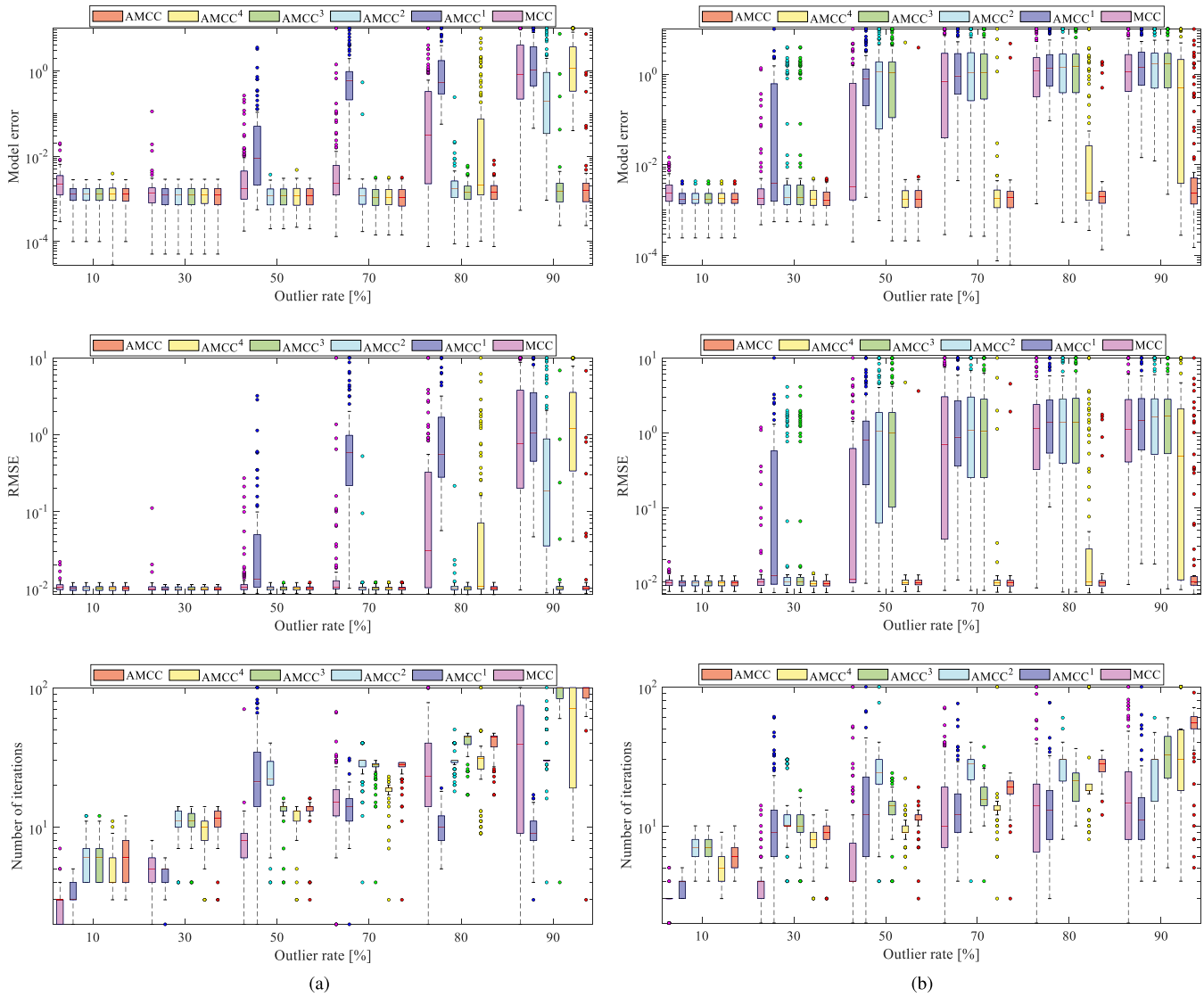


Fig. 14. Ablation study results based on LF simulations, i.e., first row – model error, middle row – RMSE, last row – number of iterations. (a) Random outliers. (b) Clustered outliers.

that of the Silverman's rule. As aforementioned, the Silverman's rule is more robust to high outlier ratios than our bandwidth estimation algorithm. However, after adding the GNC step and worst-rejection part, AMCC^4 performs much worse than our AMCC when the outlier ratio reaches 70%. 5) Our LDM step largely improves the robustness to clustered outliers without losing robustness to random outliers. From 14(b), AMCC^3 breaks down once the outlier ratio reaches 50%, while AMCC is still robust at an outlier ratio of 90%. The only difference between AMCC^3 and AMCC is the LDM step.

G. Robustness to Initialization

1) Robustness of Worst Rejection: To show the robustness of our worst rejection strategy, we perform an experiment based on the IFM task. Basic settings of the experiment are the same as in Section V-B. Differently, we provide different initializations to AMCC, and use AMCC with worst rejection (AMCC_W) and AMCC without worst rejection (AMCC_WO) as comparison methods. For initializations, we use $\boldsymbol{\theta}_{\text{ini}} = \alpha\boldsymbol{\theta}^{\circ} = (\alpha\mathbf{A}^{\circ}, \alpha\boldsymbol{t}^{\circ})$, where $\alpha = \{0.0, 0.25, 0.5, 0.75, 0.95\}$ and $\boldsymbol{\theta}^{\circ}$ is the ground truth. We report two measures S_1 and S_2 to evaluate the robustness, where S_1 indicates the ratio that AMCC_W succeeds while AMCC_WO fails (i.e., the increment of success rate when worst rejection is added), S_2 indicates the ratio that AMCC_WO succeeds while AMCC_W fails (i.e., the decrement of success rate when worst rejection is added). In total, 1000 Monte Carlo runs are performed.

From Fig. 15, we can make following observations: 1) Regardless of the initialization situation, the S_2 metric is always 0, that is, the addition of worst rejection strategy will not cause the AMCC algorithm to fail. 2) When the outlier rate reaches 70%, the S_1 metric is higher than 0, that is, the addition of worst rejection strategy will improve the success rate of AMCC.

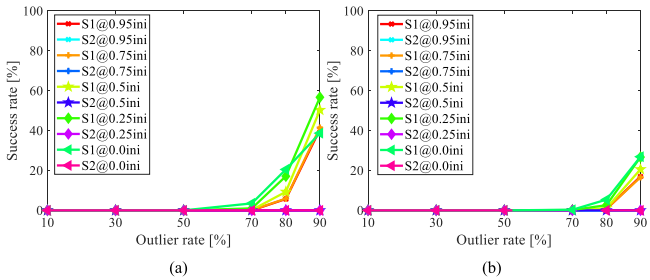


Fig. 15. Evaluation of the robustness of our worst rejection strategy. Note that some curves are identical and thus overwritten. (a) IFM with random outliers. (b) IFM with clustered outliers.

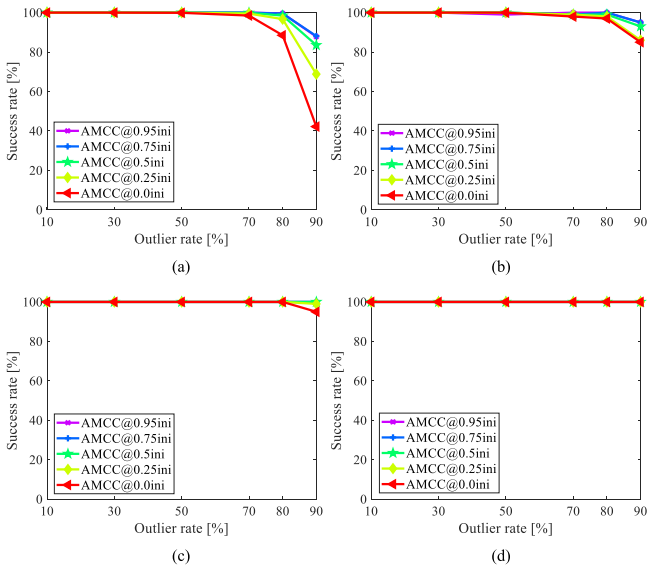


Fig. 16. Success rate results under different initializations based on the IFM and PCR tasks. (a) IFM with random outliers. (b) IFM with clustered outliers. (c) PCR with random outliers. (d) PCR with clustered outliers.

2) Robustness of AMCC: We also report the success rate of our AMCC under different initialization settings based on the IFM and PCR tasks. The initialization strategy is the same as in Section V-G1. The results are shown in Fig. 16. As can be seen: 1) Initialization has some effect on our AMCC, but this effect is relatively small and not caused by worst rejection. 2) With $\alpha = 0.25$ (a not good initialization), our success rate almost reaches 100% at 80% of outliers. So, our AMCC is not very sensitive to initializations.

H. Limitations

The limitations of our method are as follows:

1) AMCC is not as robust as the RANSAC-type methods for very high ratios of random outliers. For instance, AMCC may fail when the outlier ratio exceeds 90%.
2) AMCC is not a global optimal method, which only achieves a locally optimal solution.
3) AMCC may fail when it faces nonadversarial outliers. For example, if observations contain multiple geometric models, AMCC can only find one model at one run.

VI. Conclusion

In this article, we proposed a robust estimator, called AMCC, which is robust to both random and clustered outliers. We first presented a bandwidth estimator based on PDF matching to improve the stability of MCC. We then introduced the GNC and worst-rejection strategies into MCC to cope with high outlier ratios. We also defined a concept named LDM to measure the quality of inliers. Finally, we provided eight application examples in geometric perception and tested our AMCC on five of them, demonstrating its high efficiency, high accuracy, high robustness, and high scalability.

Appendix A
Derivation From (11) and (12)

Let $f(\chi) = \exp(-\frac{r_i^2}{2}\chi^2)$, we perform the Taylor expansion on $f(\chi)$ and take its linear part as follows:

$$f(\chi) \approx f(\chi_0) + f'(\chi_0)(\chi - \chi_0) \tag{A.1}$$

where χ_0 is a known value and

$$\begin{cases} f(\chi_0) = \exp\left(-\frac{r_i^2}{2}\chi_0^2\right) \\ f'(\chi_0) = -r_i^2\chi_0 \exp\left(-\frac{r_i^2}{2}\chi_0^2\right) \end{cases}. \tag{A.2}$$

Substituting (A.1) into $g(\chi)$, we have as follows:

$$\begin{aligned} g(\chi) &\approx \frac{\chi}{2\sqrt{\pi}} - \sqrt{\frac{2}{\pi}}\chi\frac{1}{n}\sum_{i=1}^{n}(f(\chi_0) + f'(\chi_0)(\chi - \chi_0)) \\ &\approx -\sqrt{\frac{2}{\pi}}\frac{1}{n}\sum_{i=1}^{n} f'(\chi_0)\chi^2 \\ &\quad + \left(\frac{1}{2\sqrt{\pi}} + \sqrt{\frac{2}{\pi}}\frac{1}{n}\sum_{i=1}^{n} f'(\chi_0)\chi_0 - \sqrt{\frac{2}{\pi}}\frac{1}{n}\sum_{i=1}^{n} f(\chi_0)\right)\chi. \end{aligned} \tag{A.3}$$

For notational simplicity, we further define

$$\begin{cases} a = \frac{1}{n}\sum_{i=1}^{n} f(\chi_0) = \frac{1}{n}\sum_{i=1}^{n} \exp\left(-\frac{r_i^2}{2}\chi_0^2\right) \\ b = -\frac{1}{\chi_0}\frac{1}{n}\sum_{i=1}^{n} f'(\chi_0) = \frac{1}{n}\sum_{i=1}^{n} r_i^2 \exp\left(-\frac{r_i^2}{2}\chi_0^2\right) \end{cases}. \tag{A.4}$$

Then, (11) yields a quadratic equation as follows:

$$\begin{aligned} &\arg\min_{\chi} g(\chi) \\ \approx &\arg\min_{\chi} \left\{\sqrt{\tfrac{2}{\pi}}b\chi_0\chi^2 + \left(\tfrac{1}{2\sqrt{\pi}} - \sqrt{\tfrac{2}{\pi}}b\chi_0^2 - \sqrt{\tfrac{2}{\pi}}a\right)\chi\right\} \end{aligned} \tag{A.5}$$

and its unique solution is as follows:

$$\chi^* = -\frac{\frac{1}{2\sqrt{2}} - b\chi_0^2 - a}{2b\chi_0}. \tag{A.6}$$

Appendix B
Derivation From (19) and (20)

Let $s = \frac{1}{2}$, then, $\phi(z) \doteq \rho_\mu(\sqrt{z/s}) = \rho_\mu(\sqrt{2z})$. From the definition of the Black–Rangarajan duality, we have $w = \phi'(z)$,

i.e.,

$$w = \exp\left(-\frac{z}{u\sigma^2}\right). \tag{B.1}$$

Taking the natural logarithm of (B.1) yields

$$z = -u\sigma^2 \ln w. \tag{B.2}$$

Putting (B.2) into function $\Psi_\rho(w)$, getting

$$\begin{aligned}\Psi_\rho(w) &= \phi(z) - zw \\ &= u\sigma^2\left(1 - \exp\left(-\frac{z}{u\sigma^2}\right)\right) - zw \\ &= u\sigma^2(1-w) + u\sigma^2 w \ln w \\ &= u\sigma^2 w(\ln w - 1) + u\sigma^2. \end{aligned} \tag{B.3}$$

Then, (13) becomes

$$\boldsymbol{\theta}^* = \underset{\boldsymbol{\theta}, w_i\in[0,1]}{\arg\min} \sum_{i=1}^{n} \frac{1}{2} w_i r_i^2 + u\sigma^2 w_i(\ln w_i - 1) + u\sigma^2. \tag{B.4}$$

The constant term $u\sigma^2$ can be dropped because it has no effect on the optimization results, obtaining

$$\boldsymbol{\theta}^* = \underset{\boldsymbol{\theta}, w_i\in[0,1]}{\arg\min} \sum_{i=1}^{n} \frac{1}{2} w_i r_i^2 + \mu\sigma^2 w_i (\ln w_i - 1). \tag{B.5}$$

Appendix C
GNC Optimizer

The GNC optimizer has following three main steps in an internal iteration t:

1) $\boldsymbol{\theta}$ update: Optimize (18) over $\boldsymbol{\theta}$ with given weights $\boldsymbol{w}^{(t-1)} = \{w_i^{(t-1)}\}_1^n$

$$\boldsymbol{\theta}^{(t)} = \underset{\boldsymbol{\theta}}{\arg\min} \sum_{i=1}^{n} \frac{1}{2} w_i^{(t-1)} r_i^2 \tag{C.1}$$

where the second term in (18) is a constant and is dropped. This simple weighted least-squares problem can be easily solved.

2) $\boldsymbol{w}$ update: Optimize (18) over $\boldsymbol{w}$ with given $\boldsymbol{\theta}^{(t)}$

$$\begin{aligned}\boldsymbol{w}^{(t)} &= \underset{w_i\in[0,1]}{\arg\min}\, h(w_i) \\ &= \underset{w_i\in[0,1]}{\arg\min} \sum_{i=1}^{n} \frac{1}{2} w_i \left(r_i^{(t)}\right)^2 + \mu^{(t-1)}\sigma^2 w_i (\ln w_i - 1)\end{aligned} \tag{C.2}$$

where $r_i^{(t)}$ is a known value with a given $\boldsymbol{\theta}^{(t)}$. Then, we can find the minimum of (C.2) by letting $h'(w_i) = \sum_{i=1}^{n} u^{(t-1)}\sigma^2 \ln w_i + \frac{1}{2}(r_i^{(t)})^2 = 0$ and the solution is as follows:

$$w_i^{(t)} = \exp\left(-\frac{\left(r_i^{(t)}\right)^2}{2\mu^{(t-1)}\sigma^2}\right). \tag{C.3}$$

3) μ update: Decrease the control parameter μ by a step-size τ, i.e., $\mu^{(t)} = \mu^{(t-1)}/\tau$.

References

[1] J. Ma, W. Qiu, J. Zhao, Y. Ma, A. L. Yuille, and Z. Tu, "Robust L_2E estimation of transformation for nonrigid registration," *IEEE Trans. Signal Process.*, vol. 63, no. 5, pp. 1115–1129, Mar. 2015.

[2] J. Li, Q. Hu, and M. Ai, "LAM: Locality affine-invariant feature matching," *ISPRS J. Photogrammetry Remote Sens.*, vol. 154, pp. 28–40, 2019.

[3] H. Yang, J. Shi, and L. Carlone, "TEASER: Fast and certifiable point cloud registration," *IEEE Trans. Robot.*, vol. 37, no. 2, pp. 314–333, Apr. 2021.

[4] J. Li, "A practical $O(N^2)$ outlier removal method for correspondence-based point cloud registration," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 44, no. 8, pp. 3926–3939, Aug. 2022.

[5] B. Chen, J. Cao, A. Parra, and T.-J. Chin, "Satellite pose estimation with deep landmark regression and nonlinear pose refinement," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. Workshops*, 2019, pp. 2816–2824.

[6] H. Yang and L. Carlone, "Certifiably optimal outlier-robust geometric perception: Semidefinite relaxations and scalable global optimization," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 45, no. 3, pp. 2816–2834, Mar. 2023.

[7] J. L. Schonberger and J.-M. Frahm, "Structure-from-motion revisited," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2016, pp. 4104–4113.

[8] R. M.-Artal, J. M. M. Montiel, and J. D. Tardos, "ORB-SLAM: A versatile and accurate monocular SLAM system," *IEEE Trans. Robot.*, vol. 31, no. 5, pp. 1147–1163, Oct. 2015.

[9] J. Engel, V. Koltun, and D. Cremers, "Direct sparse odometry," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 40, no. 3, pp. 611–625, Mar. 2018.

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

[11] P. J. Huber, *Robust Statistics*, vol. 523. Hoboken, NJ, USA: Wiley, 2004.

[12] P. J. Rousseeuw and A. M. Leroy, *Robust Regression and Outlier Detection*, vol. 589. Hoboken, NJ, USA: Wiley, 2005.

[13] W. Liu, P. P. Pokharel, and J. C. Principe, "Correntropy: Properties and applications in nonGaussian signal processing," *IEEE Trans. signal Process.*, vol. 55, no. 11, pp. 5286–5298, Nov. 2007.

[14] T.-J. Chin and D. Suter, "The maximum consensus problem: Recent algorithmic advances," *Synth. Lectures Comput. Vis.*, vol. 7, no. 2, pp. 1–194, 2017.

[15] A. Blake and A. Zisserman, *Visual Reconstruction*. Cambridge, MA, USA: MIT Press, 1987.

[16] H. Yang, P. Antonante, V. Tzoumas, and L. Carlone, "Graduated nonconvexity for robust spatial perception: From nonminimal solvers to global outlier rejection," *IEEE Robot. Autom. Lett.*, vol. 5, no. 2, pp. 1127–1134, Apr. 2020.

[17] W. Liu, P. Pokharel, and J. Principe, "Error entropy, correntropy and m-estimation," in *Proc. IEEE 16th Signal Process. Soc. Workshop Mach. Learn. Signal Process.*, 2006, pp. 179–184.

[18] R. He, W.-S. Zheng, and B.-G. Hu, "Maximum correntropy criterion for robust face recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 33, no. 8, pp. 1561–1576, Aug. 2011.

[19] B. W. Silverman, *Density Estimation for Statistics and Data Analysis*. Evanston, IL, USA: Routledge, 2018.

[20] Q.-Y. Zhou, J. Park, and V. Koltun, "Fast global registration," in *Proc. Eur. Conf. Comput. Vis.*, Springer, 2016, pp. 766–782.

[21] P. Antonante, V. Tzoumas, H. Yang, and L. Carlone, "Outlier-robust estimation: Hardness, minimally-tuned algorithms, and applications," *IEEE Trans. Robot.*, vol. 38, no. 1, pp. 281–301, 2021.

[22] P. H. Torr, S. J. Nasuto, and J. M. Bishop, "NAPSAC: High noise, high dimensional robust estimation-its in the bag," in *Proc. Brit. Mach. Vis. Conf.*, 2002, pp. 458–467.

[23] O. Chum and J. Matas, "Matching with prosac-progressive sample consensus," in *Proc. IEEE 2005 Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, 2005, pp. 220–226.

[24] D. Barath, M. Ivashechkin, and J. Matas, "Progressive NAPSAC: Sampling from gradually growing neighborhoods," 2019, *arXiv:1906.02295*.

[25] O. Chum, J. Matas, and J. Kittler, "Locally optimized RANSAC," in *Joint Pattern Recognition Symposium*. Berlin, Germany: Springer, 2003, pp. 236–243.

[26] K. Lebeda, J. Matas, and O. Chum, "Fixing the locally optimized RANSAC–full experimental evaluation," in *Brit. Mach. Vis. Conf.*, 2012, pp. 1–11.

[27] D. Barath and J. Matas, "Graph-cut RANSAC," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2018, pp. 6733–6741.

[28] J. Matas and O. Chum, "Randomized RANSAC with Td, d test," *Image Vis. Comput.*, vol. 22, no. 10, pp. 837–842, 2004.
[29] J. Matas and O. Chum, "Randomized RANSAC with sequential probability ratio test," in *Proc. IEEE 10th Int. Conf. Comput. Vis. Vol. 1*, 2005, vol. 2, pp. 1727–1732.
[30] O. Chum and J. Matas, "Optimal randomized RANSAC," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 30, no. 8, pp. 1472–1482, Aug. 2008.
[31] L. Moisan, P. Moulon, and P. Monasse, "Automatic homographic registration of a pair of images, with a contrario elimination of outliers," *Image Process. Line*, vol. 2, pp. 56–73, 2012.
[32] D. Barath, J. Matas, and J. Noskova, "MAGSAC: Marginalizing sample consensus," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2019, pp. 10197–10205.
[33] D. Barath, J. Noskova, and J. Matas, "Marginalizing sample consensus," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 44, no. 11, pp. 8420–8432, Nov. 2022.
[34] R. Raguram, O. Chum, M. Pollefeys, J. Matas, and J.-M. Frahm, "USAC: A universal framework for random sample consensus," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 8, pp. 2022–2038, Aug. 2013.
[35] M. Ivashechkin, D. Barath, and J. Matas, "USACV20: Robust essential, fundamental and homography matrix estimation," 2021, *arXiv:2104.05044*.
[36] M. Ivashechkin, D. Barath, and J. Matas, "VSAC: Efficient and accurate estimator for h and f," in *Proc. IEEE/CVF Int. Conf. Comput. Vis.*, 2021, pp. 15243–15252.
[37] P. Rousseeuw and V. Yohai, "Robust regression by means of S-estimators," in *Robust and Nonlinear Time Series Analysis*. Berlin, Germany: Springer, 1984, pp. 256–272.
[38] V. J. Yohai et al., "High breakdown-point and high efficiency robust estimates for regression," *Ann. Statist.*, vol. 15, no. 2, pp. 642–656, 1987.
[39] E. Olson and P. Agarwal, "Inference on networks of mixtures for robust robot mapping," *Int. J. Robot. Res.*, vol. 32, no. 7, pp. 826–840, 2013.
[40] H. Chen et al., "Robust regression with projection based M-estimators," in *Proc. 9th IEEE Int. Conf. Comput. Vis.*, 2003, pp. 878–885.
[41] S. Mittal, S. Anand, and P. Meer, "Generalized projection-based M-estimator," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 34, no. 12, pp. 2351–2364, Dec. 2012.
[42] J. Li, Q. Hu, and M. Ai, "Robust geometric model estimation based on scaled Welsch q-norm," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 8, pp. 5908–5921, Aug. 2020.
[43] J. T. Barron, "A general and adaptive robust loss function," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2019, pp. 4331–4339.
[44] B. Chen et al., "Generalized correntropy for robust adaptive filtering," *IEEE Trans. Signal Process.*, vol. 64, no. 13, pp. 3376–3387, Jul. 2016.
[45] B. Chen, X. Liu, H. Zhao, and J. C. Principe, "Maximum correntropy Kalman filter," *Automatica*, vol. 76, pp. 70–77, 2017.
[46] B. Chen, X. Wang, Y. Li, and J. C. Principe, "Maximum correntropy criterion with variable center," *IEEE Signal Process. Lett.*, vol. 26, no. 8, pp. 1212–1216, Aug. 2019.
[47] V. Vapnik, *The Nature of Statistical Learning Theory*. Berlin, Germany: Springer Sci. Bus. Media, 1995.
[48] F.-Y. Wu, K. Yang, and Y. Hu, "Sparse estimator with l0-norm constraint kernel maximum-correntropy-criterion," *IEEE Trans. Circuits Syst. II: Exp. Briefs*, vol. 67, no. 2, pp. 400–404, Feb. 2020.
[49] S. Zhao, B. Chen, and J. C. Principe, "Kernel adaptive filtering with maximum correntropy criterion," in *Proc. 2011 Int. Joint Conf. Neural Netw.*, 2011, pp. 2012–2017.
[50] S. Wang, L. Dang, B. Chen, S. Duan, L. Wang, and K. T. Chi, "Random Fourier filters under maximum correntropy criterion," *IEEE Trans. Circuits Syst. I: Reg. Papers*, vol. 65, no. 10, pp. 3390–3403, Oct. 2018.
[51] J. Peng and Q. Du, "Robust joint sparse representation based on maximum correntropy criterion for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 12, pp. 7152–7164, Dec. 2017.
[52] J. Cao, H. Dai, B. Lei, C. Yin, H. Zeng, and A. Kummert, "Maximum correntropy criterion-based hierarchical one-class classification," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 32, no. 8, pp. 3748–3754, Aug. 2021.
[53] R. He, B.-G. Hu, W.-S. Zheng, and X.-W. Kong, "Robust principal component analysis based on maximum correntropy criterion," *IEEE Trans. Image Process.*, vol. 20, no. 6, pp. 1485–1494, Jun. 2011.
[54] Y. He, F. Wang, Y. Li, J. Qin, and B. Chen, "Robust matrix completion via maximum correntropy criterion and half-quadratic optimization," *IEEE Trans. Signal Process.*, vol. 68, pp. 181–195, 2020.
[55] B. Du, T. Xinyao, Z. Wang, L. Zhang, and D. Tao, "Robust graph-based semisupervised learning for noisy labeled data via maximum correntropy criterion," *IEEE Trans. Cybern.*, vol. 49, no. 4, pp. 1440–1453, Apr. 2019.
[56] J. Liang, Y. Wang, and X. Zeng, "Robust ellipse fitting via half-quadratic and semidefinite relaxation optimization," *IEEE Trans. Image Process.*, vol. 24, no. 11, pp. 4276–4286, Nov. 2015.
[57] C. Hu, G. Wang, K. Ho, and J. Liang, "Robust ellipse fitting with Laplacian kernel based maximum correntropy criterion," *IEEE Trans. Image Process.*, vol. 30, pp. 3127–3141, 2021.
[58] Z. Wu, H. Chen, S. Du, M. Fu, N. Zhou, and N. Zheng, "Correntropy based scale ICP algorithm for robust point set registration," *Pattern Recognit.*, vol. 93, pp. 14–24, 2019.
[59] Y. Huang, H. Liu, and T. Huang, "Robust motion averaging for multi-view registration of point sets based maximum correntropy criterion," 2022, *arXiv:2208.11327*.
[60] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *Int. J. Comput. Vis.*, vol. 60, no. 2, pp. 91–110, 2004.
[61] J. Li, Q. Hu, and M. Ai, "RIFT: Multi-modal image matching based on radiation-variation insensitive feature transform," *IEEE Trans. Image Process.*, vol. 29, pp. 3296–3310, 2020.
[62] R. B. Rusu, N. Blodow, and M. Beetz, "Fast point feature histograms (FPFH) for 3D registration," in *Proc. IEEE Int. Conf. Robot. Automat.*, 2009, pp. 3212–3217.
[63] A. Zeng, S. Song, M. Nießner, M. Fisher, J. Xiao, and T. Funkhouser, "3Dmatch: Learning local geometric descriptors from RGB-D reconstructions," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2017, pp. 1802–1811.
[64] C. Choy, J. Park, and V. Koltun, "Fully convolutional geometric features," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2019, pp. 8958–8966.
[65] R. I. Hartley and P. Sturm, "Triangulation," *Comput. Vis. Image Understanding*, vol. 68, no. 2, pp. 146–157, 1997.
[66] R. Hartley and A. Zisserman, *Multiple View Geometry in Computer Vision*. Cambridge, U.K.: Cambridge Univ. Press, 2003.
[67] F. Bai, T. V.-Calleja, and G. Grisetti, "Sparse pose graph optimization in cycle space," *IEEE Trans. Robot.*, vol. 37, no. 5, pp. 1381–1400, Oct. 2021.
[68] M. J. Black and A. Rangarajan, "On the unification of line processes, outlier rejection, and robust statistics with applications in early vision," *Int. J. Comput. Vis.*, vol. 19, no. 1, pp. 57–91, 1996.
[69] J. Li, W. Xu, P. Shi, Y. Zhang, and Q. Hu, "LNIFT: Locally normalized image for rotation invariant multimodal feature matching," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5621314.
[70] M. Kisantal, S. Sharma, T. H. Park, D. Izzo, M. Märtens, and S. D'Amico, "Satellite pose estimation challenge: Dataset, competition design, and results," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 56, no. 5, pp. 4083–4098, Oct. 2020.
[71] Y. Zhong, "Intrinsic shape signatures: A shape descriptor for 3D object recognition," in *Proc. IEEE Int. Conf. Comput. Vis. Workshops*, 2009, pp. 689–696.
[72] J. Li, P. Zhao, Q. Hu, and M. Ai, "Robust point cloud registration based on topological graph and cauchy weighted lq-norm," *ISPRS J. Photogrammetry Remote Sens.*, vol. 160, pp. 244–259, 2020.

Jiayuan Li received the B.Eng., M.Eng., and Ph.D. degrees in photogrammetry and remote sensing from the School of Remote Sensing and Information Engineering, Wuhan University, Wuhan, China, in 2012, 2015, and 2018, respectively.

He is currently a Professor with Wuhan University. He has authored more than 60 peer-reviewed articles in international journals. His research interests include SLAM, image matching, and point cloud registration.

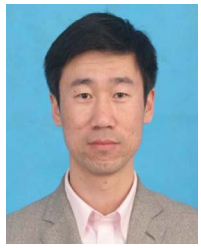
Dr. Li was the recipient of the Best Youth Author Award by ISPRS in 2021 and the Talbert Abrams Award by ASPRS in 2018.

Qingwu Hu received the B.Eng. and M.Eng. degrees in photogrammetry and remote sensing from the Wuhan Technical University of Surveying and Mapping, Wuhan, China, and the Ph.D. degree in photogrammetry and remote sensing from Wuhan University, Wuhan, in 2007.

He has authored more than 100 peer-reviewed articles in international journals. His research interests include methods, techniques, and applications of remote sensing, GIS and GPS integration, and photogrammetry.

Xinyi Liu received the B.S. and Ph.D. degrees from the School of Remote Sensing and Information Engineering, Wuhan University, Wuhan, China, in 2014 and 2020, respectively.

She is currently an Associate Researcher with Wuhan University. Her research interests include 3-D reconstruction, LiDAR and image integration, and texture mapping.



Yongjun Zhang (Member, IEEE) received the B.S. degree in geodesy, the M.S. degree in geodesy and surveying engineering, and the Ph.D. degree in geodesy and photography from Wuhan University, Wuhan, China, in 1997, 2000, and 2002, respectively.

He is currently the Dean of the School of Remote Sensing and Information Engineering, Wuhan University. He has authored or coauthored more than 180 research articles and one book. His research interests include aerospace and low-attitude photogrammetry, image matching, combined block adjustment with multisource datasets, object information extraction and modeling with artificial intelligence, integration of LiDAR point clouds and images, and 3-D city model reconstruction.

Dr. Zhang is the coEditor-in-Chief of *The Photogrammetric Record*.