Robust Ellipse Fitting Based on Maximum Correntropy Criterion With Variable Center

Wei Wang, Gang Wang®, Senior Member, IEEE, Chenlong Hu®, and K. C. Ho®, Fellow, IEEE

Abstract—The presence of radically irregular data points (RIDPs), which are referred to as the subset of measurements that represents no or little information, can significantly degrade the performance of ellipse fitting methods. We develop an ellipse fitting method that is robust to RIDPs based on the maximum correntropy criterion with variable center (MCC-VC), where an adaptable Laplacian kernel is used. For single ellipse fitting, we formulate a non-convex optimization problem and divide it into two subproblems, one to estimate the kernel bandwidth and the other the kernel center. We design sufficiently accurate convex approximation to each subproblem that will lead to computationally efficient closed-form solutions. The two subproblems are solved in an alternate manner until convergence is reached. We also investigate coupled ellipses fitting. While there exist multiple ellipses fitting methods in the literature, we develop a coupled ellipses fitting method by exploiting the underlying special structure, where the associations between the data points and ellipses are absent in the problem. The proposed method first introduces an association vector for each data point and then formulates a non-convex mixed-integer optimization problem to establish the data associations, which is approximately solved by relaxing it into a second-order cone program. Using the estimated data associations, we then extend the proposed single ellipse fitting method to accomplish the final coupled ellipses fitting. The proposed method is shown to perform significantly better than the existing methods using both simulated data and real images.

Index Terms—Ellipse fitting, radically irregular data point (RIDP), maximum correntropy criterion with variable center (MCC-VC), data association.

I. INTRODUCTION

A basic function of computer vision, ellipse fitting has been extensively studied over the years. The task of ellipse fitting is to fit a series of data points to an ellipse, which finds wide applications in the fields of aerospace industry, medical imaging, pupil recognition, biometrics, and many others [1], [2], [3], [4]. For example, ellipse fitting can detect effectively the rocky and other dangerous areas to ensure a stable and smooth landing of the lunar lander [1]. Techniques based on ellipse fitting can accurately detect whether a patient has glaucoma symptoms in the medical related fields [2].

Robust Pupil Recognition

Pupil recognition technology, which has been popular in recent years, also relies on ellipse fitting [3]. Indeed, pupil tracking through ellipse fitting has been extensively used in human-computer interactive eye tracker, and has achieved considerable success in consumer electronics industry [4].

A traditional method for ellipse fitting is based on the Hough transform (HT) [5]. It can achieve high-precision fitting through a voting mechanism in a five dimensional parameter space, but it takes a huge computational load. Owing to this drawback, more computationally efficient methods based on least squares (LS) were proposed [6], [7], [8], [9], [10], [11], [12], [13]. These methods typically work well for clean or simple instances. However, the performance of the LS based methods may degrade significantly in the presence of radically irregular data points (RIDPs). RIDPs appear very often in the extracted data points and they deviate significantly from the underlying ellipse for fitting. Robust ellipse fitting methods that are resilient against RIDPs were thus proposed, see e.g., [14], [15], [16], [17], [18], [19]. An example of such methods is the least median squares (LMS) method [14]. The motivation behind this approach is that the optimal solution should embrace the lowest median of squared residuals. Another approach is to use a portion of the sample points instead of the whole to resist the degradation from RIDPs in the fitting performance. For instance, Fischler et al. proposed a method using the random sample consensus (RANSAC) [15], which achieves the fitting by estimating a mathematical model from a collection of random subsets of the entire amount of data.

Alternatively, the ellipse fitting problem can be formulated as an optimization problem that can provide robustness. The papers [16] and [17] solved the ellipse fitting problem by using the maximum correntropy criterion (MCC), where the Gaussian and Laplacian kernels were used, respectively. The MCC method in [16] iteratively solved a number of semidefinite programs (SDPs) by applying semidefinite relaxation (SDR) to the original problem. The relaxation may lead to divergence of the iterations and cause performance loss. By contrast, the MCC method in [17] iteratively solved a set of more computationally efficient second-order cone programs (SOPCs), in which no relaxation was introduced. This, in turn, ensures convergence of the iterations. Generally speaking, the MCC ellipse fitting methods have great robustness to RIDPs in maintaining good performance. Recently, Zhao et al. [18] proposed the hierarchical Gaussian mixture model (HGM) for ellipse fitting in noisy, RIDPs-contaminated, and occluded scenarios. This method has high robustness against RIDPs and noise when the parameters for the algorithm are chosen properly. However, the results can be unsatisfactory when using one particular set of parameters for fitting ellipses with different variations in size and orientation. More recently, Thurnhofer-Hemsi et al. [19] proposed a novel ellipse fitting method based on the spatial median consensus. The idea

Copyright (c) 2023 IEEE. Personal use of this material is permitted. However, permission to reuse this material for public or commercial purposes requires IEEE permission. See https://www.ieee.org/publications/rights/index.html for more information.
behind is to combine the outputs of an ensemble of ellipse fitting methods so that the deleterious effect of suboptimal fits is alleviated.

At present, most works in this research area address the fitting of single ellipse. However, multiple ellipses with some commonality can be fitted together by exploiting the underlying special structure to improve performance. An example is the case of coupled ellipses. Coupled ellipses refer to ellipses that are concentric and have the same rotation angle; their semi-minor and semi-major axes are different but related by a proportional factor. A typical example of coupled ellipses is the inner and outer boundaries of a transmission pipe, or the inner and outer edges of an iris image. More applications related to coupled ellipses can be found in [20] and the references therein.

Compared with single ellipse fitting, coupled ellipses fitting is more challenging since it has one additional parameter for estimation. To our best knowledge, the available methods for coupled ellipses fitting are very limited in the literature. Ma and Ho [20] proposed a weighted least squares (WLS) method for the fitting problem that can reach the Kanatani-Cramer-Rao (KCR) lower bound accuracy when the data do not have RIDPs. However, the performance degrades significantly when RIDPs occur. To improve the robustness against RIDPs, the work [17] applied MCC with the Laplacian kernel and found the solution by alternately solving two subproblems. The method works well even when a large number of RIDPs are present.

It has been shown that MCC is very powerful in solving some signal processing problems when the data are corrupted by heavy-tailed impulsive noise [21], [22]. It has also been successfully applied to the ellipse fitting problem [16], [17]. However, the existing methods use zero-center kernel functions, which may not match well for the data with non-zero-mean error distributions, leading to unsatisfactory fitting performance. Moreover, the kernel bandwidths in [16] and [17] were computed according to the Silverman’s rule [23]. The resulting kernel bandwidths may not be accurate, which is particularly true for the case of small data size. Moreover, the Silverman’s rule is designed for the Gaussian kernel and provides the optimal parameter values only when the data samples are Gaussian distributed [24]. As such, Chen et al. [25] proposed the MCC with variable center (MCC-VC) to improve the performance of MCC, in which both the kernel bandwidth and center can vary to best model the data. MCC-VC is a generalization of MCC and can be used to handle a broader class of problems.

In this paper, we propose an MCC-VC based ellipse fitting method that is robust against RIDPs, in which the Laplacian kernel is used. In the proposed MCC-VC method, the kernel bandwidth and center are estimated using the error samples derived from the data. The original work [25] that proposed MCC-VC has neither explained how to obtain the objective function, nor provided an efficient method to find the kernel bandwidth and center in the objective function. We first develop a new explanation based on the kernel density estimation (KDE) and then formulate an optimization problem for the estimation of the kernel bandwidth and center parameters in MCC-VC. The proposed optimization problem is shown to be equivalent to that in [25]. We then propose to solve the optimization problem by dividing it into two subproblems, each estimating one parameter. Rather than using iterative methods that require good initialization, we design sufficiently accurate convex approximation to each subproblem that results in a computationally efficient solution. Specifically, we propose a fourth-order polynomial approximation by applying the Taylor expansion to the objective function for the estimation of kernel bandwidth. The approximate convex subproblem has a closed-form solution, thereby obtaining it has a very low computational complexity. The approximation is improved by updating with the kernel bandwidth estimate until convergence. For the subproblem of estimating the kernel center, we design a linear programming (LP) problem, whose solution turns out to be the median of the error samples, and thus is also in closed-form. Using the LP solution as a starting point, we further propose a bisection method for refining the kernel center estimate.

Armed with the estimated kernel bandwidth and center, the estimation of ellipse parameters reduces to an optimization problem by MCC with the Laplacian kernel, which can be solved by using the iterative method in [17]. It is worth noting that this work formulates a different second-order cone (SOC) constraint from the one in [17] to ensure that the fitting results in an ellipse, which enables us to solve the SOCP problem only once in each iteration, instead of twice in the previous MCC method [17]. Hence, the new SOC constraint reduces the computational complexity of solving the MCC problem. The estimated ellipse parameters are used to form better error samples for updating the estimates of the kernel bandwidth and center. The estimation of the kernel and ellipse parameters are repeated alternately until convergence.

Coupled ellipses can be viewed as multiple ellipses with some special structure, and hence, the multiple ellipses fitting methods, e.g., those in [5], [15], [18], [26], [27], [28], [29] are applicable to coupled ellipses fitting. However, the inherent special structure of coupled ellipses is not exploited in these methods, which leads to sub-optimal performance. Furthermore, existing methods for coupled ellipses fitting [17], [20] require the prior knowledge of the associations between the data points and the ellipses, which is not straightforward or impossible to obtain in practice. To alleviate this issue, we introduce a length 2 association vector, one for each data point having elements equal to 0 or 1, for indicating the association of every data point to either ellipse, and formulate a mixed integer optimization problem to jointly estimate the ellipse parameters and the association vectors. It is commonly known that the mixed integer problem is very difficult to solve.

To make the problem tractable, we relax each association vector to a probability vector, yielding a convex SOCP. The estimate of each association vector can be deduced from the SOCP solution. Using the estimated association vectors, the proposed single ellipse MCC-VC fitting method is extended to coupled ellipses fitting. It is worth noting that the incorrectly associated data points can be interpreted as RIDPs during the fitting, implying the fitting method needs to be robust to RIDPs.

The contributions are summarized as follows, including:

1) We propose a robust ellipse fitting method based on MCC-VC, in which the kernel bandwidth and center are estimated by a well explained and properly formulated optimization problem through KDE. Moreover, this problem is efficiently solved by an accurate convex approximation.

2) We formulate a new SOC constraint to guarantee the fitting result is an ellipse, which greatly improves the computational efficiency of the proposed method.

3) We propose a data association method for the coupled ellipses fitting problem when the associations between the data points and the ellipses are not known. With the estimated data associations, we extend the proposed
MCC-VC single ellipse fitting method to achieve the final coupled ellipses fitting.

In this paper, Section II gives the measurement models for single ellipse, and coupled ellipses without the knowledge of data association. In Section III, we present the optimization problems for single ellipse fitting based on MCC-VC, and derive an iterative method to solve the optimization problems. Section IV develops the data association method and presents the MCC-VC method for coupled ellipses fitting. Section V demonstrates the performance of the proposed fitting methods by several experiments using both simulated data and real images. Section VI concludes the paper. Regarding notations, vectors and matrices are represented by boldface lowercase and boldface uppercase letters, respectively. (v)\textsuperscript{\_i} is the true value of (v). E[\*] is the mathematical expectation of the vector (v). (v)((k)) is the k\textsuperscript{th} element of (v). |\*| and \|\*\| are the \(\ell_1\)-norm and \(\ell_2\)-norm, respectively. Other mathematical symbols are defined when they first appear.

II. SYSTEM MODELS
A. Single Ellipse

In the 2D Euclidean space, an ellipse can be uniquely determined by five parameters: the center \((g, h)\), semi-major axis \(a\), semi-minor axis \(b\) and counter-clockwise rotation angle \(\theta\). The ellipse constructed by these five parameters is [17]

\[
\frac{((x - g) \cos \theta + (y - h) \sin \theta)^2}{a^2} + \frac{(-(x - g) \sin \theta + (y - h) \cos \theta)^2}{b^2} = 1,
\]

(1)

where \((x^o, y^o)\) denotes a regular point on the ellipse.

The available data points, including RIDPs, can be modeled by

\[x_i = x_i^o + v_i^x + n_i^x, \quad y_i = y_i^o + v_i^y + n_i^y,\]

(2)

where \(n_i^x\) and \(n_i^y\) are the measurement noise, and \(v_i^x\) and \(v_i^y\) are zero if the pair is a noise only contaminated point and have larger values if it is an RIDP. We assume that the measurement noise follows a zero-mean Gaussian distribution, but the distributions of \(v_i^x\) and \(v_i^y\) are difficult to model. Indeed, RIDPs in the ellipse fitting problem are referred to as the subset of measurements that represents no or much less information compared to the rest for the underlying estimation problem. Such a definition of RIDPs inherently assumes we know what are the parameters of interest for modeling the shape of an object, in particular, ellipse in this work. RIDPs typically come from corruption or the presence of other objects in the image that we extract the data points for ellipse fitting. RIDPs can appear as outliers that occur sparsely and randomly in the data, or it can emerge as anomalies that are grouped together, or both. Hence, the forms of RIDPs are problem specific, and the distributions of \(v_i^x\) and \(v_i^y\) cannot be described by generic models.

Substituting (2) into (1) and after simple manipulations, we can rewrite (1) as the following implicit second-order polynomial equation [8], [16], [17]:

\[Ax_i^2 + Bx_iy_i + Cy_i^2 + Dx_i + Ey_i + F = \delta_i, \quad i = 1, \ldots, N,\]

(3)

where \(\delta_i\) is the measurement error induced by noise and RIDPs. For simplicity, we introduce the following vectors:

\[v = [A, B, C, D, E, F]^T, \quad q = [g, h, a, b, \theta]^T, \quad u_i = [x_i^2, x_iy_i, y_i^2, x_i, y_i]^T.\]

(4)
The model (3) can be expressed in the vector form

\[v^T u_i = \delta_i, \quad i = 1, \ldots, N.\]

(5)

To guarantee that the model (3) represents an ellipse but not a hyperbola, the condition \(B^2 - 4AC < 0\), i.e., \(v_{(2)}^2 - 4v_{(1)}v_{(3)} < 0\), must be satisfied. The objective is to solve the ellipse fitting problem by obtaining an optimal estimate of \(v\) with the condition \(v_{(2)}^2 - 4v_{(1)}v_{(3)} < 0\) being satisfied. From the estimate of \(v\), we can deduce the estimate of the ellipse parameter vector \(q\) by the common conversion formulas [17].

B. Coupled Ellipses

The equations for the outer and inner coupled ellipses can be expressed as [20]

\[
\begin{aligned}
\delta_i \left( \frac{((x - g) \cos \theta + (y - h) \sin \theta)^2}{a^2_j} + \frac{(-(x - g) \sin \theta + (y - h) \cos \theta)^2}{b^2_j} \right) = \beta, & \quad j = 1, 2.
\end{aligned}
\]

(6)

Assume that we have collected \(N\) noisy data points \((x_i, y_i)\), \(i = 1, 2, \ldots, N\), where the associations between the data points and the ellipses are not known. After substituting \((x_i, y_i)\) to both equations of the outer and inner ellipses in (7), the model equations become [17]

\[
A_{x_i}^2 + B_{x_i}x_iy_i + C_{y_i}^2 + Dx_i + Ey_i + F = \delta_i, \quad j = 1, 2,
\]

(8a)

\[
A_{x_i}^2 + B_{x_i}x_iy_i + C_{y_i}^2 + Dx_i + Ey_i + F + \eta = \delta_i,
\]

(8b)

where \(\eta = \beta(1 - \mu^2)\) and \(\delta_i\) for \(j = 1, 2\) are the equation errors of the \(i\)th data point \((x_i, y_i)\) corresponding to the outer and inner ellipses. Typically, \(|\delta_{i,1}| < |\delta_{i,2}|\) if the \(i\)th point belongs to the outer ellipse, and \(|\delta_{i,1}| > |\delta_{i,2}|\) otherwise.

The equations in (8a) and (8b) can be further represented by the following vector form:

\[v^T u_i + v^T u_i = \tau = \delta_i, \quad i = 1, \ldots, N,\]

(9)

where

\[
\begin{aligned}
v = [A, B, C, D, E, F]^T, \quad \tau = [\eta, 0]^T, \quad u_i = [x_i^2, x_iy_i, y_i^2, x_i, y_i]^T, \quad \delta_i = [\delta_{i,1}, \delta_{i,2}]^T.
\end{aligned}
\]

(10)
To explicitly express the model equation for coupled ellipses fitting that includes data association, we introduce an association vector \( \phi_i \) to represent the association between either ellipse and the \( i \)th noisy data point. It will take the value of \( \phi_i = [1, 0]^T \) if the \( i \)th data point is associated with the outer ellipse and \( \phi_i = [0, 1]^T \) if it is associated with the inner ellipse. \( \phi_i \) is not known and to be estimated. The model for coupled ellipses fitting after data association is obtained by multiplying the association vector \( \phi_i \) to both sides of (9):
\[
\phi_i^T \delta_i = v_i^T u_i + \phi_i^T \tau, \quad i = 1, \ldots, N, \quad (11)
\]
where the left term is the model error of the \( i \)th data point.

III. SINGLE ELLIPSE FITTING BASED ON MCC-VC

In this section, we propose a robust formulation for single ellipse fitting by MCC-VC, to achieve better robustness against RIDPs. Moreover, we propose a more efficient procedure than the one in [17], to guarantee the condition for forming an ellipse, i.e., \( v_i^2 - 4v_{(1)}v_{(3)} < 0 \), is fulfilled.

A. MCC-VC With Laplacian Kernel

MCC can be utilized for robust estimation of unknowns. According to [17], the unknown vector \( \alpha \), which is related to the error vector \( \delta \), can be estimated using the MCC by
\[
\max_{\alpha} \mathbb{E}[\kappa_{\sigma}(\delta(\alpha))], \quad (12)
\]
where \( \kappa_{\sigma}(\bullet) \) is the kernel function with bandwidth \( \sigma \).

The value of \( \sigma \) is usually determined according to the Silverman’s rule [23]. Such an obtained value may not be accurate especially when the number of samples used for evaluating the expectation is small or the kernel function is not Gaussian. For this reason, we propose an MCC with adaptive kernel bandwidth \( \sigma \), which involves KDE.

In statistics, KDE can be considered as using a non-parametric approach to estimate the probability density function of a random variable. Given a finite number of data samples, KDE is a fundamental data smoothing problem. Specifically, let \( \delta_i \) be the independent and identically distributed samples drawn from some univariate distribution of an unknown density \( p \). The kernel density estimator \( \hat{p}_{\alpha}(\delta) \) for a given sample point \( \delta \) can be expressed as
\[
\hat{p}_{\alpha}(\delta) = \frac{1}{N} \sum_{i=1}^{N} \kappa_{\sigma}(\delta - \delta_i). \quad (13)
\]

In KDE, the kernel bandwidth estimation is key to the performance. The commonly used criterion for selecting \( \sigma \) is to minimize the mean integrated squared error (MISE) [30]:
\[
\text{MISE}(\sigma) = \mathbb{E} \left[ \int (\hat{p}_{\alpha}(\delta) - p(\delta))^2 d\delta \right], \quad (14)
\]
where \( p(\delta) \) is the probability density function (PDF) of the true error distribution. However, minimizing the MISE is generally not feasible since \( p(\delta) \) is not available. To make the problem tractable, we make approximations in the following.

Since there is only one set of samples, we first remove the expectation operation, i.e., we minimize the integrated squared error (ISE) instead of the MISE. It is not difficult to observe from (13) and (14) that minimizing ISE is still an intractable problem owing the square of the summation. Thus, we further approximate \( \hat{p}_{\alpha}(\delta) \) by using one sample, i.e., \( \hat{p}_{\alpha}(\delta) \approx \kappa_{\sigma}(\delta - c) \), where \( c \) is the an unknown representative sample. \( c \) is often called as the center of the kernel function [25].

Finally, we seek the optimal \( c \) and \( \sigma \) by minimizing the ISE. The term \( \int (p(\delta))^2 d\delta \) is independent of \( c \) and \( \sigma \), hence
\[
\begin{align*}
(c^*, \sigma^*) &= \arg \min_{c, \sigma} \int \left[ \kappa_{\sigma}(|\delta - c| - p(\delta)) \right]^2 d\delta \\
&= \arg \min_{c, \sigma} \left\{ \int [\kappa_{\sigma}(|\delta - c|)]^2 d\delta - 2E[\kappa_{\sigma}(|\delta - c|)] \right\}. \quad (15)
\end{align*}
\]

The mathematical expectation in problem (15) is approximated by sample averaging over the \( N \) samples \( \{\delta_i\}_{i=1}^{N} \). In this paper, the kernel function is Laplacian for its robustness to RIDPs. Substituting the Laplacian kernel function \( \kappa_{\sigma}(|\delta - c|) = \frac{1}{\sigma} e^{-\frac{|\delta - c|}{\sigma}} \) into problem (15) gives
\[
\begin{align*}
(c^*, \sigma^*) &= \arg \min_{c, \sigma} \left\{ \frac{1}{4\sigma^2} - \frac{1}{N\sigma} \sum_{i=1}^{N} e^{-\frac{|\delta_i - c|}{\sigma}} \right\}. \quad (16)
\end{align*}
\]

It is worth noting that problem (16) is obtained differently from [25] although the final form is exactly the same as that in [25] when the Laplacian kernel function is used.

With the estimated \( c^* \) and \( \sigma^* \) by problem (16), MCC follows to estimate the unknown parameters in the model by
\[
\arg \min_{\alpha} \left\{ \frac{1}{N} \sum_{i=1}^{N} \kappa_{\sigma^*}(\delta_i(\alpha) - c^*) \right\} = \arg \min_{\alpha} \left\{ \frac{1}{\sigma^*} \sum_{i=1}^{N} e^{-\frac{|\delta_i(\alpha) - c^*|}{\sigma^*}} \right\}, \quad (17)
\]
where the \( \frac{1}{N} \) is discarded as it is a constant. To summarize, the MCC-VC method consists of solving problems (16) and (17).

B. Single Ellipse Fitting Based on MCC-VC

In the ellipse fitting problem, \( v \) is the unknown vector to be estimated. We shall deduce an optimization problem to estimate \( v \) based on MCC-VC.

It is seen from (5) that \( \delta_i \) is related to the unknown vector \( v \). Thus, we express \( \delta_i = v_i^T u_i \) for \( i = 1, \ldots, N \). Using an estimate of \( v \), denoted by \( \hat{v} \), we can construct the samples \( \delta_i(\hat{v}) \). Problem (16) becomes
\[
(c^*, \sigma^*) = \arg \min_{c, \sigma} \left\{ \frac{1}{4\sigma^2} - \frac{1}{N\sigma} \sum_{i=1}^{N} e^{-\frac{|\delta_i(\hat{v}) - c|}{\sigma^*}} \right\}. \quad (18)
\]

Using the estimated \( c \) and \( \sigma \), denoted by \( \hat{c} \) and \( \hat{\sigma} \), the ellipse parameter vector can be estimated by solving
\[
\hat{v} = \arg \min_{v} \left\{ -\frac{1}{\hat{\sigma}} \sum_{i=1}^{N} e^{-\frac{|v_i^2 u_i^T u_i - \hat{c}|}{\hat{\sigma}}} \right\} \quad \text{s.t.} \quad v_i^2 - 4v_{(1)}v_{(3)} < 0, \quad (19)
\]
where the condition \( v_i^2 - 4v_{(1)}v_{(3)} < 0 \) is included as a constraint to ensure the solution is an ellipse.

In the following, we develop specific methods to solve problems (18) and (19) iteratively for estimating the unknown parameters \( c, \sigma, \) and \( v \) in an alternate manner.
1) Estimation of the Kernel Bandwidth and Center: Joint estimation of the kernel bandwidth and center in problem (18) may lead to local convergence owing to the non-convex nature of the problem. To avoid the local convergence issue, we propose to estimate the two parameters by dividing problem (18) into two subproblems, with one parameter estimated by one subproblem. The subproblems are not solved using the routine gradient based method, such as the gradient decent method or the Newton’s method. Instead, we design a sufficiently accurate convex problem for each subproblem, and the solution of the convex problem is used as the starting point to obtain the optimal solution of the corresponding subproblem by local search. Since the designed convex problems are sufficiently accurate to provide good starting points, we expect to be able to obtain the optimal solutions of the original non-convex subproblems, although global convergence is not guaranteed.

a) Estimation of the kernel bandwidth: We first present a procedure to estimate the kernel bandwidth $\sigma$ when fixing $v$ and $c$ to their estimates $\hat{v}$ and $\hat{c}$ from the previous iteration. For notational simplicity, let $\delta_i = (\hat{v}^T u_i)$ for $i = 1, \ldots, N$. Using $\delta_i$ and $\hat{c}$, we can estimate $\sigma$ by solving

$$\min_{\sigma} \left\{ \frac{1}{4\sigma} - \frac{1}{N\sigma} \sum_{i=1}^{N} e^{-\frac{|\delta_i - \hat{c}|}{\sigma}} \right\}. \quad (20)$$

By letting $r = \frac{1}{\sigma}$, problem (20) can be written into an optimization problem with $r$ as the variable:

$$\min_{r} \left\{ h(r) := \frac{r}{4} - \frac{r}{N} \sum_{i=1}^{N} e^{-\frac{|\delta_i - \hat{c}|}{r}} \right\}. \quad (21)$$

Once the optimal solution of problem (21) is obtained, the optimal solution of problem (20) is also readily available.

By computing the second-order derivative of $h(r)$, it can be proven that problem (21) is non-convex [31], which may lead to possible local convergence if it is solved by an iterative algorithm. We shall approximate (21) as a sufficiently accurate convex problem through truncating the Taylor expansion. For brevity, we denote $\hat{a}_i = |\delta_i - \hat{c}|$ for $i = 1, \ldots, N$. After introducing a known positive constant $r_0$, Appendix A shows that $h(r)$ can be approximated by the following fourth order polynomial function, i.e.,

$$h(r) \approx f(r) = \frac{b_4}{6} (r - r_0)^4 + \frac{b_3}{2} (r - r_0)^3 + b_2 (r - r_0)^2 + (b_2 r_0 - b_1 + \frac{1}{4}) (r - r_0) + (\frac{1}{4} r_0^2 - b_1 r_0),$$

where $b_1, b_2, b_3,$ and $b_4$ are defined in Appendix A. The approximation is typically accurate around $r_0$.

It is straightforward to see that the optimal solution of problem (20) can be obtained by sequentially minimizing $f(r)$. In particular, starting from $r_0 = 0$, we can minimize $f(r)$ to obtain an updated $r_0$, then we minimize $f(r)$ again using the updated $r_0$ and repeat the process until convergence.

It turns out that the problem of minimizing $f(r)$ has a closed-form solution. According to the Karush-Kuhn-Tucker (KKT) condition, the solution of minimizing $f(r)$ can be obtained by solving the univariate cubic equation $f'(r) = 0$, whose real root has a closed-form expression. Appendix B shows the existence, uniqueness, and the expression of the closed-form solution. The solution in closed-form implies a lower computational complexity than the gradient based methods to solve problem (20). Finally, the optimal estimation of $\sigma$ can be obtained by taking the reciprocal of $r$.

b) Estimation of the kernel center: We will present the procedure to estimate the kernel center $c$ when fixing bandwidth $\sigma$ to its estimate $\hat{\sigma}$ from the previous iteration. Problem (16) now becomes

$$\min_{c} - \sum_{i=1}^{N} e^{-\frac{|\hat{\delta}_i - c|}{\hat{\sigma}}}, \quad (23)$$

which can be used to obtain the kernel center $c$.

However, problem (23) is non-convex owing to the non-convex objective function. To simplify the problem, we shall approximate the objective function of problem (23) by a convex function. Specifically, we keep only the first-order Taylor expansion of $e^{-\frac{|\hat{\delta}_i - c|}{\hat{\sigma}}}$, i.e., $e^{-\frac{|\hat{\delta}_i - c|}{\hat{\sigma}}} \approx 1 - \frac{|\hat{\delta}_i - c|}{\hat{\sigma}}$. As a result, we can approximate problem (23) by the LP problem

$$\min_{c} - \sum_{i=1}^{N} \left( 1 - \frac{|\hat{\delta}_i - c|}{\hat{\sigma}} \right) = \min_{c} \sum_{i=1}^{N} |\hat{\delta}_i - c|. \quad (24)$$

In fact, problem (24) is the maximum likelihood estimation of the center of the Laplacian distribution from the samples that follow the Laplacian distribution [32], and its solution is the median of the samples. Hence, the solution would be a reasonably approximated one even when the samples are not Laplacian distributed, although it is not optimal to the original problem (23). On the other hand, the optimal solution of problem (23) can be obtained through a simple procedure using the LP solution as the starting point. Noting that the objective of problem (23) is non-differentiable, the gradient based method is not applicable. We propose a simple bisection method to solve problem (23). To this end, we first determine an appropriate interval of $c$ for the bisection method, containing the LP solution and the sample around the LP solution having smallest objective value, and then perform the bisection method. When the LP solution is accurate enough, the bisection method will give the optimal solution of problem (23).

2) Estimation of the Ellipse Parameters: When keeping $c$ and $\sigma$ to the values $\hat{c}$ and $\hat{\sigma}$, the MCC-VC problem reduces to the following MCC problem:

$$\min_{v} - \sum_{i=1}^{N} e^{-\frac{v^T u_i}{\hat{\delta}_i}}, \text{ s.t. } v_{(2)}^2 - 4v_{(1)}v_{(3)} < 0, \quad (25)$$

where the constant scaling parameter $1/\hat{\sigma}$ has been dropped. The previous work [17] presented a method for solving problem (25). Specifically, the solution to (25) is obtained by solving the following two subproblems iteratively:

a) Subproblem 1: Assuming that an estimate of the weight vector $w$ is available, denoted by $\hat{w}$, the first subproblem is:

$$\min_{v, \xi} \sum_{i=1}^{N} (-\hat{w}_i \xi_i) \quad \text{s.t. } \|\hat{v}^T u_i - \hat{c}\| \leq \xi_i, \ i = 1, \ldots, N, \quad (26a)$$

$$v_{(2)}^2 - 4v_{(1)}v_{(3)} < 0. \quad (26b)$$
b) Subproblem 2: If an estimate of \( v \), denoted by \( \hat{v} \), is available, we can obtain the optimal estimate of the weight vector \( w = [w_1, \ldots , w_N]^T \) by the property of convex conjugate functions: 
\[
\hat{w}_i = -e^{\frac{1}{\sigma} v^T u_i - \frac{1}{\sigma}}.
\]

The solution method typically takes the “\( \leq \)“ as “\( \leq \)“ when solving Subproblem 1. A procedure was proposed to guarantee the sign “\( \leq \)“ in [17], where an SOCP problem was solved twice in one iteration. Different from [17], we propose here a more efficient procedure for the condition \( u_i^2 - 4u(1)v(3) < 0 \) to be fulfilled. Specifically, we replace the constraint in (26b) by \( u_i^2 + \varepsilon^2 \leq 4u(1)v(3) \), where \( \varepsilon \) is an arbitrary constant.

It is straightforward to write the constraint (26b) as the following SOC constraint:
\[
\left[ u(2), v(1) - v(3) \right]^T \leq (v(1) + v(3)).
\]

Replacing the constraint in (26b) by that in (27), Subproblem 1 becomes the following SOCP:
\[
\min_{\hat{w}, \hat{c}} \sum_{i=1}^{N} (-\hat{w}_i \hat{c}_i) \\
\text{s.t.} \left[ v^T u_i - \hat{c} \right] \leq \hat{c}_i, \ i = 1, \ldots , N.
\]

Different from [17], in which an SOCP problem was solved twice in one iteration, the procedure proposed in this work solves the SOCP (28) only once.

After obtaining the estimate of \( v \) by solving problem (25), the kernel center and bandwidth estimates can be updated using the procedure in Section III-B1. The process repeats until the stopping criterion is reached. From the estimate of \( v \), we can recover the estimate of \( q \) according to the relationship between \( v \) and \( q \) [17].

The entire MCC-VC method for single ellipse fitting is given in Algorithm 1. It is worth noting that the MCC fitting method in [17] cannot fit the ellipse correctly in some cases owing to the predefined kernel center and bandwidth. Therefore, a clustering technique is used in [17] to detect the failed fitting. Once the failed fitting is detected, the iterations will be restarted by setting a different initialization, which obviously increases the computational load. By contrast, the proposed MCC-VC method has very rare failed fittings when the number of data points is sufficiently large and the proportion of RIDPs is generally not needed. Thus, the proposed method is easier to use. On the other hand, the proposed method may benefit from the clustering step if the number of data points is small or the proportion of RIDPs is large.

Before closing this section, we would like to discuss the weaknesses of the proposed method. First, the proposed method is iterative, which implies that local convergence or divergence, although rarely occurs under typical conditions, can happen. Another weakness is the approximations used in the proposed solution method, which may cause performance loss especially when the proportion of RIDPs is high or the number of data points is small.

C. Complexity Analysis

The proposed MCC-VC method contains an outer loop and an inner loop, where the outer loop determines the kernel center \( c \) and bandwidth \( \sigma \) while the inner loop solves the MCC problem that iteratively estimates the ellipse parameter vector \( v \).

In the estimation of kernel center \( c \) and bandwidth \( \sigma \), computations of exponential values are involved when evaluating the values of the objective functions. The complexity of the estimation of \( c \) comes mainly from the bisection method, and it is on the order of \( J_c = O(3L_1N) \), where \( N \) is the number of data points, \( L \) is the highest order of the polynomial for approximating the exponential function (\( L = 10 \) is typically sufficient), and \( L_1 \) is the number iterations when implementing the bisection method.\(^1\) The estimation of \( \sigma \) involves iteratively solving a problem having the closed-form solution expressed in (48). Computing the closed-form solution requires \( O((3L + 5)N) \) multiplications. Assuming that the number of iterations required to solve the minimum of \( f(r) \) defined in (22) is \( L_2 \) (with a typical value 20), the computational complexity of estimating the kernel bandwidth is \( J_{\sigma} = O(L_2(3L + 5)N) \). Thus, the complexity of estimating \( (c, \sigma) \) is \( J = J_c + J_{\sigma} \).

Next, let us detail the complexity required in each iteration of the inner loop. In each iteration, two subproblems are solved in an alternate manner, one for the weight vector \( w \) and the other for the ellipse parameter vector \( v \). The subproblem w.r.t. \( w \) has a closed-form solution, and hence, its complexity relative to the overall complexity is negligible. Thus, the complexity of the MCC problem comes mainly from solving an SOCP to obtain the estimate of the ellipse parameters. Generally, the worst-case complexity of solving an SOCP is on the order of \( \sqrt{\frac{m}{N}} \sum_{i=1}^{N} (m_{i}^{soc} + m_{i}^{soc}) \times \ln(1/\epsilon) \) [33], where \( m \) is the number of equality constraints, \( N \) is the number of second-order cone constraints, \( m_{i}^{soc} \) is the dimension of the \( i \)th second-order cone, \( \lambda = 2N_{soc} \) is the barrier parameter that measures the geometric complexities of the cones involved, and \( \epsilon > 0 \) is the solution precision. If we write the SOCP problem (28) into the standard primal form, there are \( N + 3 \) equality constraints, \( N \) second-order cone constraints of size 2, and one second-order cone constraint of size 4. Hence, its worst-case complexity is on the order of \( J_{socp} = O(\sqrt{2N} ((4N + 16)(N + 3) + (N + 3)^3) \ln(1/\epsilon)) \). Note that this is the worst-case complexity, and the actual complexity can be significantly less by exploiting the special structure, e.g., the sparsity of the problem [34].

Algorithm 1 The MCC-VC Method for Single Ellipse Fitting

**Input:**
- \( e^0 \): initial kernel center; \( u^0 \): initial weights;
- \( \{u_i\} \): collected data points; \( L \): maximum number of iterations;
- \( \epsilon \): constant to guarantee \( v(2) - 4v(3) < 0 \);

**Steps:**
1. Solve problem (25) to obtain an initial estimate \( \delta^0 \);
2. Solve problem (23) to obtain the estimate of \( \sigma \); \( \delta^2 \); \( \epsilon \);
3. Solve problem (20) to obtain the estimate of \( \sigma \), \( \delta^4 \);
4. Obtain the vector \( \hat{q} \) using \( \delta^4 \), according to the relationship between \( v \) and \( q \).

**Output:** The ellipse parameters \( \{\hat{q}, \hat{h}, \hat{a}, \hat{b}, \hat{\beta}\} \).

---

\(^1\) \( L_1 \) is related to the search interval and it satisfies \( L_1 > \log_{2}(\max(\delta_1, \ldots , \delta_N) - \min(\delta_1, \ldots , \delta_N))/\gamma \), where \( \gamma \) is solution precision and \( \delta_i \) for \( i = 1, \ldots , N \) are the error samples to start the iteration.
Furthermore, let the numbers of iterations in the inner and outer loops when solving the MCC-VC problem be \( L_3 \) and \( L_4 \), respectively. The complexity of solving the entire MCC-VC problem is \( O \left( L_4 \left( J_e + J_o + L_3 J_{SOPC} \right) \right) \). From their expressions, \( J_e \) and \( J_o \) are relatively insignificant compared to \( L_3 J_{SOPC} \). \( L_3 \) and \( L_4 \) have no explicit expressions and are problem specific. From our simulations, their typical values are both 4.

### IV. Coupled Ellipses Fitting With Unknown Data Association

In this section, we investigate the coupled ellipses fitting problem, where the associations between the ellipses and the data points are not available, i.e., we do not know which data point belongs to which ellipse. To overcome this difficulty, we propose a practical approach consisting of two steps: Step 1: Associations of the data points to the ellipses, and Step 2: Fitting of the associated data points to the coupled ellipses. The task of Step 1 is to estimate the association vectors \( \phi_i \) for \( i = 1, \ldots, N \), and Step 2 extends the proposed MCC-VC method to coupled ellipses fitting by using the estimated association vectors in Step 1. In Step 2, the incorrectly associated data points in Step 1 are treated as RIDPs.

#### A. Data Association

According to (11) and imposing the condition \( v_i^2 + \varepsilon^2 - 4v_i(1) v_i(3) < 0 \) to ensure the result is an ellipse, we formulate the following optimization problem to estimate the association vectors, where the coupled ellipses parameters are estimated in conjunction as:

\[
\begin{align*}
\min_{\{\phi_i\} \in \mathbb{R}^{2 \times 1}, \ v, \ \tau} & \sum_{i=1}^{N} |v_i^T u_i + \phi_i^T \tau| \\
\text{s.t.} & \ \tau_{(1)} = 0, \ (27), \\
\end{align*}
\]

where \( \prod_{2 \times 1} \) is the set of all possible \( 2 \times 1 \) association vectors.

Problem (29) is very difficult to solve owing to the integer variables and the inner product of the unknown vectors \( \phi_i \) and \( \tau \). It should be emphasized that the main purpose of problem (29) is to estimate the association vectors but not for coupled ellipses fitting, because the final coupled ellipses will be obtained in the second step. Keeping this in mind, we may still be able to correctly estimate the association vectors by making some approximations, even though the ellipse parameters may not be accurate owing to these approximations. The idea is to make the fitted inner and outer ellipses be located between the true inner and outer ellipses, which will generate correct associations for most noisy data points. To this end, we set \( \eta \) that is defined below (8) to unity such that \( \tau \) shown in (10) is a constant vector. By doing so, the inner product of the two unknown vectors becomes the product with only one unknown vector \( \phi_i \), and problem (29) is partially simplified.

The value of \( \varepsilon \) in (29) cannot be arbitrarily chosen any more since \( \eta \) is fixed to 1. Let us imagine that the association vectors can be accurately estimated when the estimated ellipses by (29) are located between the true inner and outer ellipses, i.e., the corresponding \( \mu \) should be greater than its true value. It follows from \( \eta = \beta(1 - \mu^2) = 1 \) that \( \beta \) should be large. To make this happen, we can intentionally choose a larger \( \varepsilon^2 \). Our simulation shows that \( \varepsilon \) can be chosen from a very large range, without affecting the estimation accuracy of the association vectors.

Even after setting \( \eta = 1 \), problem (29) is still a mixed integer problem and very difficult to solve. It will become more tractable if we relax each association vector \( \phi_i \) into a probability vector characterized by \( 0 \leq \phi_{i,j} \leq 1, \sum_{j=1}^{2} \phi_{i,j} = 1 \), where \( \phi_{i,j} \) is the \( j \)-th element of \( \phi_i \).

By doing so, problem (29) can be relaxed to the form

\[
\begin{align*}
\min_{\{\phi_i\} \in \mathbb{R}^{2 \times 1}, \ v, \ \{\beta_i\}} & \sum_{i=1}^{N} \beta_i \\
\text{s.t.} & \ \phi_{i,j} \leq 1, \sum_{j=1}^{2} \phi_{i,j} = 1, \ i = 1, \ldots, N; \\
& |v_i^T u_i + \phi_i^T \tau| \leq \beta_i. \ (27) \\
\end{align*}
\]

Problem (30) is a convex SOCP problem, which can be solved using off-the-shelf software. Let us represent the solution of (30) as \( \hat{\phi}_i \). An estimate of the association vector, denoted by \( \tilde{\phi}_i \), can be obtained by setting the larger element of the vector \( \hat{\phi}_i \) to 1 and the smaller element to 0. It is possible that the case of \( \hat{\phi}_i \) equal to \([0.5, 0.5]^T \) occurs, which may lead to wrong data association for the corresponding point. Nonetheless, these points tend to be those located in the middle between the inner and outer ellipses, and the number of such points is typically very small. On the other hand, if wrong data association occurs, the incorrectly associated data points can be interpreted as RIDPs. Owing to the robustness of the proposed MCC-VC method, the effect of the wrong associations is minor.

After obtaining \( \hat{\phi}_i \), the robust ellipse fitting methods can be used to obtain the ellipse parameters. In the next subsection, we shall present a new coupled ellipse fitting method by extending the proposed MCC-VC method.

#### B. Coupled Ellipses Fitting Based on MCC-VC

Replacing \( \phi_i \) with \( \hat{\phi}_i \) in (11), the model equation becomes:

\[
\phi_i^T \delta_i = v_i^T u_i + \phi_i^T \tau, \ i = 1, \ldots, N. \ (31)
\]

The model in (31) may not be ideal owing to the fact that \( \hat{\phi}_i \) may not be equal to the true value \( \phi_i \).

By introducing \( \tilde{v} = [v^T \eta]^T \) and

\[
\tilde{u}_i = \begin{cases} 
[u_i^T 0]^T & \text{if } \phi_{i,1} = [1, 0]^T; \\
[u_i^T 1]^T & \text{if } \phi_{i,1} = [0, 1]^T,
\end{cases}
\]

the model (31) becomes the concise form that \( \tilde{\phi}_i^T \tilde{\delta}_i = \tilde{v}^T \tilde{u}_i, \ i = 1, \ldots, N \).

By letting \( \tilde{\delta}_i = \tilde{v}^T \tilde{u}_i \) be the error vector and \( \tilde{v} \) be the unknown variable vector, respectively, we can similarly formulate the optimization problems based on MCC-VC, and the resulting problems can be solved in a similar manner as the single ellipse fitting case. The parameter estimates of the two ellipses can be recovered from the solution of the MCC-VC problem as in [17].

### V. Numerical Results

This section verifies the robustness of the proposed MCC-VC method for single ellipse and coupled ellipses.
fittings using both simulated data and real images. In order to test the single ellipse fitting performance, we compare the proposed MCC-VC method using the Laplacian kernel (denoted by “MCC-VC-Laplacian”) with the RANSAC method having LS as the fitting algorithm (denoted by “RANSAC+LS”) [15], and the RANSAC method having MCC-VC-Laplacian as the fitting algorithm (denoted by “RANSAC+(MCC-VC)”). The maximum number of iterations for both RANSAC methods is set to 1000. The other methods included for comparison are the SAREfit method (denoted by “SAREfit”) [35], the MCC method using the Gaussian kernel (denoted by “MCC-Gaussian”) [16], the MCC method using the Laplacian kernel (denoted by “MCC-Laplacian”) [17], the HGMM method (denoted by “HGMM”) [18], the Szpak method (denoted by “Szpak”) [36], the least median squares method (denoted by “LMS”) [14], and the spatial median consensus based fitting method (denoted by “SMC”) [19]. For coupled ellipses fitting, we first test the data association performance using the percentage of successful data associations, and then further compare the fitting performance of the proposed MCC-VC-Laplacian and the MCC-Laplacian [17] methods. The associated SOCP problems in MCC-Laplacian and MCC-VC-Laplacian are solved using the toolbox “ECOS” [37] and the SDP problem in MCC-Gaussian is solved using the Matlab toolbox “CVX” [38], where the solver is SeDuMi [39].

A. Single Ellipse Fitting: Simulated Data

In the following, we randomly generate the simulated ellipses for fitting as in [17]. The true five ellipse parameters are set according to the following distributions: $g \sim \mathcal{U}(0, 20)$, $h \sim \mathcal{U}(0, 20)$, $b \sim \mathcal{U}(10, 50)$, $a \sim \mathcal{U}[b + 5, 55]$, and $\theta \sim \mathcal{U}(-90^\circ, 90^\circ)$. The noise follows the zero-mean Gaussian distribution with variance $(0.005b)^2$. We use three metrics for performance evaluation, i.e., the normalized root mean square error (NRMSE), the Hausdorff distance [40], and the Euclidean ellipse comparison metric (EECM) [41]. The definitions of the Hausdorff distance and EECM are given in [40] and [41], respectively, and the NRMSE is defined by $\text{NRMSE} = \sqrt{\frac{1}{KM} \sum_{k=1}^{K} \sum_{m=1}^{M} \Vert \hat{q}_{mk} - q_k \Vert^2}$, where $K$ and $M$ are the numbers of the generated ellipses and the Monte Carlo (MC) runs for each ellipse, and $\hat{q}_{mk}$ and $q_k$ represent the estimated and true parameters of the $k$th ellipse in the $m$th MC run. In the following, we set $K = 100$ and $M = 500$ to compute the NRMSE.

Due to the existence of RIDPs, the fitting possibly fails in a few runs. To obtain meaningful NRMSE, Hausdorff distance, and EECM, they are computed by discarding the results of these runs. Several conditions are used to identify the failed runs. Here, we adopt the same rule as in [17] to identify a failed fitting. In the following, we assume that RIDPs are originated from outliers, grouped noise, or both.

1) Scenario 1 - Outliers Only RIDPs: In this scenario, we assume that RIDPs are outliers only. The total number of data points is $N = 100$, and the proportion of outliers varies from 5% to 20%. The values $v_i^*$ and $v_i$ for the outliers are uniformly generated according to $\mathcal{U}(-2b, 2a)$ for the simulated scenarios. Figs. 1(a)-(c) show the NRMSE, Hausdorff distance, and EECM performance as the proportion of outliers increases, respectively. Generally speaking, the proposed MCC-VC-Laplacian method performs better than the other methods, including the MCC-Laplacian method, especially when the proportion of outliers increases from 10% to 20%. Correspondingly, Table I shows the rates of successful fittings for different methods. In this simulation scenario, the proposed MCC-VC-Laplacian method almost always successfully fits the ellipse even when the proportion of outliers is 20%. By contrast, the other methods may fail, especially when the proportion of outliers is large. As an illustration, Fig. 1(d) shows the ellipses fitted by the ten methods in a typical MC run when the proportion of RIDPs is 20%. It can be seen that the ellipse generated by MCC-VC-Laplacian fits the true ellipse very well, and better than the other methods. MCC-Gaussian performs much worse than the other two MCC based methods. It is likely resulting from the fact that the subproblem w.r.t. the ellipse parameters in MCC-Gaussian is approximately solved by applying SDR, which may cause divergence since the relaxed SDP by applying SDR cannot reach the optimal solution of the original subproblem. This, in turn, causes significant performance loss. Interestingly, after replacing the fitting algorithm in the RANSAC method with MCC-VC-Laplacian (RANSAC+(MCC-VC)), it is clear from in Fig. 1 and Table I that its performance is significantly improved compared with the original RANSAC method (RANSAC+LS), which even completely fails when the proportion of outliers is 20%. This further indicates that the good performance of the proposed method. Some observations can be found in the following experiments.

2) Scenario 2 - Grouped Noise Contaminated RIDPs: In this scenario, we assume that RIDPs are contaminated by grouped noise, and these points are randomly distributed around the true ellipse. Specifically, we generate five set of grouped noise contaminated RIDPs. The center for each set is first generated similarly to the outliers in Scenario 1, and the other points of this set are then randomly generated in the $15 \times 15$ square area located at the center. Each set has the same number of RIDPs. The grouped noise pattern varies for each of the 100 randomly generated ellipses and in each MC run. Considering the characteristics of grouped noise in practice, we vary the proportion of grouped noise contaminated points from 20% to 50%. The NRMSE, Hausdorff distance, and EECM are shown in Figs. 2(a), (b), and (c), respectively, and the rates of successful fittings are given in Table II. Although HGMM has better fitting performance than MCC-VC-Laplacian in NRMSE, Hausdorff distance, and EECM when the proportion of RIDPs is 50%, its successful fitting rate is far from satisfactory. In general, the proposed

| Method                  | Prop. (%) | 5    | 10   | 15   | 20   |
|-------------------------|----------|------|------|------|------|
| MCC-VC-Laplacian        | 99.97    | 99.36| 98.92| 98.36|      |
| MCC-Laplacian           | 98.19    | 97.48| 96.18| 92.74|      |
| MCC-Gaussian            | 81.65    | 59.08| 48.05| 38.99|      |
| SAREfit                 | 92.59    | 94.39| 92.62| 84.44|      |
| HGMM                    | 60.91    | 25.36| 22.66| 21.89|      |
| Szpak                   | 84.90    | 66.69| 52.18| 33.35|      |
| RANSAC+LS               | 2.47     | 0.17 | 0.02 | 0     |      |
| RANSAC+(MCC-VC)         | 89.84    | 85.73| 83.36| 76.02|      |
| LMS                     | 17.09    | 10.89| 2.89 | 0.89 |      |
| SMC                     | 92.14    | 20.10| 15.81| 11.79|      |
MCC-VC-Laplacian still significantly outperforms the other methods. Comparison between the results for the outliers and the cluster-like grouped noise contaminated RIDPs reveals that the latter will result in larger fitting errors and higher rate of failed fittings, indicating that the ellipse is more difficult to fit in this scenario. Fig. 2(d) illustrates the fitting results in a typical MC run when the proportion of RIDPs is 40%, which also shows the distribution of RIDPs. It clearly illustrates the better fitting performance of the proposed method.

3) Scenario 3 - Mixed RIDPs: In this scenario, we simulate a more general case in real applications, where both outliers and grouped noise contaminated RIDPs are present. The two kinds of RIDPs are generated in similar manners as in Scenarios 1 and 2. We vary the proportion of RIDPs from 10% to 80%, where the ratio of outliers to grouped noise contaminated RIDPs is 2:3 when the proportion is below 50%, and the proportion of outliers does not increase when it reaches 20%. The NRMSEs, Hausdorff distance, and EECM are shown in Figs. 3(a), (b), and (c), respectively. Fig. 3(d) illustrates the fitting results in a typical MC run when the proportion of RIDPs is 45%, which also shows the distribution of RIDPs. It clearly demonstrates the better fitting performance of the proposed method.

In order to gain some insight about the accuracy of the approximations applied to the proposed solution method when estimating the kernel bandwidth and center, we include the performance of the proposed method when solving the sub-problems w.r.t. the kernel bandwidth and center by grid search (denoted by “MCC-VC-Grid Search”). It is well known that the grid search method is able to reach the global solution of a non-convex problem when the grid size is sufficiently small. Hence, its performance can be used as a benchmark. The search intervals for the kernel center $c$ and the kernel bandwidth $\sigma$ are $[-10, 10]$ and $[0.01, 20]$, respectively.

Table II

| Method         | $\sigma$ (%) | 20   | 30   | 40   | 50   |
|----------------|--------------|------|------|------|------|
| MCC-VC-Laplacian | 100          | 38.95| 37.61| 35.34| 33.34|
| MCC-Laplacian   | 95.44        | 31.55| 30.97| 29.53| 28.48|
| MCC-Gaussian    | 63.31        | 30.55| 29.31| 28.25| 27.26|
| SAREfit         | 73.17        | 38.22| 37.24| 36.27| 35.37|
| RMSMM           | 17.12        | 11.16| 11.14| 11.12| 11.10|
| Sjöpik          | 70.05        | 50.05| 38.99| 35.48| 33.34|
| RANSAC+L+S      | 13.57        | 7.47 | 7.22 | 7.13 | 7.05 |
| RANSAC+MCC-V    | 20.10        | 79.22| 66.15| 60.05| 55.00|
| LMS             | 29.48        | 17.12| 16.61| 15.42| 14.42|
| SMC             | 56.66        | 24.32| 20.66| 18.72| 16.72|

Fig. 2. Performance of different fitting methods as the proportion of outliers varies. Fig. 3 includes the performance of MCC-VC-Grid Search for comparison, which shows that the proposed solution method performs very closely to the grid search method unless the proportion of RIDPs is very high, implying the approximations in the proposed solution method are sufficiently accurate. The grid search method, however, uses the brute force approach that is very computationally intensive. Furthermore, it operates much less efficiently and its accuracy is closely related to the grid size.

Fig. 4(d) illustrates the fitting results in a typical MC run when the proportion of RIDPs is 70%, the rate of successful fittings drops quickly to below 10% to 80%, where the ratio of outliers to group noise contaminated RIDPs is set to 2:3. It is demonstrated in Figs. 4(a)-(c) that the proposed MCC-Laplacian method remains to have best fitting accuracy.

Table IV lists the average run times of the compared methods when the proportion of RIDPs is 30%. Evidently, the proposed MCC-VC-Laplacian method is computationally competitive. Among the nine other algorithms tested, only SAREfit, RANSAC+L+S, and LMS have lower complexities but they also have worse fitting accuracy. Note that MCC-Laplacian consumes more time than MCC-VC-Laplacian owing to the fact that the SOCP problem is solved twice and the clustering procedure is included in MCC-Laplacian.

4) Scenario 4 - Number of Data Points: In this scenario, we vary the number of data points from 10 to 90 in the presence of 30% RIDPs to test the effect of the number of data points on the performance of various fitting methods. The ratio of outliers to group noise contaminated RIDPs is 2:3. It is demonstrated in Figs. 4(a)-(c) that the proposed MCC-Laplacian method remains to have best fitting accuracy. Fig. 4(d) illustrates the fitting results in a typical MC run when the number of data points is 10. The performance of MCC-VC-Grid Search is also included in Fig. 4 for comparison, and it shows that the proposed solution method has slight performance loss when the number of data points is small. Except MCC-VC-Grid Search, Table V indicates that the proposed method has a higher successful fitting rate than the other methods. Although RANSAC+L+(MCC-VC) has a higher successful fitting rate as well, its estimation accuracy is far less satisfactory in Fig. 4. The proposed method is able to have more than 50% successful fittings with as few as 30 data points at 30% RIDPs. This demonstrates the applicability of the proposed method when a very small number of data points is available.

$^3$The proposed method is implemented using Matlab R2014b on a PC with a 3.6 GHz Intel CPU and 16 GB RAM.
Fig. 2. Performance comparison for different fitting methods as the proportion of grouped noise contaminated RIDPs varies.

Fig. 3. Performance comparison for different fitting methods as the proportion of RIDPs varies.

Fig. 4. Performance comparison for different fitting methods as the number of data points varies.

5) Scenario 5 - Occlusion Level: In practice, the ellipse can be occluded and only the data points of one or several segments appear. Here, occlusion refers to the situation where some portions of the true ellipse are missing and the data points are extracted from the remaining ellipse segments. In this scenario, we test the effect of occlusion on the fitting methods, where the occlusion level is defined as the proportion of missing arc segments to the complete ellipse. We change the occlusion level from 30% to 60% in the presence of 30% RIDPs where the ratio of outliers to grouped noise contaminated RIDPs is 2:3. The results as the occlusion level varies are given in Fig. 5 and Table VI. Fig. 5(d) illustrates the fitting results in a typical MC run when the occlusion level is 40%. It is seen from the results that the proposed method still performs better than the compared methods in terms of the four performance measures, indicating its good performance in dealing with this difficult scenario.

B. Coupled Ellipses Fitting: Simulated Data

In the coupled ellipses fitting, the ellipse parameters \( \{g, h, a, b, \theta\} \) are generated in the same way as the single ellipse fitting case. The proportional parameter \( \mu \) is created randomly according to the uniform distribution \( \mu \sim \mathcal{U}(0, 1) \). 100 data points are collected for each ellipse and hence, the number of total data points is \( N = 200 \), with RIDPs possibly included. The edge data points are generated by uniformly sampling over the ellipses, and RIDPs are simulated according to the uniform distribution \( \mathcal{U}(-b, b) \). The associations between the data points and the coupled ellipses are not available to the fitting.

Due to the possible incorrect association of data points in Step 1 and the presence of RIDPs, we take the same measures as mentioned before to discard the failed fitting results when obtaining the performance metrics NRMSE, Hausdorff distance and EECM.
Fig. 5. Performance comparison for different fitting methods as the occlusion level varies.

TABLE V
RATE (%) OF SUCCESSFUL FITTINGS FOR SINGLE ELLIPSE (FROM 50000 MC RUNS): NUMBER OF DATA POINTS

| Method                  | 10  | 30  | 50  | 70  | 90  |
|-------------------------|-----|-----|-----|-----|-----|
| MCC-CV-Grid Search     | 23.39 | 35.58 | 38.44 | 60.50 | 85.21 |
| MCC-CV-Laplacian        | 13.83 | 50.24 | 55.44 | 57.50 | 64.21 |
| MCC-Genus            | 5.64 | 14.14 | 26.72 | 45.01 | 72.01 |
| MCC-Gaussian          | 5.72 | 8.04 | 27.72 | 35.25 | 46.35 |
| SAREE                  | 0.28 | 18.25 | 24.22 | 37.71 | 55.03 |
| HOPE                   | 0.01 | 1.62 | 15.45 | 34.27 | 43.60 |
| Spah                   | 0.01 | 6.66 | 24.73 | 28.92 | 33.34 |
| SPAN                   | 7.72 | 5.57 | 10.49 | 14.40 | 22.21 |
| RANSAC+MC-Vc            | 11.41 | 68.66 | 83.11 | 88.85 | 89.49 |
| RANSAC+MC-Laplacian    | 11.41 | 68.66 | 83.11 | 88.85 | 89.49 |
| LMS                    | 4.37 | 5.52 | 22.85 | 41.50 | 51.47 |
| SMC                    | 3.98 | 26.18 | 36.07 | 44.78 | 80.08 |

TABLE VI
RATE (%) OF SUCCESSFUL FITTINGS FOR SINGLE ELLIPSE (FROM 50000 MC RUNS): OCCLUSION LEVEL

| Method                  | 30  | 50  | 60  |
|-------------------------|-----|-----|-----|
| MCC-CV-Laplacian        | 99.89 | 95.33 | 87.65 |
| MCC-Genus            | 49.36 | 57.91 | 19.92 |
| MCC-Gaussian          | 36.82 | 18.01 | 30.55 | 13.47 |
| SAREE                  | 32.97 | 14.26 | 3.28 |
| HOPE                   | 10.62 | 3.31 | 8.77 |
| Spah                   | 8.12 | 7.25 | 3.51 | 0.55 |
| SPAN                   | 0.30 | 0.32 | 0.32 |
| RANSAC+MC-Vc            | 72.81 | 38.36 | 33.43 |
| RANSAC+MC-Laplacian    | 72.81 | 38.36 | 33.43 |
| LMS                    | 11.79 | 3.78 | 8.25 | 0.03 |
| SMC                    | 9.16 | 6.02 | 2.81 | 0.92 |

Fig. 6. Sensitivity of the value of $\epsilon$ on the association accuracy.

1) Effect of $\epsilon$ on Association: As aforementioned, $\epsilon$ cannot be arbitrarily chosen owing by fixing $\eta$ to 1 during the data association step. In this experiment, we study the effect of choosing different values of $\epsilon$ on data association. The percentages of incorrect associations of both edge data points and RDPs are examined, as the proportion of RDPs increases from 0 to 40%, when $\epsilon$ takes the values of 1, 10, 100, 1000 or 10000. Again, the ratio of outliers to group noise contaminated RDPs is set to 2:3. The results are shown in Fig. 6, which indicates that the choice of $\epsilon$ has nearly no effect on the association vector estimation, although the ellipse parameter vector may not be accurately estimated. In the following, $\epsilon$ is set to 1.

2) Performance of Coupled Ellipses Fitting by Varying the Proportion of RDPs: In this experiment, we consider the scenario that RDPs are present inherently in the data points. The noise standard deviation is fixed at 0.005b, and the proportion of RDPs varies from 10% to 40%, where the ratio of outliers to group noise contaminated RDPs is still set to 2:3. The results are given in Table VII.

The results remain encouraging. Let us take the last column as an example. Taking the incorrectly associated points into account, the actual proportion of RDPs is greater than 40% under this setting. Even in this challenging case, the rate of successful fitting is still greater than 97.7%, implying that the proposed data association method is quite effective for use in a practical harsh environment. Moreover, the proposed MCC-VC method outperforms MCC in terms of the three fitting accuracy metrics, where the kernel is Laplacian for both.

C. Single Ellipse Fitting: Real Data

In this subsection, we apply ten methods to fit the ellipses in real images, including the Voyager aircraft, Mars, globe, basketball rim, and butterfly frame images. Before the fitting, the data points of these images are extracted through a series of preprocessing steps including the image segmentation, the morphological operations, and the edge detection. It is not difficult to imagine there are a large number of RDPs in the extracted data points, and RDPs do not necessarily follow a zero-mean distribution.

1) Voyager Aircraft: The Voyager aircraft image [42] fitting process and results are shown in row (a) of Fig. 7, in which the proportion of RDPs is about 21.47%. The proportion is larger than that in [17], which is generated by setting the Sobel operator parameter to a smaller value of 0.17 as compared to 0.2 in [17]. The fitting results of the MCC-VC-Laplacian method and the other methods are shown in the third column and fourth column of row(a), respectively. Obviously, MCC-Laplacian [17] fails to fit the ellipse but MCC-VC-Laplacian is still successful and more robust to larger amount of RDPs.

2) Mars: The Mars image is selected from the Caltech 256 dataset [43], labeled as 137_0008. The fitting process and results are shown in row (b) of Fig. 7. The Sobel operator parameter is set to 0.042 for this image to include more RDPs as compared to performance evaluation purpose. The proportion of RDPs is about 54.51%. The fitting results of MCC-VC-Laplacian and the other methods are shown in the
third column and fourth column, of row (b), respectively, we see that the proposed method successfully fits the ellipse but the others fail.

3) Globe: The globe image is selected from [43], labeled as 053_0080. The fitting performance is shown in row (c) of Fig. 7. Different from the previous two images, the Canny detector is used for edge extraction of this image, where the correlation coefficient is set to 0.5. RIDPs in the extracted data points form a one-sided distribution, due to the base holding the globe. The proportion of RIDPs is about 33.27%. For this image, the proposed MCC-VC-Laplacian method remains to have the best fitting performance compared to the other methods, as shown in the third column and fourth column of row (c), confirming the robustness of the proposed method to one-sided RIDPs.

4) Basketball Rim: The basketball rim image is selected from the Caltech 256 dataset [43], labeled as 006_0033. The performance is shown in row (d) of Fig. 7. The Canny detector is also used for edge extraction of this image, where the correlation coefficient is set to 0.5. The presence of the damaged basketball net not only introduces a large amount of grouped noise contaminated RIDPs that are distributed in one side of the ellipse, but also causes a missing arc segment, resulting in a more difficult case for ellipse fitting. The proportion of RIDPs in this fitting scenario is about 31.44%. For this image, the proposed method also has the best fitting performance compared to the other methods.

5) Butterfly Frame: The butterfly frame image is also selected from the Caltech 256 dataset [43], labeled as 024_0103. The fitting performance and results are shown in row (e) of Fig. 7. The correlation coefficient of the Canny detector is set to 0.3. The presence of the butterfly and flower inside the frame introduces a large number of concentrated RIDPs, whose proportion is about 39.64%. From the fitting results, in the third column and fourth column of row (e), the proposed method fits the ellipse very well. However, all of the other methods have more or less fitting performance loss caused by RIDPs.

D. Coupled Ellipses Fitting: Real Data

In this subsection, the proposed method is applied to fit the coupled ellipses in an iris image. To demonstrate the robustness of the proposed method, two scenarios without and with RIDPs in the data points are investigated. The data points extracted from the iris image in both scenarios are shown in Figs. 8 (a) & (d). In the case of having RIDPs, RIDPs account for 24.98% of the total data points. The results of the data association and the fitting are illustrated in Figs. 8 (b) & (e) and Figs. 8 (c) & (f), respectively. In Figs. 8 (b) & (e), the blue dots and the red stars represent the data points associated with the inner and outer ellipses, respectively. Clearly, there exist some incorrectly associated points, and they are regarded as RIDPs. Hence, the percentage of actual RIDPs is greater than 24.98% when doing the fitting. Nonetheless, the fitting is still successful, indicating the incorrectly associated data points are handled without difficulty by the fitting method in Step 2.

E. Multiple Ellipses Fitting

In this subsection, we show that the proposed method has the ability of accurately fitting multiple ellipses, with the help of ellipse detection. Most ellipse detection methods [26], [27], [28], [29] involve ellipse fitting of sampled data points. Specifically, the principle of many detection methods is to detect arcs by the position relationship among pixel points. After grouping or deleting the arcs through basic fitting methods such as the direct LS method [8], the candidate ellipses are formed, and then the qualified ellipses are selected from the candidates through screening. The detection methods focus on the detection performance, and their fitting performance may not be satisfactory. Our aim is to first apply the ellipse detection method to find the ellipses in a given image, and then further improve the fitting performance based on the coarse position and size information of the detected ellipses. More specifically, the idea is to divide the data points into groups, each corresponding to one detected ellipse. To this end, we first form the equations of the detected ellipses using the parameters of the detected ellipses obtained from the detection methods. We then apply the coordinates of each data point to the equations and compute the errors of each data point relative to all ellipse equations. For each ellipse, we can extract the data points with relatively small errors by setting a threshold. By doing so, the extracted points belong to the ellipse with a high probability. This process can be regarded as association, i.e., associate the data points to particular ellipses. Apparently, RIDPs may be introduced in the association process, and the value of the threshold determines the number of RIDPs. The smaller the threshold, the less the number of RIDPs, which, however, may filter out useful points.

To validate the fitting performance improvement by the proposed method, we select several images, each containing multiple ellipses. We first use Lu’s detection method proposed recently in [27] to detect the ellipses and then apply
ellipse fitting methods to further fit the ellipses. We conducted corresponding comparative experiments, from the data points extracted after detection as input, by using the proposed MCC-VC-Laplacian, MCC-Laplacian [17], HGMM [18], SAREfit [35], for fitting. All of them use the same set of data points from ellipse detection. The results are shown in Fig. 9, where the threshold is set to 1.5. For comparison, the original fitting results from Lu’s detector are also included in the second column of Fig. 10 that Lu’s method cannot detect the ellipses owing to some losing portions. In comparison, the proposed coupled ellipses fitting method successfully fits the coupled ellipses. Fig. 11(a) shows an owl’s iris image containing coupled ellipses. For this image, Lu’s detection method is only able to detect the inner ellipse; see Fig. 11(c). The data points associated to the inner ellipse can be determined using the detected inner ellipse, and the remaining data points are associated to the outer ellipse. The association results based on detection and using the proposed SOCP method are given in Figs. 11(e) & (g), respectively. The proposed MCC-VC coupled ellipses fitting method successfully fits the coupled ellipses. Figs. 11(f) & (h), respectively. Clearly, the proposed SOCP method for association has better fitting performance.

Before closing this subsection, we would like to make the following discussion about the use of the ellipse detection method in multiple ellipses fitting. When there are multiple ellipses in the image, we need to resort to some other techniques to roughly divide the overall data points into several subsets, each subset of data points corresponding to one ellipse. Ellipse detection is one of such techniques. It should be noted that ellipse detection is not the only technique that can complete this task. Hence, to be exact, the proposed method does not depend on ellipse detection, but requires a pre-processing step of dividing the data points into subsets for different ellipses. Besides ellipse detection, the traditional k-means method and the machine learning method are able to complete the data pre-processing task. The k-means method has been used in [44], and an example of machine learning method can be found in [45]. Moreover, we would like to clarify that the proposed ellipse fitting method does not require ellipse detection when there is only one ellipse or there are only coupled ellipses in the image. For example, the fittings

![Fig. 9. Fitting performance comparison for multiple ellipses fitting after ellipse detection. (1) The first column of images are synthetic images; (2) The second column shows the fitting results from Lu’s detection method; (3) The last column shows the fitting results of the proposed coupled ellipses fitting method.](image)

| Method          | NRMSE | Hausdorff | EECD | Rate (%) |
|-----------------|-------|-----------|------|----------|
| Detection by Lu | 11.91 | 13.03     | 3.19 | 72.83    |
| MCC-Laplacian   | 0.44  | 4.90      | 3.53 | 93.83    |
| MCC-Laplacian   | 0.63  | 6.36      | 2.40 | 63.89    |
| MCC-Laplacian   | 0.89  | 8.23      | 6.13 | 40.38    |
| MCC-Laplacian   | 9.31  | 9.05      | 7.09 | 31.75    |
| HGMM            | 9.95  | 9.65      | 3.34 | 36.20    |
| SAREfit         | 0.16  | 0.79      | 3.15 | 58.85    |
| RANSAC-LLS      | 17.11 | 17.09     | 7.51 | 40.80    |
| RANSAC-MCC-VC   | 15.37 | 15.73     | 1.90 | 40.40    |
| LMS             | 2.22  | 6.95      | 3.44 | 41.87    |
| SMC             | 0.76  | 8.56      | 3.00 | 51.47    |
in Fig. 7 do not apply ellipse detection first since there is no need to divide the data points into subsets.

VI. CONCLUSION

In this paper, we have presented a new ellipse fitting method based on MCC-VC, which offers strong robustness against RIDPs. By the iterative optimization of the kernel center, kernel bandwidth and ellipse parameter vector, one at a time through convex approximations, the proposed MCC-VC method is computationally efficient and more flexible, and is applicable to more challenging scenarios. Furthermore, we have proposed a data association method for coupled ellipses fitting without knowing the data association between the ellipses and the data points, and extended the proposed MCC-VC single ellipse fitting method to coupled ellipses fitting. Both simulated data and real images have confirmed the superior fitting performance of the proposed methods over several competitive fitting methods in the literature.

APPENDIX A

APPROXIMATE CONVEX FUNCTION OF h(r)

By introducing a known positive constant \( r_0 \), we can rewrite \( h(r) \) as \( h(r) = h_1(r) + h_2(r) \), where

\[
h_1(r) = -\frac{r - r_0}{N} \sum_{i=1}^{N} e^{-\hat{a}_i r}, \quad h_2(r) = \frac{r}{4} - \frac{r_0}{N} \sum_{i=1}^{N} e^{-\hat{a}_i r}.
\]

To approximate \( h(r) \) by a convex function, we approximate the two non-convex functions \( h_1(r) \) and \( h_2(r) \) to convex functions through appropriate Taylor expansions, respectively. For the function \( h_1(r) \), we perform the Taylor expansion to \( e^{-\hat{a}_i r} \) up to the third order at \( r_0 \), giving

\[
e^{-\hat{a}_i r} \approx e^{-\hat{a}_i r_0} - \hat{a}_i e^{-\hat{a}_i r_0} (r - r_0) + \frac{\hat{a}_i^2}{2} e^{-\hat{a}_i r_0} (r - r_0)^2 - \frac{\hat{a}_i^3}{6} e^{-\hat{a}_i r_0} (r - r_0)^3.
\]

The value of \( \hat{a}_i \) may be large in the presence of RIDPs, and keeping up to the third-order in the expansion is to guarantee a sufficiently accurate approximation.

Using (34) in \( h_1(r) \) yields the approximate function \( f_1(r) \):

\[
f_1(r) = -\frac{r - r_0}{N} \sum_{i=1}^{N} e^{-\hat{a}_i r_0} - \hat{a}_i e^{-\hat{a}_i r_0} (r - r_0) + \frac{\hat{a}_i^2}{2} e^{-\hat{a}_i r_0} (r - r_0)^2 - \frac{\hat{a}_i^3}{6} e^{-\hat{a}_i r_0} (r - r_0)^3.
\]

For notational simplicity, we further define

\[
b_1 = \frac{1}{N} \sum_{i=1}^{N} e^{-\hat{a}_i r_0}, \quad b_2 = \frac{1}{N} \sum_{i=1}^{N} \hat{a}_i e^{-\hat{a}_i r_0},
\]

\[
b_3 = \frac{1}{N} \sum_{i=1}^{N} \hat{a}_i^2 e^{-\hat{a}_i r_0}, \quad b_4 = \frac{1}{N} \sum_{i=1}^{N} \hat{a}_i^3 e^{-\hat{a}_i r_0},
\]

which are known constants. With these notations, \( f_1(r) \) can be rewritten as

\[
f_1(r) = -\frac{b_4}{6} (r - r_0)^4 - \frac{b_3}{2} (r - r_0)^3 + b_2 (r - r_0)^2 - b_1 (r - r_0).
\]

Regarding the convexity of \( f_1(r) \), we have the following proposition.

Proposition 1: \( f_1(r) \) is a strictly convex function in the domain \((0, +\infty)\).

Proof: By letting \( t = r - r_0 \), we form a function \( f_1(t) \). To prove the convexity of \( f_1(r) \), we first prove that \( f_1(t) \) is strictly convex. \( f_1(t) \) is a strictly convex function if and only if the second derivative is strictly greater than zero, i.e., \( f''_1(t) > 0 \) [31]. It follows from the expression of \( f_1(t) \) that

\[
f''_1(t) = 2b_4 t^2 - 3b_3 t + 2b_2,
\]

which is a quadratic function. Since \( b_4 > 0 \), \( f''_1(t) \) is convex and has a minimum. The minimum value of \( f''_1(t) \) is

\[
\min_{t=r-r_0} f''_1(t) = \frac{b_2 b_4 - \frac{9}{16} b_3^2}{\frac{1}{2} b_4},
\]

and it is reached at \( t = \frac{3b_3}{4b_4} \).

Since \( \frac{b_4}{b_3} > 0 \), we only need to show the numerator term \( b_2 b_4 - \frac{9}{16} b_3^2 > 0 \). Using the expressions of \( b_2, b_3, b_4 \) in (36),

\[
\begin{align*}
b_2 b_4 - \frac{9}{16} b_3^2 &= \frac{1}{N} \sum_{i=1}^{N} \hat{a}_i e^{-\hat{a}_i r_0} \sum_{i=1}^{N} \hat{a}_i^2 e^{-\hat{a}_i r_0} - \frac{9}{16} N^2 \left( \sum_{i=1}^{N} \hat{a}_i^2 e^{-\hat{a}_i r_0} \right)^2. \quad (38)
\end{align*}
\]

According to the Cauchy-Schwarz inequality, we have

\[
\begin{align*}
&\sum_{i=1}^{N} \hat{a}_i^4 e^{-\hat{a}_i r_0} + \sum_{i=1}^{N} \hat{a}_i^2 e^{-\hat{a}_i r_0}^2 \\
&\geq \left( \sum_{i=1}^{N} \hat{a}_i^4 e^{-\hat{a}_i r_0} \right) \left( \sum_{i=1}^{N} \hat{a}_i^2 e^{-\hat{a}_i r_0}^2 \right) \\
&= \left( \sum_{i=1}^{N} \hat{a}_i^2 e^{-\hat{a}_i r_0} \right)^2 \geq \frac{9}{16} \left( \sum_{i=1}^{N} \hat{a}_i^2 e^{-\hat{a}_i r_0} \right)^2. \quad (40)
\end{align*}
\]

The last inequality holds when \( \sum_{i=1}^{N} \hat{a}_i^2 e^{-\hat{a}_i r_0} \neq 0 \), which obviously is the case since \( \hat{a}_i = |\hat{\delta}_i - \hat{c}| \), and \( \hat{\delta}_i \) (\( i = 1, \ldots, N \)) are random samples and cannot have the same value of \( \hat{c} \). We conclude from (41) that \( f''_1(t) > 0 \) holds, indicating that \( f_1(t) \) is a strictly convex function. Since \( f_1(r) \) is the composition of the convex function \( f_1(t) \) and the affine function \( t = r - r_0 \), it is also strictly convex.

Next, we will focus on the approximation of function \( h_2(r) \). For \( e^{-\hat{a}_i r} \) in \( h_2(r) \), we only perform the first-order Taylor expansion at \( r_0 \), since it is difficult to prove the convexity of the higher order approximations. By doing so, we obtain the approximate function \( f_2(r) \) of \( h_2(r) \):

\[
f_2(r) = \frac{r}{4} - \frac{r_0}{N} \sum_{i=1}^{N} \hat{a}_i e^{-\hat{a}_i r_0} (r - r_0).
\]

which is an affine function of \( r \) and thus convex.
Finally, the approximate convex function of $h(r)$ is
\[ f(r) = f_1(r) + f_2(r) \]
\[ = \frac{b_4}{6} (r - r_0)^4 - \frac{b_3}{2} (r - r_0)^2 + b_2 (r - r_0)^2 \]
\[ + (b_2 r_0 - b_1 + \frac{1}{4} (r - r_0) + \frac{1}{4} r_0 - b_1 r_0). \] (43)

**APPENDIX B**

**CLOSED-FORM SOLUTION OF $f'(r) = 0$**

In this appendix, we show the existence and uniqueness of the real root of $f'(r) = 0$, and give the expression of the closed-form solution.

Similar to Appendix A, we first obtain the function $f(t)$ by letting $t = r - r_0$. $f'(t)$ can be expressed as
\[ f'(t) = \frac{2b_4}{3} t^3 - \frac{3b_3}{2} t^2 + 2b_2 t + (b_2 r_0 - b_1 + \frac{1}{4}). \] (44)

By defining the following notations,
\[ d_1 = \frac{2b_4}{3}, \quad d_2 = -\frac{3b_3}{2}, \quad d_3 = 2b_2, \quad d_4 = b_2 r_0 - b_1 + \frac{1}{4}, \]
\[ p = \frac{3d_1 d_3 - d_2^2}{3d_1^2}, \quad q = \frac{27d_1^2 d_4 - 9d_1 d_2 d_3 + 2d_2^3}{27d_1^2}, \] (45)

the equation $f'(t) = 0$ can be transformed into $\tilde{t}^3 + pt + q = 0$ by letting $\tilde{t} = t - \frac{d_2}{3d_1}$. According to the Cardano’s formula [46], the discriminant of the root is $\Delta = (\frac{2}{3})^2 + (\frac{q}{p})^3$, which can be used to verify the existence and uniqueness of the solution. Similar to proving the non-negativity of (40), we can validate the non-negativity of $p$, which means $\Delta > 0$, and the equation $\tilde{t}^3 + pt + q = 0$ has only one real root, so does the equation $f'(t) = 0$. The real root of $f'(t) = 0$ can be expressed as [47]
\[ t = \frac{-d_2 - (k_1 + k_2^\frac{1}{2})}{3d_1}, \] (46)

where
\[ k_1 = e_1 d_2 + 3d_1 \left[ \frac{-e_2 + (e_3^2 - 4e_1 e_3)^{1/2}}{2} \right], \]
\[ k_2 = e_1 d_2 + 3d_1 \left[ \frac{-e_2 - (e_3^2 - 4e_1 e_3)^{1/2}}{2} \right], \] (47)

with $e_1 = d_2 - 3d_1 d_3, e_2 = d_2 d_3 - 9d_1 d_4$, and $e_3 = d_2^2 - 3d_1 d_4$.

Finally, the solution of the equation $f'(r) = 0$ can be obtained from (46) as follows:
\[ r = \frac{-d_2 - (k_1^{1/2} + k_2^{1/2})}{3d_1} + r_0. \] (48)

**ACKNOWLEDGMENT**

The MATLAB source code for this work can be downloaded from https://github.com/nbgwang/MCC-VC-for-ellipse-fitting.
[26] D. K. Prasad, M. K. Leung, and S.-Y. Cho, “Edge curvature and convexity based ellipse detection method,” Pattern Recognit., vol. 45, no. 9, pp. 3204–3221, Sep. 2012.

[27] C. Lu, S. Xia, M. Shao, and Y. Fu, “Arc-support line segments revisited: An efficient high-quality ellipse detection,” IEEE Trans. Image Process., vol. 29, pp. 768–781, 2020.

[28] P. K. Das, S. Meher, R. Panda, and A. Abraham, “An efficient blood-cell segmentation for the detection of hematological disorders,” IEEE Trans. Cybern., vol. 52, no. 10, pp. 1–12, Oct. 2022.

[29] C. Meng, Z. Li, X. Bai, and F. Zhou, “Arc adjacency matrix-based fast ellipse detection,” IEEE Trans. Image Process., vol. 29, pp. 4406–4420, 2020.

[30] J. S. Marron and M. P. Wand, “Exact mean integrated squared error,” Ann. Statist., vol. 20, no. 2, pp. 712–736, 1992.

[31] A. Domahidi, E. Chu, and S. Boyd, “ECOS: An embedded conic solver,” Cambridge Univ. Press, 2004.

[32] N. C. Schwertman, A. J. Gilks, and J. Cameron, “A simple noncalculus proof that the median minimizes the sum of the absolute deviations,” Amer. Statistician, vol. 44, no. 1, pp. 38–39, Feb. 1990.

[33] A. Ben-Tal and A. Nemirovski, Lectures on Modern Convex Optimization: Analysis, Algorithms, and Engineering Applications (MPS-SIAM Series on Optimization). Philadelphia, PA: SIAM, 2001.

[34] M. S. Lobo, L. Vandenberghe, S. Boyd, and H. Lebret, “Applications of second-order cone programming,” Linear Algebra Appl., vol. 284, nos. 1–3, pp. 193–228, Nov. 1998.

[35] K. Thornber-New, H. M. Klanderman, and W. J. Rucklidge, “Ellipse fitting by spatial averaging of random ensembles,” Pattern Recognit., vol. 106, pp. 1–15, Oct. 2020.

[36] Z. L. Szpak, W. Chojnacki, and A. van den Hengel, “Guaranteed ellipse fitting with a confidence region and an uncertainty measure for centre, axes, and orientation,” J. Math. Imag. Vis., vol. 52, no. 2, pp. 173–199, Jun. 2015.

[37] Z. Dang, K. M. Yi, Y. Hu, F. Wang, P. Fua, and M. Salzmann, “An efficient high-quality ellipse detection,” Pattern Recognit., vol. 106, pp. 3071–3076.

[38] A. Ben-Tal and A. Nemirovski, Lectures on Modern Convex Optimization: Analysis, Algorithms, and Engineering Applications (MPS-SIAM Series on Optimization). Philadelphia, PA: SIAM, 2001.

[39] J. F. Sturm, “Using SeDuMi 1.02, a MATLAB toolbox for optimization over symmetric cones,” Optim. Methods Softw., vol. 11, nos. 1–4, pp. 625–653, 1999.

[40] D. P. Hutenlocher, G. A. Klanderman, and W. J. Rucklidge, “Comparing images using the Hausdorff distance,” IEEE Trans. Pattern Anal. Mach. Intell., vol. 15, no. 9, pp. 850–863, Sep. 1993.

[41] H. I. Kakir and C. Topal, “An Euclidean ellipse comparison metric for quantitative evaluation,” in Proc. IEEE Int. Conf. Acoust. Speech Signal Process. (ICASSP), Apr. 2018, pp. 1263–1267.

[42] Voyager-2. Accessed: Nov. 1, 2020. [Online]. Available: http://mobile mag.com/2013/06/28/voyager-1-spacezone/

[43] G. Griffin, A. Holub, and P. Perona. (Mar. 2007). Caltech-256 Object Category Dataset. Caltech. [Online]. Available: https://authors.library.caltech.edu/7094/1/CNS-TR-2007-001.pdf

[44] T. Maros and R. M. Monma, “Multiple ellipse fitting by center-based clustering,” Croatian Oper. Res. Rev., vol. 6, no. 1, pp. 43–53, Mar. 2015.

[45] Z. Dang, K. M. Yi, Y. Hu, F. Wang, P. Fua, and M. Salzmann, “Eigendecomposition-free training of deep networks for linear least-square problems,” IEEE Trans. Pattern Anal. Mach. Intell., vol. 43, no. 9, pp. 3167–3182, Sep. 2021.

[46] R. Wittula and D. Sõtma, “Cardanos formula, square roots, Chebyshev polynomials and radicals,” J. Math. Anal. Appl., vol. 363, no. 2, 2010, Art. no. 639647.

[47] S. Fan, “A new extracting formula and a new distinguishing means on the one variable cubic equation,” Natural Sci. J. Hainan Teachers College, vol. 2, no. 2, pp. 91–98, Dec. 1989.

Wei Wang received the B.Eng. degree in electronic and information engineering from West Anhui University, Lu'an, China, in 2020. He is currently pursuing the master's degree in electronics and communications engineering with Ningbo University, Ningbo, China. His research interests include robust ellipse fitting and robust localization under non-line-of-sight (NLOS) conditions.

Gang Wang (Senior Member, IEEE) received the B.Eng. degree in electronic engineering from Shandong University, Jinan, China, in 2006, and the Ph.D. degree from Xidian University, Xi'an, China, in 2011.

He joined Ningbo University, Ningbo, China, in January 2012, where he is currently a Full Professor. From June 2018 to June 2019, he was a Visiting Scholar with the University of Missouri, Columbia, MO, USA. His research interests include target localization and tracking in wireless networks.

He is elected as a member of the Sensor Array and Multichannel (SAM) Technical Committee of the IEEE Signal Processing Society. He was a recipient of the Natural Science Funds for Outstanding Young Scholars from the National Natural Science Foundation of China and the Natural Science Funds for Distinguished Young Scholars from Zhejiang Provincial Natural Science Foundation. He was the Handling Editor of Digital Signal Processing (Elsevier) from September 2019 to October 2022. He serves as an Associate Editor for IEEE TRANSACTIONS ON AEROSPACE AND ELECTRONIC SYSTEMS and the Handling Editor for Signal Processing (Elsevier).

Chenlong Hu received the B.Eng. degree in electronic and information engineering from the Ningbo University of Technology, Ningbo, China, in 2018, and the master's degree in electronics and communications engineering from Ningbo University, Ningbo, in 2021. His research interests include robust ellipse fitting.

Wei Wang received the B.Eng. degree in electronic and information engineering from West Anhui University, Lu'an, China, in 2020. He is currently pursuing the master's degree in electronics and communications engineering with Ningbo University, Ningbo, China. His research interests include robust ellipse fitting and robust localization under non-line-of-sight (NLOS) conditions.

K. C. Ho (Fellow, IEEE) was born in Hong Kong. He received the B.Sc. degree (Hons.) in electronics and the Ph.D. degree in electronic engineering from The Chinese University of Hong Kong, Hong Kong, in 1988 and 1991, respectively.

He was a Research Associate with the Royal Military College of Canada from 1991 to 1994. He joined Bell-Northern Research, Montreal, QC, Canada, as a member of Scientific Staff, in 1995. He was a Faculty Member with the Department of Electrical Engineering, University of Saskatchewan, Saskatoon, SK, Canada, from September 1996 to August 1997. Since September 1997, he has been with the University of Missouri, Columbia, MO, USA, where he is currently a Professor with the Electrical Engineering and Computer Science Department. He was active in the development of the ITU-T Standard Recommendations from 1995 to 2012. He was the Rapporteur of ITU-T Q15/S16: Voice Gateway Signal Processing Functions and Circuit Multiplication Equipment/Systems (during 2009–2012) and the Associate Rapporteur of ITU-T Q16/S16: Speech Enhancement Functions in Signal Processing Network Equipment in 2012. His research interests include sensor array processing, source localization, subsurface object detection, wireless communications, and adaptive processing.

Dr. Ho was a recipient of the Senior Faculty Research Award, the Junior Faculty Research Award, and the Teaching Award from the College of Engineering, University of Missouri. He was the Technical Chair of the IEEE International Conference on Acoustics, Speech and Signal Processing in 2016. He was the Past Chair (2015), the Chair (during 2013–2014), and the Vice-Chair (during 2011–2012) of the Sensor Array and Mutilchannel Technical Committee of the IEEE Signal Processing Society. He was in the organizing committees of several IEEE sponsored conferences and workshops, including the VCIAP 2020, APSIPA 2015, CAMSAP 2011, and SAM 2008. He was an Associate Editor of the IEEE TRANSACTIONS ON SIGNAL PROCESSING (during 2009–2013 and during 2003–2006) and the IEEE SIGNAL PROCESSING LETTERS (during 2004–2008). He was the Editor of the ITU-T Recommendations G.160: Voice Enhancement Devices (during 2006–2012), the G.168: Digital Network Echo Cancellers (during 2000–2012), and the G.799.2: Mechanism for Dynamic Coordination of Signal Processing Network Equipment (during 2004–2009).