

Point Set Registration With Similarity and Affine Transformations Based on Bidirectional KMPE Loss

Yang Yang^{ID}, Member, IEEE, Dandan Fan, Shaoyi Du^{ID}, Muqi Wang, Badong Chen^{ID}, Senior Member, IEEE, and Yue Gao, Senior Member, IEEE

Abstract—Robust point set registration is a challenging problem, especially in the cases of noise, outliers, and partial overlapping. Previous methods generally formulate their objective functions based on the mean-square error (MSE) loss and, hence, are only able to register point sets under predefined constraints (e.g., with Gaussian noise). This article proposes a novel objective function based on a bidirectional kernel mean p -power error (KMPE) loss, to jointly deal with the above nonideal situations. KMPE is a nonsecond-order similarity measure in kernel space and shows a strong robustness against various noise and outliers. Moreover, a bidirectional measure is applied to judge the registration, which can avoid the ill-posed problem when a lot of points converges to the same point. In particular, we develop two effective optimization methods to deal with the point set registrations with the similarity and the affine transformations, respectively. The experimental results demonstrate the effectiveness of our methods.

Index Terms—Affine, bidirectional kernel mean p -power error (KMPE) loss, outliers, point set registration, similarity.

I. INTRODUCTION

SHAPE point set is a common feature to represent the object of interest. In many applications, in computer vision, such as object reconstruction [1], image retrieval [2], and medical image analysis [3], point set registration is a fundamental problem for feature matching. In general, the point set registration techniques need to define a certain loss function, such that a spatial transformation which maps one point set to the other can be estimated. However, the correspondence between the two point sets is unknown. The problem is usually solved under an iterative framework, such as the iterative

Manuscript received June 18, 2018; revised March 9, 2019; accepted September 14, 2019. Date of publication October 18, 2019; date of current version February 17, 2021. This work was supported in part by the National Key Research and Development Program of China under Grant 2018AAA0102500, and in part by the National Natural Science Foundation of China under Grant 61971343, Grant 91648208, Grant 61573274, and Grant 61873199. This article was recommended by Associate Editor G. C. Anagnostopoulos. (Corresponding authors: Shaoyi Du; Yue Gao.)

Y. Yang and M. Wang are with the School of Electronic and Information Engineering, Xi'an Jiaotong University, Xi'an 710049, China (e-mail: yyang@mail.xjtu.edu.cn; wangmuyi@mail.xjtu.edu.cn).

D. Fan is with the School of Software Engineering, Xi'an Jiaotong University, Xi'an 710049, China (e-mail: kris88@stu.xjtu.edu.cn).

S. Du and B. Chen are with the Institute of Artificial Intelligence and Robotics, Xi'an Jiaotong University, Xi'an 710049, China (e-mail: dushaoyi@gmail.com; chenbd@mail.xjtu.edu.cn).

Y. Gao is with the School of Software, Tsinghua University, Beijing 100084, China (e-mail: gaoyue@tsinghua.edu.cn).

Color versions of one or more of the figures in this article are available online at <http://ieeexplore.ieee.org>.

Digital Object Identifier 10.1109/TCYB.2019.2944171

closest point (ICP) algorithm [4], [5], which builds up the correspondence and calculates the transformation parameters in turn.

To achieve registration, ICP equally uses all corresponding points to estimate the transformation. Such an approach is efficient in general, but not accurate in the presence of outliers or for two point sets partially overlapping. Furthermore, some methods attempt to select part of reliable points for registration by trimming some points with greater corresponding distances [6]. To adjust the trimming ratio automatically, an overlapping ratio parameter was applied in [7]–[9]. Furthermore, to extend ICP to nonrigid registration, Du *et al.* [10]–[12] incorporated some scale and affine transformations into the ICP framework and provided an iterative solution for parameter updating. However, the above-mentioned methods all used the typical mean-square error (MSE) cost function, which is sensitive to non-Gaussian noise and outliers. To improve the robustness against noise and outliers, Liu [13] added the expectation-maximization (EM) principles to the ICP and used SoftAssign for the registration of overlapping point sets. Nevertheless, this kind of approach is time consuming due to more uncertainties in the soft-assignment. Recently, more structure features were preserved as context constraints in the registration model [14]. To seek reliable correspondence, Ma *et al.* [15] removed mismatches between two initial feature sets by locality preserving matching. However, their assumption is based on correspondences that are known in advance.

Besides, some probabilistic models [16], [17] were integrated into registration approaches to improve the robustness. With the help of the Gaussian mixture model (GMM), these methods modeled the point set as a probability density function (PDF) [18]. Then, the point set registration was solved by aligning two mixture models. Probabilistic point matching was solved by the EM updating [19]. Weighted matching was built by the normalized Gaussian model [20]. Chui *et al.* [21], [22] proposed a robust point matching algorithm for nonrigid registration, where the deformation relation is formulated by the thin-plate splines. Ma *et al.* [23], [24] and Bai *et al.* [25] proposed their robust registration methods based on GMM with manifold regularization and local connectivity. But the outliers were assumed to satisfy a uniform distribution. In multiple point sets [26], points were realized as a unique GMM and the registration was cast into a clustering problem. The above methods assign a soft correspondence between two point sets to reduce mismatching. Nevertheless, due to

stronger uncertainties, they need to estimate full correspondences for all points. In addition, they are not stable for various outliers.

Recently, employing a more robust loss function has attracted attention for point set registration. Since the loss function using a second-order statistical measure can only be robust to the Gaussian noise, Ma *et al.* [27] provided a robust nonrigid registration method based on L2E estimation. In information-theoretic learning (ITL) [28], [29], several researchers have demonstrated that nonsecond-order loss functions have a significant effect in robust learning and signal processing [30]–[33]. A typical example is correntropic loss (C-Loss) [34], [35] which is an MSE in kernel space. The shape matching method [36] minimized the C-Loss of spatial distances between two point sets. With C-Loss, the adverse effects of outliers can be automatically eliminated since they are generally assigned with small and nonlinear weights. This article showed significant performance improvements especially for the case with impulsive noise. But the algorithm assumes that a good initial correspondence should be assigned in advance. More recently, Du *et al.* [37] and Wu *et al.* [38] applied the C-Loss into the ICP framework to set up the correspondence and find the transformation concurrently. But they are only suitable for some simple noise distributions.

To jointly tackle the issues posed by severe changes in noise, outliers, and partial overlapping, a robust point set registration method is proposed in this article. The main idea is to introduce a state-of-the-art nonsecond-order similarity measure in kernel space, known as the kernel mean p -power error (KMPE) [39] loss, to formulate the objective function. Compared with the C-Loss, the KMPE is more flexible and robust to general noisy data with outliers. Actually, the C-Loss is contained in the KMPE loss as a special case when the power $p = 2$. Different from our previous work in [40], we extend the registration problem from a rigid transformation to two nonrigid forms: similarity and affine transformations, which brings more challenges. To prevent the scale factor converging to zero in both similarity and affine registrations, we use a bidirectional measure [41]–[43] to maximum overlapping parts. This also avoids the ill-posed problem when a lot of points converge to the same point. Furthermore, we develop two effective optimization algorithms under the ICP framework to deal with the point set registration with the similarity and the affine transformations, respectively. Since the proposed algorithm is based on the bidirectional KMPE loss and adopts the one-to-one hard correspondence, it achieves low computational complexity, high accuracy, and robustness.

The main contributions of this article are as follows.

- 1) A bidirectional KMPE loss measure is proposed for the point set registration problem. It can jointly deal with the issues posed by the non-Gaussian noises, outliers, and partial overlapping.
- 2) After modeling both the similarity and affine registrations with the bidirectional KMPE loss measure, we solved these two problems under the ICP framework by optimizing the objective function with two groups of transformation parameters, respectively.

- 3) Some proofs are given to highlight the properties of the proposed objective function and BiK-ICP algorithm.
- 4) The experimental evaluations are conducted in various nonideal noisy conditions and reveal the superiority of our methods compared with some state-of-the-art methods.

The remainder of this article is organized as follows. Section II briefly reviews two preliminaries: 1) ICP and 2) KMPE. In Section III, the problem formulation based on the bidirection KMPE loss is proposed. Section IV presents the proposed BiK-ICP algorithm in detail, and Section V discusses some important properties of our method. Section VI reports the experimental settings and results. Finally, we give the conclusion in Section VII.

II. PRELIMINARIES

To make this article self-contained, we briefly review the classic ICP algorithm [4] and the nonsecond-order loss function KMPE [39] in this section.

A. ICP Algorithm

Suppose we have two point sets, one is the model point set $\mathbf{X} \stackrel{\Delta}{=} \{\vec{x}_i\}_{i=1}^{N_x}$ and the other is the data point set $\mathbf{Y} \stackrel{\Delta}{=} \{\vec{y}_j\}_{j=1}^{N_y}$ in \mathbb{R}^2 . The goal is to find a spatial transformation for the model point set such that the distance between two point sets can reach minimum. Traditionally, the problem is formulated with an MSE measure

$$\min_{j \in \{1, 2, \dots, N_y\}} \frac{1}{N_x} \sum_{i=1}^{N_x} \|\mathbf{T}(\vec{x}_i) - \vec{y}_j\|_2^2 \quad (1)$$

where \mathbf{T} is a nonsingular matrix.

To solve (1), it is necessary to impose a regularity condition on transformation parameters. The ICP algorithm assumes \mathbf{T} to be a rigid transformation, that is, a combination of a rotation matrix \mathbf{R} and a translation vector \vec{t} . The optimization of (1) consists of an iterative process. In each iteration, according to the current spatial location, the correspondence between the data and the model point set is built up and then transformation parameters are optimized with the resulting correspondence. Especially, in the k th iteration, the above two steps can be presented as follows.

Step 1: Map all points in \mathbf{X} based on the $(k-1)$ th rigid transformation $(\mathbf{R}_{k-1}, \vec{t}_{k-1})$, and find the corresponding point in \mathbf{Y} for each \vec{x}_i

$$c_k(i) = \arg \min_{j \in \{1, 2, \dots, N_y\}} \left(\|(\mathbf{R}_{k-1}\vec{x}_i + \vec{t}_{k-1}) - \vec{y}_j\|_2^2 \right), i = 1, \dots, N_x. \quad (2)$$

Step 2: Optimize the transformation parameters $(\mathbf{R}_k, \vec{t}_k)$ with the known correspondence $\{i, c_k(i)\}$

$$(\mathbf{R}_k, \vec{t}_k) = \arg \min_{\mathbf{R}^T \mathbf{R} = I_2, \det(\mathbf{R}) = 1, \vec{t}} \left(\sum_{i=1}^{N_x} \|\mathbf{R}\vec{x}_i + \vec{t} - \vec{y}_{c_k(i)}\|_2^2 \right). \quad (3)$$

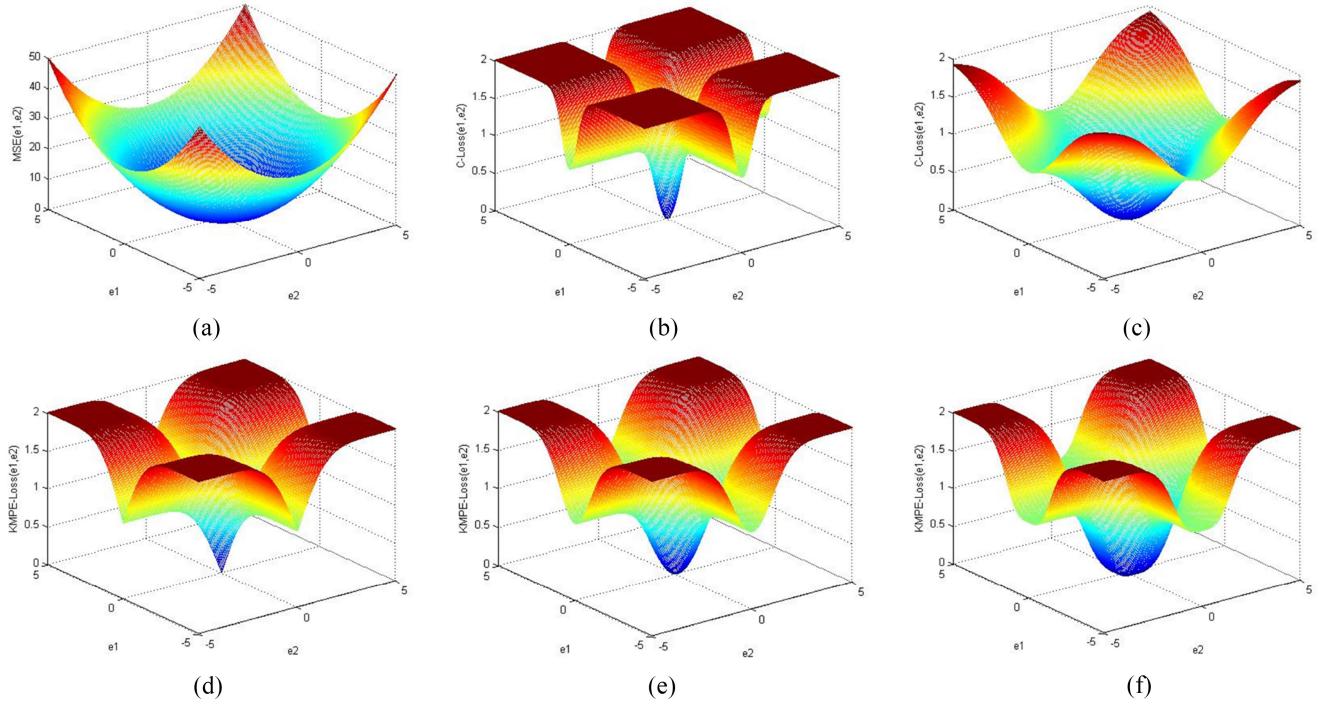


Fig. 1. Bidirectional loss function with different parameters: e_1 and e_2 denote the errors from two directions, respectively. (a) Bidirectional MSE: $e_1^2 + e_2^2$; (b) and (c) Bidirectional C-loss with Gauss kernel: $(1 - \kappa_\sigma(e_1)) + (1 - \kappa_\sigma(e_2))$, $\sigma = 0.5$ and 2 ; and (d)–(f) Bidirectional KMPE loss: $BiK(e_1, e_2)$, $\sigma = 1$ and $p = 1, 2, 3$.

B. KMPE

Nonsecond-order loss functions have made great success in the ITL. As a representation of these kinds of loss functions, C-Loss can be considered as the MSE in kernel space and has been successfully applied in robust signal processing. Recently, KMPE, a new nonsecond-order statistical measure in kernel space, has embodied its superiority by generalizing C-Loss to the case with arbitrary power. Given two random variables X and Y , the general KMPE is defined by

$$C_p(X, Y) = \mathbb{E}[(1 - \kappa_\sigma(X - Y))^{p/2}] \quad (4)$$

where $p > 0$ is the power parameter. Without being mentioned otherwise, the kernel function is specified as a Gaussian kernel and can be described by

$$\kappa_\sigma(X - Y) = \exp(-\|X - Y\|_2^2/2\sigma^2)$$

with σ being the kernel bandwidth.

In practice, when a set of pairwise samples $(x_i, y_i)_{i=1}^N$ is given, the empirical KMPE can be estimated by

$$\hat{C}_p(X, Y) = \frac{1}{N} \sum_{i=1}^N (1 - \kappa_\sigma(x_i - y_i))^{p/2}. \quad (5)$$

It is easy to find that $C_p(X, Y)$ is symmetric, positive, and bounded. Especially, when $p = 2$, it converts to the C-Loss measure. Nevertheless, it may show better performance when p takes other values [39].

III. BIDIRECTIONAL KMPE

In this article, we assume that two shape point sets \mathbf{X} and \mathbf{Y} are composed of independent and identically distributed (i.i.d.)

samples drawn from the same random variable in \mathbb{R}^2 with noise and outliers. The goal is to register them with a transformation matrix \mathbf{W} and a translation vector \vec{t} , which can be modeled by a minimum KMPE loss function.

If we just consider the simple unidirectional transformation, a \mathbf{W} with $\det(\mathbf{W}) \rightarrow 0$ will lead to an ill-posed function, making all points in \mathbf{X} be aligned to one specific point in \mathbf{Y} . To address this issue, we define a bidirectional KMPE loss function as follows:

$$BiK(e_1, e_2) = (1 - \kappa_\sigma(e_1))^{p/2} + (1 - \kappa_\sigma(e_2))^{p/2} \\ e_1 = \|\vec{x}_i - \vec{y}_{c(i)}\|, \quad e_2 = \|\vec{y}_j - \vec{x}_{d(j)}\| \quad (6)$$

where $c(.)$ and $d(.)$ denote the indices of corresponding points matching two given points.

For comparison, Fig. 1 shows the plots of three kinds of bidirectional loss functions with different parameters. Fig. 1(a) exhibits the surface of the bidirectional MSE loss: $e_1^2 + e_2^2$. It is very smooth but without boundary and then a big noise will bring a strong influence to the loss function. Fig. 1(b) and (c) shows the surfaces of bidirectional C-Loss functions with different kernel bandwidths: $(1 - \kappa_\sigma(e_1)) + (1 - \kappa_\sigma(e_2))$, $\sigma = 0.5$ and 2 . We can see that these functions are bounded and the influence from big errors will be automatically weakened. The parameter σ mainly controls the kernel bandwidth. Fig. 1(d) and (f) shows the surfaces of the bidirectional KMPE loss functions with different power parameters: $BiK(e_1, e_2)$, $\sigma = 1$ and $p = 1, 2, 3$. We can see that the bidirectional KMPE loss gives more emphasis along the area around $e_1 = 0$ and $e_2 = 0$, but filters out the behaviors when e_1 or e_2 is big. This means that if two points relate more to each other, they will have more contributions to the objective function. Thus,

other noisy points and outliers are penalized naturally. It can be observed that bidirectional KMPE loss further improves C-Loss by adjusting its “Sharpness” near the origin with a variable parameter p , and the smaller p is, the sharper the surface is. This makes it more flexible to deal with different kinds of noises. We can also see that $BiK(e_1, e_2) = 0$ if and only if both e_1 and e_2 are equal to 0. Since one nonzero error will continue to drive the optimization process even when the other error is equal to zero, the bidirectional KMPE loss can avoid the ill-posed situation mentioned above.

Based on the aforementioned analysis, we can minimize the objective function for point set registration as follows:

$$J_{BiK}(\mathbf{X}, \mathbf{Y}) = \frac{1}{N_x + N_y} \left(\sum_{i=1}^{N_x} (1 - \kappa_\sigma(\mathbf{W}\vec{x}_i + \vec{t} - \vec{y}_{c(i)}))^{p/2} + \sum_{j=1}^{N_y} (1 - \kappa_\sigma(\mathbf{W}\vec{x}_{d(j)} + \vec{t} + \vec{y}_j))^{p/2} \right) \quad (7)$$

where \mathbf{W} and \vec{t} are the transformation matrix and the translation vector, respectively. The first main part is the sum of the forward distance from \mathbf{X} to \mathbf{Y} . And the second part is the backward distance from \mathbf{Y} to \mathbf{X} .

IV. BiK-ICP ALGORITHM

There are two kinds of unknown factors in the objective function (7): 1) the point correspondence between two point sets and 2) the specific transformation parameters for registration. To minimize this objective function, we develop an algorithm under the ICP framework, called BiK-ICP. In the k th iteration, two steps are as follows.

Step 1: Find the bidirectional correspondences between the two point sets \mathbf{X} and \mathbf{Y} obtained in the $(k-1)$ th iteration

$$\begin{cases} c_k(i) = \arg \min_{j \in \{1, 2, \dots, N_y\}} \|(\mathbf{W}_{k-1}\vec{x}_i + \vec{t}_{k-1}) - \vec{y}_j\|_2^2 \\ d_k(j) = \arg \min_{i \in \{1, 2, \dots, N_x\}} \|(\mathbf{W}_{k-1}\vec{x}_i + \vec{t}_{k-1}) - \vec{y}_j\|_2^2. \end{cases} \quad (8)$$

Several searching strategies can be used to find each pair of matching points. The nearest-neighbor searching strategy based on k -d tree is employed here to ensure the searching speed.

Step 2: Optimize the k th transformation parameters with the known correspondences $\{i, c_k(i)\}_{i=1}^{N_x}$ and $\{j, d_k(j)\}_{j=1}^{N_y}$. Letting $N = N_x + N_y$, we redefine two point sets $\mathbf{U} = \{\vec{u}_i\}_{i=1}^N$ and $\mathbf{V} = \{\vec{v}_i\}_{i=1}^N$ as

$$\begin{aligned} \vec{u}_i &= \begin{cases} \vec{x}_i, & 1 \leq i \leq N_x \\ \vec{x}_{d_k(i-N_x)}, & N_x + 1 \leq i \leq N \end{cases} \\ \vec{v}_i &= \begin{cases} \vec{y}_{c_k(i)}, & 1 \leq i \leq N_x \\ \vec{y}_{i-N_x}, & N_x + 1 \leq i \leq N. \end{cases} \end{aligned}$$

Then, the transformation parameters can be obtained by minimizing the following function:

$$(\mathbf{W}_k, \vec{t}_k) = \arg \min_{W, \vec{t}} \frac{1}{N} \sum_{i=1}^N (1 - \kappa_\sigma(\mathbf{W}\vec{u}_i + \vec{t} - \vec{v}_i))^{p/2}. \quad (9)$$

Now, the key lies in how to effectively estimate the transformation parameters \mathbf{W} and \vec{t} in (9).

A. Estimation of the Similarity Transformation Parameter

For two point sets with similarity transformation, the matrix \mathbf{W} can be considered as a combination of a scale factor s and a rotation matrix \mathbf{R} , that is, $\mathbf{W} = s\mathbf{R}$, where $\mathbf{R}^T\mathbf{R} = \mathbf{I}_2$ and $\det(\mathbf{R}) = 1$. Then, the goal of step 2 becomes obtaining a group of optimal parameters (s, \mathbf{R}, \vec{t}) based on current corresponding points.

To achieve this goal, we first deduce the derivative of (9) with respect to \vec{t} and obtain

$$\vec{t}_k = \sum_{i=1}^N \varphi(e_i)(\vec{v}_i - s\mathbf{R}\vec{u}_i) / \sum_{i=1}^N \varphi(e_i) \quad (10)$$

where $\varphi(e_i) \triangleq (1 - \kappa_\sigma(e_i))^{[(p-2)/2]}\kappa_\sigma'(e_i)$ and $e_i = s\mathbf{R}\vec{u}_i + \vec{t} - \vec{v}_i$.

Substituting \vec{t}_k into (9) and defining

$$\begin{cases} \vec{p}_i \triangleq \vec{u}_i - \sum_{i=1}^N \varphi(e_i)\vec{u}_i / \sum_{i=1}^N \varphi(e_i) \\ \vec{q}_i \triangleq \vec{v}_i - \sum_{i=1}^N \varphi(e_i)\vec{v}_i / \sum_{i=1}^N \varphi(e_i) \end{cases}$$

we can obtain a simplified objective function as

$$(\mathbf{R}_k, s_k) = \arg \min_{\mathbf{R}^T\mathbf{R}=\mathbf{I}_2, \det(\mathbf{R})=1, s} \frac{1}{N} \sum_{i=1}^N (1 - \kappa_\sigma(s\mathbf{R}\vec{p}_i - \vec{q}_i))^{p/2}. \quad (11)$$

Next, we estimate s_k and \mathbf{R}_k in a serial way.

- 1) For any given \mathbf{R}_k , the optimal value of the factor s can be obtained by setting the derivative of (11) with respect to s to be zero, and we can obtain

$$s_k = \sum_{i=1}^N \vec{q}_i^T \mathbf{R}_k \varphi(e_i) \vec{p}_i / \sum_{i=1}^N \vec{p}_i^T \varphi(e_i) \vec{p}_i. \quad (12)$$

- 2) For any given s_k , it is still challenging to obtain the optimal rotation matrix. Here, we resort to the method proposed by Arun *et al.* [44]. This method estimates the rotation matrix for the MSE-based objective function by singular value decomposition (SVD). Inspired by this method, we first construct a matrix \mathbf{H} for our objective function

$$\mathbf{H} = \frac{1}{N} \sum_{i=1}^N s_k \vec{p}_i \varphi(e_i) \vec{q}_i^T$$

and then decompose \mathbf{H} by SVD as follows: $\mathbf{H} = \mathbf{S}\Lambda\mathbf{D}$. Accordingly, \mathbf{R}_k can be estimated as

$$\mathbf{R}_k = \tilde{\mathbf{S}}\tilde{\mathbf{D}}^{-1}\mathbf{T} \quad (13)$$

where

$$\tilde{\mathbf{I}} = \begin{cases} \mathbf{I}_2 & \det(\mathbf{H}) > 0 \\ \text{diag}(1, -1) & \det(\mathbf{H}) < 0. \end{cases}$$

Note that $\varphi(e_i)$ that is of concern in (12) and (13) is also related to s and \mathbf{R} . For the solution of (12) and (13), we make an assumption that the kernel term in the k th iteration is known $\varphi(e_i) = \varphi(s_{k-1}\mathbf{R}_{k-1}\vec{u}_i + \vec{t}_{k-1} - \vec{v}_i)$, that is, replacing

the unknown kernel term in the current iteration with the one obtained in the last iteration. Since the difference between the optimized parameters in two successive iterations is small, this approximation has little impact on the registration result.

Finally, with the calculated parameters s_k and \mathbf{R}_k , the translation parameter \vec{t}_k can be acquired according to (10).

B. Estimation of the Affine Transformation Parameter

For the point set registration with affine transformation, the matrix \mathbf{W} in (9) can be defined as an invertible matrix \mathbf{A} . The goal is to obtain the optimal transformation (\mathbf{A}, \vec{t}) with the current correspondence.

With the similar deducing process of (10), we can obtain the optimal \vec{t}_k for a given \mathbf{A}

$$\vec{t}_k = \sum_{i=1}^N \varphi(e_i)(\vec{v}_i - \mathbf{A}\vec{u}_i) / \sum_{i=1}^N \varphi(e_i) \quad (14)$$

where $e_i = \mathbf{A}\vec{u}_i + \vec{t} - \vec{v}_i$.

After substituting \vec{t}_k into (9), a new objective function similar to (11) can be obtained

$$\mathbf{A}_k = \arg \min_{\det(\mathbf{A}) \neq 0} \frac{1}{N} \sum_{i=1}^N (1 - \kappa_\sigma(\mathbf{A}\vec{p}_i - \vec{q}_i))^{p/2}. \quad (15)$$

Next, we generate two matrices \mathbf{P} and \mathbf{Q} by taking points \vec{p}_i and \vec{q}_i as rows, respectively. Let \mathbf{g} be a column vector with $g_i = \kappa_\sigma(\mathbf{A}\vec{p}_i - \vec{q}_i)$. We set a diagonal matrix $G(\mathbf{Q} - \mathbf{P}\mathbf{A}^T) = D(\mathbf{g}) = \text{diag}(\mathbf{g})$.

Then, (15) can be written as its matrix form

$$\mathbf{A}_k = \arg \min_{\det(\mathbf{A}) \neq 0} \frac{1}{N} (\mathbf{I}_N - G(\mathbf{Q} - \mathbf{P}\mathbf{A}^T))^{p/2} \quad (16)$$

where \mathbf{I}_N is an identity matrix of size $N \times N$.

Setting the derivative of (16) with respect to \mathbf{A} to be zero, we have

$$\mathbf{P}^T D(\mathbf{g}) \mathbf{K} \mathbf{P} \mathbf{A}_k - \mathbf{P}^T D(\mathbf{g}) \mathbf{K} \mathbf{Q} = 0 \quad (17)$$

where $\mathbf{K} = (\mathbf{I}_N - D(\mathbf{g}))^{[(p-2)/2]}$. Thus

$$\mathbf{A}_k = (\mathbf{P}^T D(\mathbf{g}) \mathbf{K})^{-1} \mathbf{P}^T D(\mathbf{g}) \mathbf{K} \mathbf{Q}. \quad (18)$$

Finally, we can estimate the affine transform parameters according to (14) and (18), respectively.

Table I summarizes the process of BiK-ICP. Two termination conditions are provided in Algorithm 1. One is to judge whether the difference between the registration errors e_k and e_{k-1} in two successive iterations is smaller than a given small value ε , and the other is to judge whether the number of iterations reaches the given maximum value.

As in the other kernel methods, the kernel width σ mostly affects the performance of BiK-ICP. A proper kernel width is able to effectively distinguish noise and outliers and, thus, restrains their influence. At the beginning of the registration, most points are not well aligned, and we wish to utilize as much point information as possible, which requires a large bandwidth. Gradually, the alignment becomes more accurate, and it is beneficial to filter out those noisy points through a smaller bandwidth. Therefore, the kernel width σ should be

TABLE I
BiK-ICP ALGORITHM

Algorithm 1: BiK-ICP

Require: Two shape point sets $\mathbf{X} \triangleq \{\vec{x}_i\}_{i=1}^{N_x}$ and $\mathbf{Y} \triangleq \{\vec{y}_i\}_{i=1}^{N_y}$; Initial parameters: s_0 , \mathbf{R}_0 (or \mathbf{A}_0) and \vec{t}_0 , σ_0 and p .

- 1: **For** $k = 1, \dots, K$.
- 2: Set up correspondences $c_k(i)$ and $d_k(j)$ via (8).
- 3: Compute $\varphi(e_i)$ by s_{k-1} , \mathbf{R}_{k-1} , and t_{k-1} .
- 4: Reconstruct matrices \mathbf{P} and \mathbf{Q} .
- 5: Solve s_k and \mathbf{R}_k via (12) and (13) or \mathbf{A}_k via (18).
- 6: Solve the translation vector \vec{t}_k via (10) or (14).
- 7: Update σ .
- 8: Compute the bidirectional MSE e_k .
- 9: **If** $|e_k - e_{k-1}| \leq \varepsilon$.
- 10: **break**;
- 11: **end if**.
- 12: **end for**.

Return: the transformation parameters s , \mathbf{R} or \mathbf{A} and \vec{t} .

adjusted from big to small according to the current registration error. There are some adaptive techniques to determine σ . Here, we employ the one presented in [30].

For the KMPE loss, another free parameter p needs to be determined for robust registration. From the loss function surfaces shown in Fig. 1, we can see that parameter p adjusts the sharpness of the KMPE loss, which may make the proposed method adapt to the various changes of noise. Through some preliminary experiments, we find that a proper p is related to the relative noise of the point data. The general rule is when the noise is big, p should be big, corresponding to a “wide” loss around the right registration result. On the contrary, when the noise is small, p should be small, corresponding to a “sharp” loss. Generally, good registration results can be achieved when p is in the range [0.2, 8]. For a concrete registration task, we can search for a proper p in this bounded range with a greedy searching strategy. The details are given in Section VI.

V. DISCUSSION

In this section, we mainly discuss some properties of the proposed objective function and the BiK-ICP algorithm.

Property 1: The proposed objective function $J_{BiK}(\mathbf{X}, \mathbf{Y})$ is positive and bounded: $0 \leq J_{BiK}(\mathbf{X}, \mathbf{Y}) < 1$.

Proof: Given two point sets \mathbf{X} and \mathbf{Y} , straightforwardly since $0 < \kappa_\sigma(\vec{x}_i - \vec{y}_{c(i)}) \leq 1$ and $0 < \kappa_\sigma(\vec{x}_{d(j)} - \vec{y}_j) \leq 1$, $J_{BiK}(\mathbf{X}, \mathbf{Y})$ is positive and bounded. ■

Property 2: For any given initial parameters, BiK-ICP for point set registration with similarity or affine transformation monotonically converges to a minimum.

Proof: For two point sets \mathbf{X} and \mathbf{Y} tackled by BiK-ICP, we assume that the transformation parameters \mathbf{W}_{k-1} and \vec{t}_{k-1} are known in the k th iteration. Then, after solving (8) in step 1, we have $J_{BiK}(\mathbf{U}_k, \mathbf{V}_k, \mathbf{W}_{k-1}, \vec{t}_{k-1}) \leq J_{BiK}(\mathbf{U}_{k-1}, \mathbf{V}_{k-1}, \mathbf{W}_{k-1}, \vec{t}_{k-1})$. As for step 2, it enables us to acquire a group of better parameters \mathbf{W}_k and \vec{t}_k , satisfying $J_{BiK}(\mathbf{U}_k, \mathbf{V}_k, \mathbf{W}_k, \vec{t}_k) \leq J_{BiK}(\mathbf{U}_k, \mathbf{V}_k, \mathbf{W}_{k-1}, \vec{t}_{k-1})$. ■

Consequently, for all $k = 1, 2, 3, \dots$, we have

$$\begin{aligned} J_{BiK}(\mathbf{U}_k, \mathbf{V}_k, \mathbf{W}_k, \vec{t}_k) &\leq J_{BiK}(\mathbf{U}_k, \mathbf{V}_k, \mathbf{W}_{k-1}, \vec{t}_{k-1}) \\ &\leq J_{BiK}(\mathbf{U}_{k-1}, \mathbf{V}_{k-1}, \mathbf{W}_{k-1}, \vec{t}_{k-1}). \end{aligned}$$

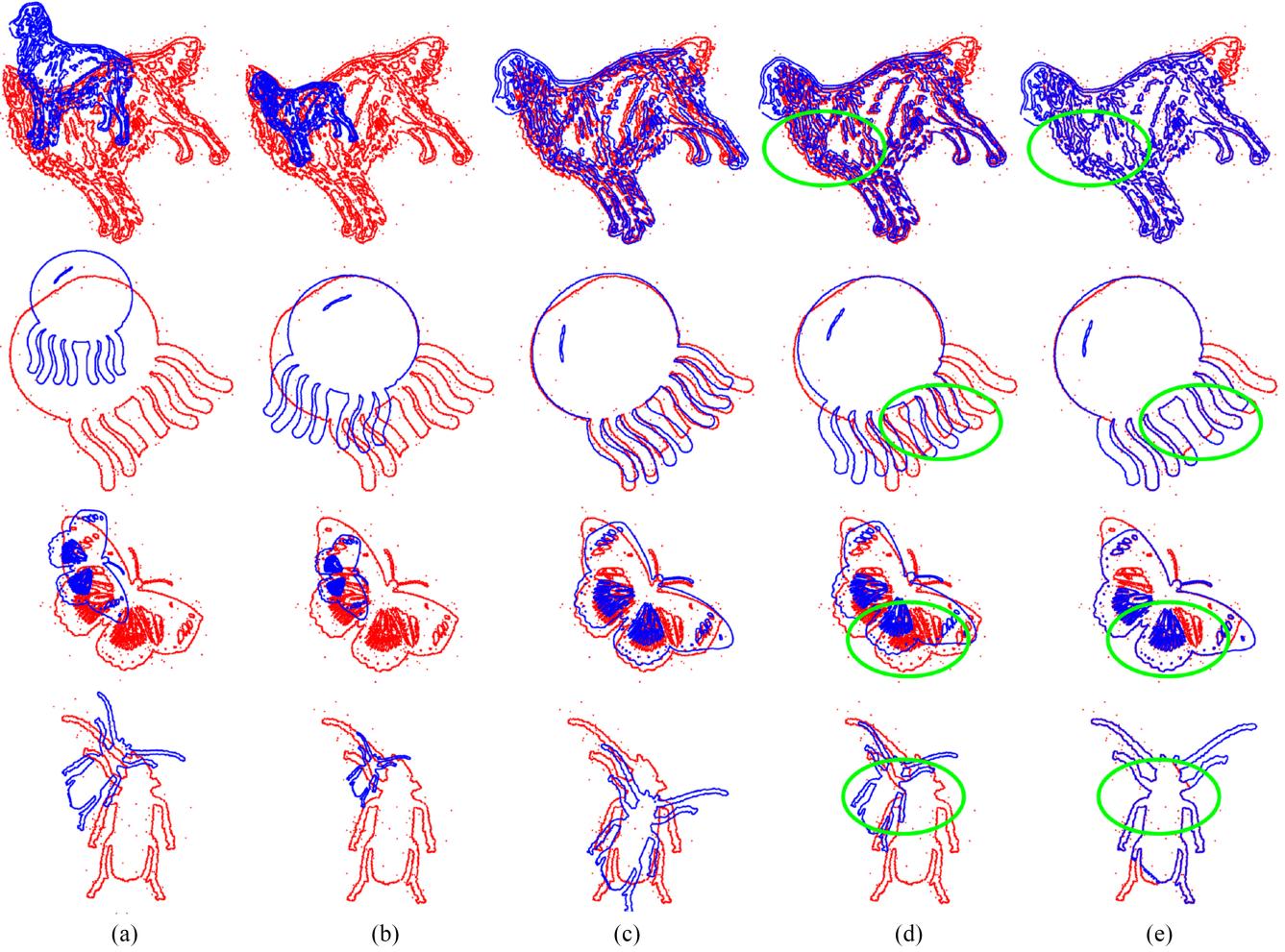


Fig. 2. Registration results on trimmed and noisy shapes with similarity transformation: rotation angle $\theta = \pi/5$, $s = 1.5$, $\vec{t} = [20, 10]^T$ under the uniform background noises and outliers. (a) Initialization, (b) result of SICP, (c) result of Bi-SICP, (d) result of BiC-SICP, and (e) result of our BiK-SICP.

This means that the objective function decreases at each alternating minimization step. Besides, Property 1 indicates that $J_{BiK}(\mathbf{X}, \mathbf{Y})$ is bounded. According to the monotonic sequence theorem that “every bounded monotonic sequence of real numbers is convergent,” it can be concluded that the BiK-ICP algorithm converges to a minimum.

Computational Complexity: For step 1 of BiK-ICP, we build the correspondence between two point sets by means of the k-d tree method, the computational complexity of which is $O(N_x \ln N_y) + O(N_y \ln N_x)$. Since the matrices of concern in (13) have the same size of 2×2 , the calculation quantity of estimating \mathbf{R} mainly lies in the calculation of \mathbf{H} , the computational complexity of which is $O(N)$. From (10) and (12), it can be known that the computational complexity to estimate s and \vec{t} also is $O(N)$. Thus, the total computational complexity to estimate the similarity transformation parameters is $O(N)$. The matrices \mathbf{P} and \mathbf{Q} in (18) both have the size of $N \times 2$ and can be calculated at a cost of $O(N)$. Profiting from the diagonal property of the matrices $D(\mathbf{g})$ and \mathbf{K} , the computational complexity of the multiplication operation in (18) is $O(N)$. Further taking the computational complexity to estimate the translation parameter \vec{t} ($O(N)$) into account, we can know

that the total computational complexity to estimate the affine transformation parameters is $O(N)$.

VI. EXPERIMENTAL RESULTS

In this section, the performance of the proposed BiK-ICP algorithm is evaluated for partial overlapping point sets with various noise and outliers. All shape point sets are obtained from the 2-D shape dataset in CE-Shape-1 [45] and normalized into 500×500 size. First, to evaluate the registration with similarity transformation, we compare the registration results by the ICP algorithm with a scale factor (SICP) [10]; the bidirectional SICP algorithm (Bi-SICP) [46]; the Bi-SICP using C-Loss (BiC-SICP) [36], [37]; and our method using bidirectional KMPE loss (BiK-SICP). And then, to evaluate the registration with affine transformation, we compare the results by the ICP algorithm (ICP), the bidirectional ICP algorithm with affine transformation (Bi-AICP) [41], Bi-AICP using C-Loss (BiC-AICP) [38], and our method using bidirectional KMPE loss (BiK-AICP). Finally, we compare the robustness testing on different kinds of noises.

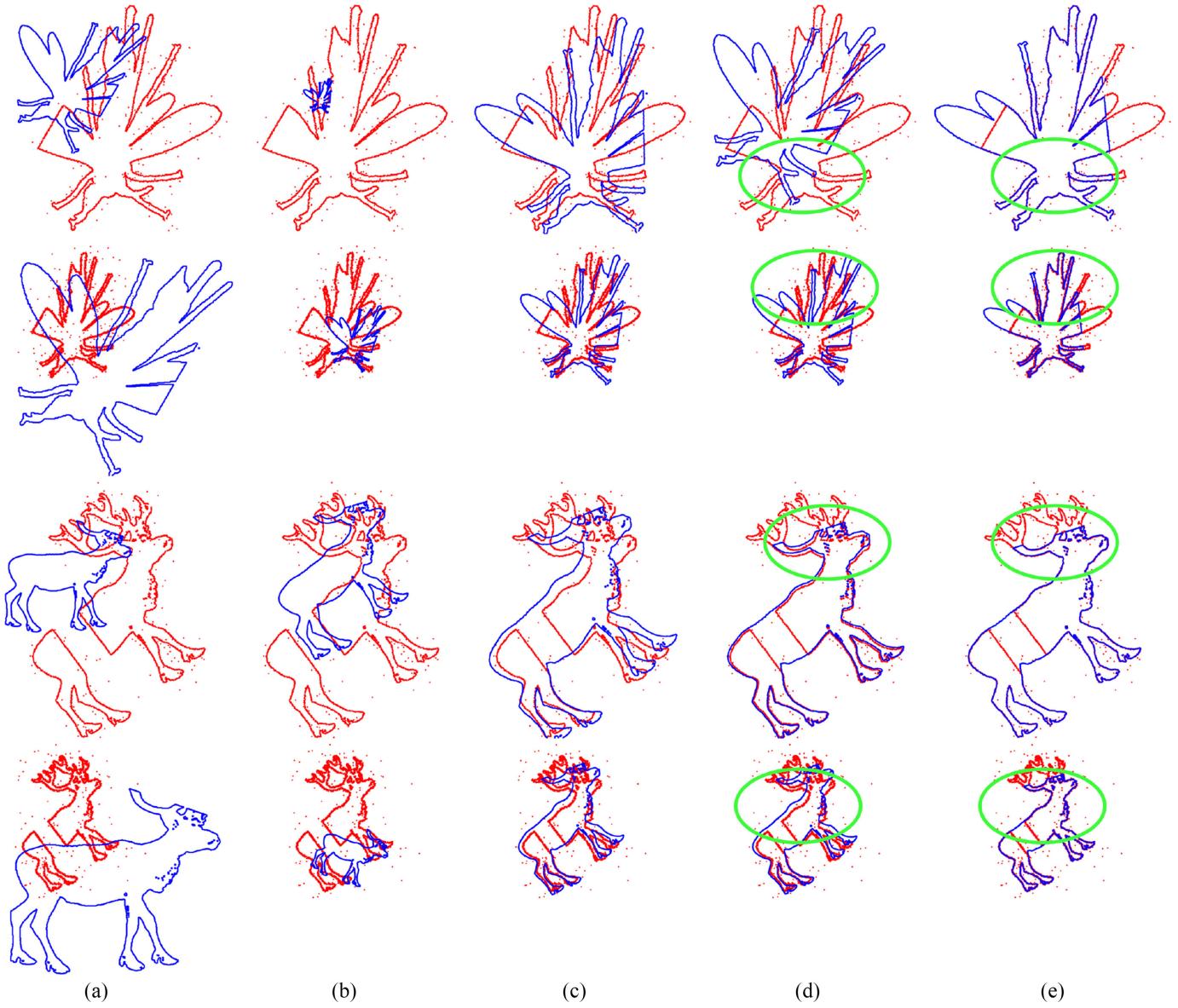


Fig. 3. Registration results on trimmed and noisy shapes of “fly” and “deer” with similarity transformation: $s = 1.5$, rotation angle $\theta = \pi/5, \vec{\tau} = [20, 10]^T$ (the first and third rows) and $s = 0.5$, rotation angle $\theta = \pi/5, \vec{\tau} = [20, 10]^T$ (the second and fourth rows). (a) Initialization, (b) result of SICP, (c) result of Bi-SICP, (d) result of BiC-SICP, and (e) result of our BiK-SICP.

A. Evaluation on Similarity Registration

To evaluate the performance on similarity registration, testing data are simulated in particular ways with various cases of noise and outliers. First, the testing point sets are transformed by a particular similarity transformation. And then, we trim parts of points randomly and add the non-Gaussian noise in the form as $v(i) = (1 - a(i))A(i) + a(i)B(i)$, where $A(i)$ denotes the background noise and $B(i)$ denotes the large outliers. The background noises $A(i)$ are considered as a uniform distribution over $[-\sqrt{3}, \sqrt{3}]$. $B(i)$ is a Gaussian noise with a big variance, that is, $N(0, 30)$. $a(i)$ is a binary i.i.d. process, whose probability is set as $\Pr\{a(i) = 1\} = 0.08$ and $\Pr\{a(i) = 0\} = 0.92$.

The power p in the BiK-SICP is a free parameter. Generally, good registration results can be achieved when p is in the range $[0.2, 8]$. For unknown point sets, it is selected by a greedy

searching strategy by setting $p = [0.2, 0.4, 0.6, \dots, 8]$ to achieve desirable performance. And the identity transformation matrix is used as the initiation parameter for similarity registration.

Fig. 2 demonstrates the registration results by different algorithms in an intuitive way. Four kinds of shapes are selected. We can see that since the unidirectional direction measure cannot constrain the scale factor, the SICP algorithm obtains the totally wrong correspondence and mismatch. The results of the Bi-SICP algorithm are badly influenced by partial overlapping and outliers. Furthermore, from the local marked regions in green circles, we can observe that some complex noisy situations affect the accuracy of the BiC-SICP algorithm. Nevertheless, by our BiK-SICP algorithm, the corresponding points are completely aligned.

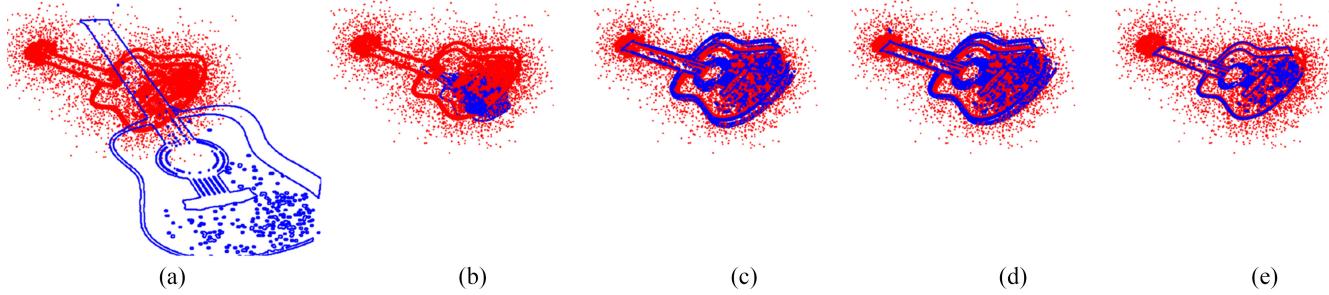


Fig. 4. Accumulated 10 times repeating the registration results on trimmed and a noisy “guitar” point set with similarity transformation: $s = 0.5$, rotation angle $\theta = \pi/5$, $\vec{t} = [20, 10]^T$. (a) Initialization, (b) result of SICP, (c) result of Bi-SICP, (d) result of BiC-SICP, and (e) result of our BiK-SICP.

TABLE II
BIDIRECTIONAL MSE'S AND STANDARD DEVIATION OF SIMILARITY REGISTRATION OBTAINED BY 20 TIMES REPEATING

Transform	Data	Bi-SICP	BiC-SICP	BiK-SICP
$s = 0.5$ $\theta = \pi/5$ $\vec{t} = [20, 10]^T$	'dog'	2.24 ± 0.10	2.24 ± 0.09	1.66 ± 0.13
	'octopus'	3.62 ± 0.34	9.70 ± 4.42	3.34 ± 0.81
	'butterfly'	2.69 ± 0.06	2.65 ± 0.06	2.59 ± 0.04
	'beetle'	6.16 ± 0.80	5.60 ± 0.68	5.27 ± 0.89
	'fly'	7.28 ± 0.50	6.92 ± 0.58	6.45 ± 1.06
	'deer'	6.48 ± 0.10	6.27 ± 0.09	5.92 ± 0.06
$s = 1.5$ $\theta = \pi/5$ $\vec{t} = [20, 10]^T$	'guitar'	3.06 ± 0.16	2.90 ± 0.12	2.43 ± 0.04
	'dog'	5.66 ± 0.11	4.70 ± 0.20	3.70 ± 0.08
	'octopus'	11.85 ± 1.22	18.25 ± 4.44	9.45 ± 2.36
	'butterfly'	8.22 ± 0.37	9.83 ± 0.71	7.12 ± 0.95
	'beetle'	21.53 ± 1.61	32.44 ± 8.74	17.53 ± 1.36
	'fly'	22.81 ± 0.68	26.61 ± 8.13	15.37 ± 2.15
	'deer'	18.85 ± 0.25	16.95 ± 0.14	16.12 ± 0.12
	'guitar'	7.94 ± 0.19	7.54 ± 1.01	6.23 ± 0.13

We analyze the reasons of the above results. In the previous ICP method, each corresponding point pair has the same contribution to the objective function. When the point sets have non-Gaussian noise, they will drag the entire point set to be far away from the correct position for averaging the registration errors. To handle the non-Gaussian noise, some nonsecond-order-loss functions are effective. But the BiC-SICP method is only suitable for a simple noise situation. Its results become bad especially for the heavy-tailed noise and mixed noise distributions. The core reason of the better performance of our method is that corresponding point pairs are set by different nonlinear weights. The result is that the adverse effects from noisy points can be eliminated. Moreover, there is one more free parameter p to adjust the loss function, as shown in Fig. 1. So it is more flexible to deal with different kinds of noise.

To compare the registration accuracy quantitatively, bidirectional MSE is used to compute the distances between corresponding points. Although there is a lacking part of correspondence on trimmed parts, the bidirectional MSE is still suitable to compare the relative accuracies among different algorithms. In our experiments, the noise is randomly added to the testing data. Thus, every experimental condition is repeated 20 times. And we record the bidirectional MSE's mean and standard deviation. The statistical quantitative results for seven point sets are demonstrated in Table II. We can see that our BiK-SICP algorithm achieves the best performance among these algorithms with the smallest bidirectional MSE.

TABLE III
BIDIRECTIONAL MSE'S MEAN AND STANDARD DEVIATION OF AFFINE REGISTRATION OBTAINED BY 20 TIMES REPEATING

Data	Bi-AICP	BiC-AICP	BiK-AICP
'apple'	10.36 ± 0.26	8.67 ± 1.32	7.90 ± 0.15
'chicken'	18.35 ± 0.76	21.86 ± 3.15	16.83 ± 0.67
'turtle'	3.96 ± 0.26	9.28 ± 25.99	3.06 ± 0.20
'elephant'	10.24 ± 0.28	10.01 ± 2.96	7.00 ± 1.43
'horse'	16.39 ± 0.28	14.30 ± 0.20	14.16 ± 0.10

Moreover, Fig. 3 shows the similarity registration results by different scale factors ($s = 0.5$ and $s = 1.5$) for fly and deer shapes. Fig. 4 shows the results of the above repeated experiments for guitar shape. From the comparison results of accumulated 10 times with random noise, we can see the stability of our method.

B. Evaluation on Affine Registration

Similar to the evaluation on similarity registration, point sets are transformed using a particular affine transformation and parts of them are trimmed. And we add the non-Gaussian noise in the form as $v(i) = (1 - a(i))A(i) + a(i)B(i)$, where $A(i)$ denotes the background noise and $B(i)$ denotes large outliers. The background noises $A(i)$ are considered as a uniform distribution over $[-\sqrt{3}, \sqrt{3}]$. $B(i)$ is a Gaussian noise

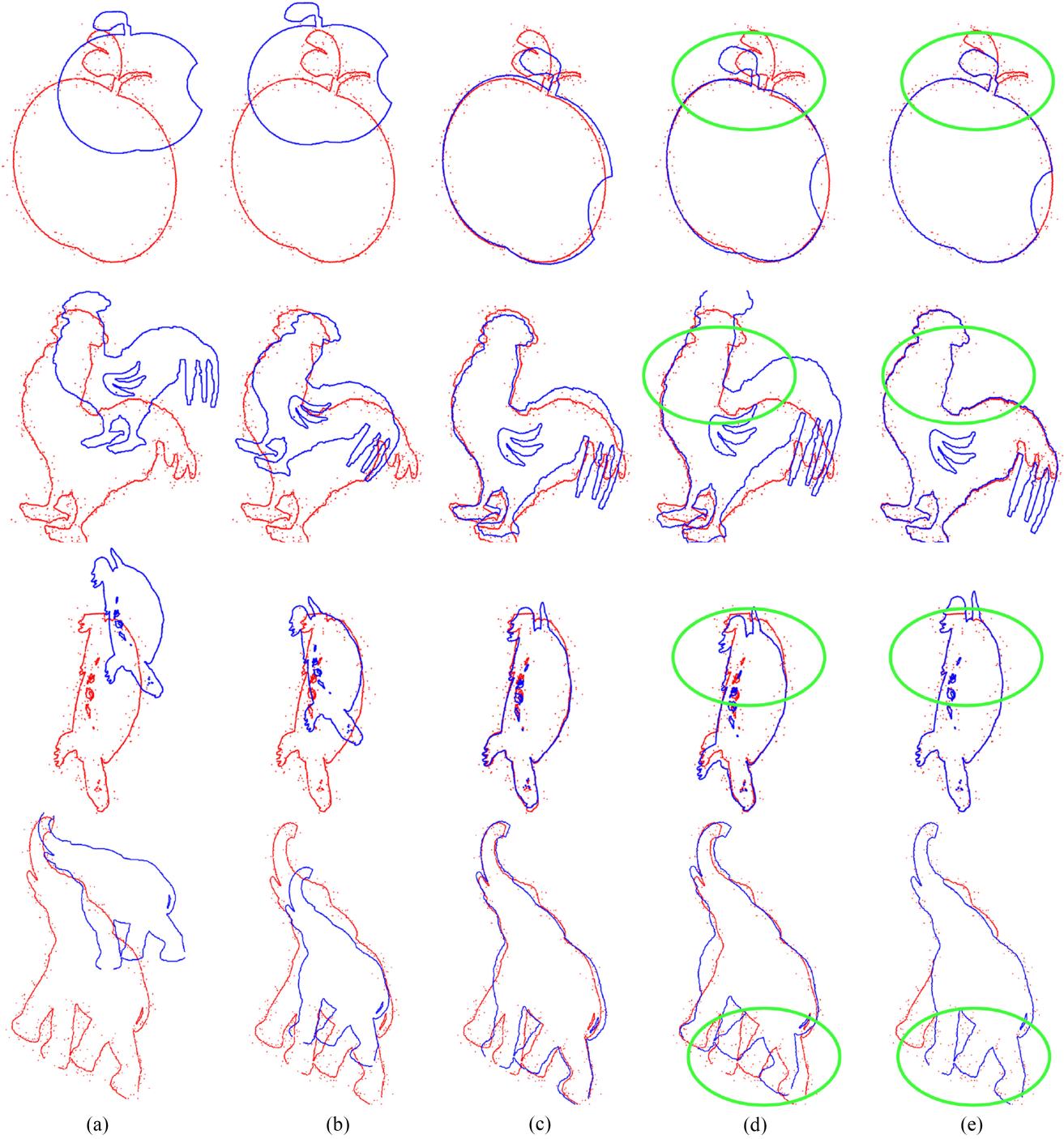


Fig. 5. Affine registration results on trimmed shapes with outliers. (a) Initialization, (b) result of ICP, (c) result of Bi-AICP, (d) result of BiC-AICP, and (e) result of our BiK-AICP.

$N(0, 15)$ with a big variance. $a(i)$ is a binary i.i.d. process, whose probability is set as $\Pr\{a(i) = 1\} = 0.25$ and $\Pr\{a(i) = 0\} = 0.75$.

There are some effective ways [11], [47] to provide an initial value for the affine registration. Here, we use the standard ICP results with rigid transformation as the initial state for further affine registration. And the power p in BiK-AICP is a free parameter for the nonlinear measure. Generally, good registration results can be achieved when p is in the range $[0.2, 8]$. For unknown

point sets, it is selected by a greedy searching strategy by setting $p = [0.2, 0.4, 0.6, \dots, 8]$ to achieve desirable performance.

First, another four kinds of shapes are selected. Fig. 5 demonstrates the registration results by different algorithms in an intuitive way. We can see that the ICP only includes a rigid transformation which can provide initial and coarse registration results. The Bi-AICP results are badly influenced by the trimmed shapes with outliers. We also observe that the complex noisy situation affects the accuracy of the BiC-AICP

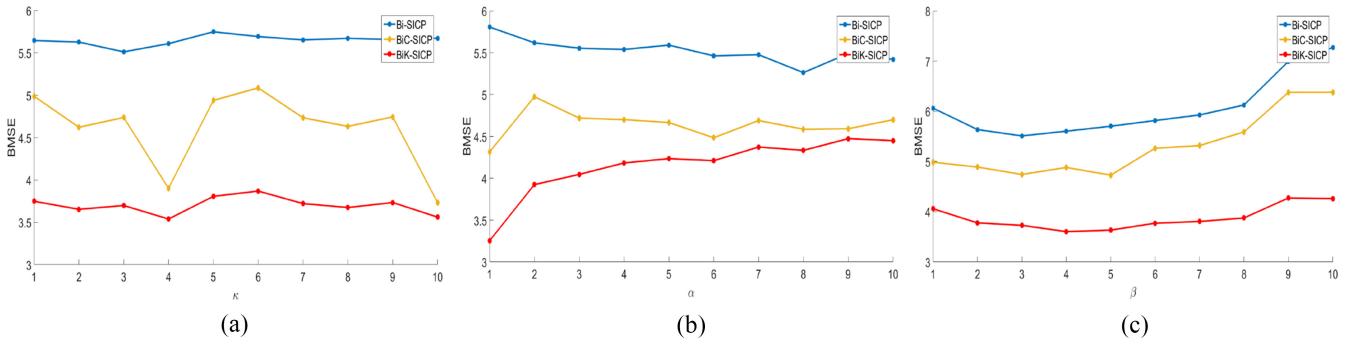


Fig. 6. Performance comparison of three different similarity registration algorithms, controlling parameters α, κ, β of sine wave noises yields to “dog” shape point set. (a) $\alpha = 2, \kappa = 1:10, \beta = 3$; (b) $\alpha = 1:10, \kappa = 1, \beta = 3$; and (c) $\alpha = 2, \kappa = 1, \beta = 1:10$.

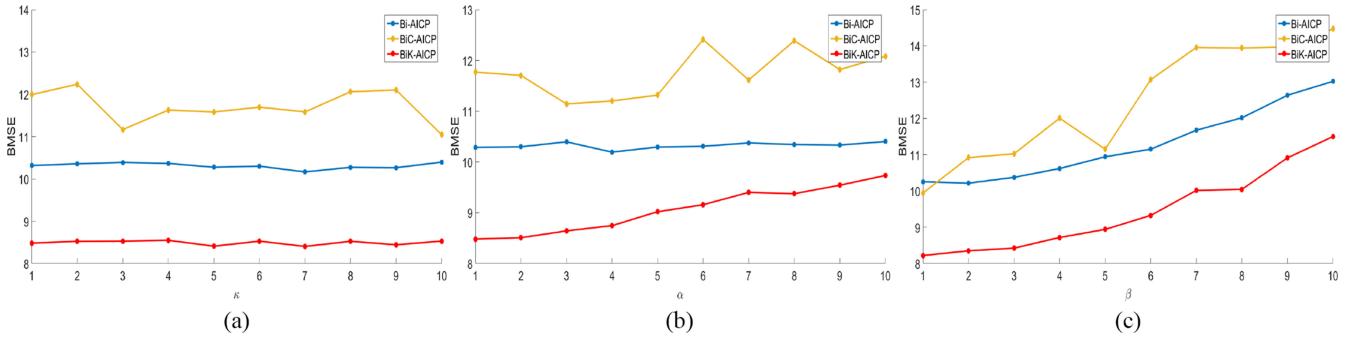


Fig. 7. Performance comparison of three different affine registration algorithms, controlling parameters α, κ, β of sine-wave noises yield to the “apple” shape point set. (a) $\alpha = 2, \kappa = 1:10, \beta = 3$; (b) $\alpha = 1:10, \kappa = 1, \beta = 3$; and (c) $\alpha = 2, \kappa = 1, \beta = 1:10$.

TABLE IV

REGISTRATION MEAN ERRORS ε_s , ε_R , AND $\varepsilon_{\vec{t}}$ OF THE SIMILARITY TRANSFORMATION $s = 0.5, \theta = \pi/5, \vec{t} = [20, 10]^T$, OBTAINED FROM THE DEER SHAPE BY 20 TIMES REPEATING EXPERIMENTS

Algorithm	Uniform			Binary			Sine Wave			Gauss		
	ε_s	ε_R	$\varepsilon_{\vec{t}}$									
SICP	0.1916	0.8082	11.5050	0.1719	0.8177	11.4742	0.1835	0.7296	11.7267	0.1793	0.8231	11.5750
Bi-SICP	0.0560	0.0120	5.1965	0.0600	0.0149	5.3878	0.0598	0.0136	5.3826	0.0601	0.0150	5.4117
BiC-SICP	0.0441	0.0052	4.5358	0.0495	0.0074	4.8085	0.0474	0.0057	4.7080	0.0475	0.0069	4.7417
BiK-SICP	0.0063	0.0000	1.7086	0.0017	0.0000	1.2284	0.0102	0.0001	2.1776	0.0058	0.0000	1.6767

algorithm. But corresponding points are completely aligned by our BiK-AICP algorithm.

We use the bidirectional MSE to compare the registration accuracy of different algorithms. Every experimental condition is repeated 20 times. We record the bidirectional MSE's mean and standard deviation. The statistical quantitative results for five point sets are demonstrated on Table III with the affine transformation $\vec{t} = [1, 2]^T$ and $\mathbf{A} = \begin{bmatrix} 1.3 & -0.4 \\ 0.6 & 1.04 \end{bmatrix}$. We can see that our BiK-AICP algorithm achieves the best performance among these algorithms with the smallest bidirectional MSE.

C. Robustness Test on Noise, Outliers, and Occlusion

In this section, we test the robustness of some algorithms on different kinds and levels of noise distributions.

First, the outliers are added the same as before. But we design different kinds of background noises $A(i)$. The following four kinds of noise cases are considered.

Case A: Uniform distribution over $[-\sqrt{3}, \sqrt{3}]$.

Case B: Binary distribution with $\Pr\{A(i) = -1\} = 0.5$ and $\Pr\{A(i) = 1\} = 0.5$.

Case C: Sine-wave noise $2 \sin(\omega)$, with ω uniformly distributing over $[0, 2\pi]$.

Case D: Zero-mean Gaussian noise with variance 1.0.

An alternative way to measure the performance is the parameter errors between estimated values and ground-truth parameters. For the similarity transformation parameters, they can be defined as $\varepsilon_s = |s - s_G|$, $\varepsilon_R = \|\mathbf{R} - \mathbf{R}_G\|_2$ and $\varepsilon_{\vec{t}} = \|\vec{t} - \vec{t}_G\|_2$. The compared results for the deer shape point set are illustrated in Table IV. We can see that for the above four kinds of noises, our BiK-SICP achieves the best performance among the four algorithms.

Then, to verify the influence of algorithm under different levels of noise and outliers, we specifically design the background noises $A(i)$ as a sine-wave $\alpha \sin(\kappa \omega)$, with ω uniformly distributed over $[0, 2\pi]$. And the outliers $B(i)$ are assumed to be a Gaussian noise with zero mean and $10 * \beta$ variance. We set α , κ , and β independently from 1 to 10, respectively. The

similarity registration results of bidirectional MSE curves for a dog shape point set with different parameters are displayed in Fig. 6. We can see that our BiK-SICP algorithm performs best in all situations. These results demonstrate that the BiK-SICP algorithm is very robust to noise parameters κ and β . But the registration error increases with noise parameter α .

Similarly, the affine registration results of the bidirectional MSE curves for the apple shape point set with different parameters are displayed in Fig. 7. We can see that our BiK-AICP algorithm also performs best in all situations. These results demonstrate that the BiK-AICP algorithm is very robust to the noise parameter κ . But the registration error increases with noise parameters β and α .

VII. CONCLUSION

This article proposed a novel registration method for point sets with the similarity and the affine transformations based on a newly defined bidirectional KMPE loss. Benefiting from its nonlinear similarity measure, the bidirectional KMPE loss function focuses on the regions of high density and is rather robust to the presence of outliers and noise. Two related optimization techniques are presented to estimate the registration parameters. Experiments on similarity and affine registrations indicate that the proposed method is more capable of dealing with the alignment under tough conditions, including the non-Gaussian noises and large outliers. These merits are obtained at the cost of introducing one more free parameter, that is, the power p . How to adaptively set this parameter will be an interesting subject for our future study.

REFERENCES

- [1] Y. Gao, M. Wang, D. Tao, R. Ji, and Q. Dai, “3-D object retrieval and recognition with hypergraph analysis,” *IEEE Trans. Image Process.*, vol. 21, no. 9, pp. 4290–4303, Sep. 2012.
- [2] Y. Yang, Q. Kou, S. Du, S. Luo, Y. Liu, and B. Wu, “An iterative feature-pair updating framework for rigid template matching with outliers,” in *Proc. IEEE Int. Symp. Multimedia*, 2017, pp. 200–207.
- [3] S. Du, Y. Guo, G. Sanroma, D. Ni, G. Wu, and D. Shen, “Building dynamic population graph for accurate correspondence detection,” *Medical Image Anal.*, vol. 26, no. 1, pp. 256–267, 2015.
- [4] P. J. Besl and N. D. McKay, “A method for registration of 3-D shapes,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 14, no. 2, pp. 239–256, Feb. 1992.
- [5] Z. Zhang, “Iterative point matching for registration of free-form curves and surfaces,” *Int. J. Comput. Vis.*, vol. 13, no. 2, pp. 119–152, Oct. 1994.
- [6] S. Rusinkiewicz and M. Levoy, “Efficient variants of the ICP algorithm,” in *Proc. Int. Conf. 3D Digit. Imag. Model.*, 2002, pp. 145–152.
- [7] D. Chetverikov, D. Stepanov, and P. Krsek, “Robust Euclidean alignment of 3D point sets: The trimmed iterative closest point algorithm,” *Image Vis. Comput.*, vol. 23, no. 3, pp. 299–309, Mar. 2005.
- [8] G. Godin, M. Rioux, and R. Baribeau, “Three-dimensional registration using range and intensity information,” in *Proc. SPIE Videometrics*, 2005, pp. 279–290.
- [9] J. M. Phillips, R. Liu, and C. Tomasi, “Outlier robust ICP for minimizing fractional RMSD,” in *Proc. Int. Conf. 3D Digit. Imag. Model.*, 2007, pp. 427–434.
- [10] S. Du, N. Zheng, S. Ying, and J. Wei, “ICP with bounded scale for registration of M-D point sets,” in *Proc. IEEE Int. Conf. Multimedia Expo.*, 2007, pp. 1291–1294.
- [11] S. Du, N. Zheng, G. Meng, and Z. Yuan, “Affine registration of point sets using ICP and ICA,” *IEEE Signal Process. Lett.*, vol. 15, pp. 689–692, Nov. 2008.
- [12] S. Du, N. Zheng, S. Ying, and J. Liu, “Affine iterative closest point algorithm for point set registration,” *Pattern Recognit. Lett.*, vol. 31, no. 9, pp. 791–799, Jul. 2010.
- [13] Y. Liu, “Automatic registration of overlapping 3D point clouds using closest points,” *Image Vis. Comput.*, vol. 24, no. 7, pp. 762–781, Jul. 2006.
- [14] J. Ma, J. Zhao, and A. L. Yuille, “Non-rigid point set registration by preserving global and local structures,” *IEEE Trans. Image Process.*, vol. 25, no. 1, pp. 53–64, Jan. 2016.
- [15] J. Ma, J. Zhao, J. Jiang, H. Zhou, and X. Guo, “Locality preserving matching,” *Int. J. Comput. Vis.*, vol. 127, no. 5, pp. 512–531, Sep. 2018.
- [16] M. Sofka, G. Yang, and C. V. Stewart, “Simultaneous covariance driven correspondence (CDC) and transformation estimation in the expectation maximization framework,” in *Proc. IEEE Int. Conf. Comput. Vis. Pattern Recognit.*, 2007, pp. 17–22.
- [17] S. Gold, C. Lu, A. Rangarajan, S. Pappu, and E. Mjolsness, “New algorithms for 2D and 3D point matching: Pose estimation and correspondence,” *Pattern Recognit.*, vol. 31, no. 8, pp. 1019–1031, Aug. 1998.
- [18] B. Jian and B. C. Vemuri, “Robust point set registration using Gaussian mixture models,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 33, no. 8, pp. 1633–1645, Aug. 2011.
- [19] G. McNeill and S. Vijayakumar, “A probabilistic approach to robust shape matching,” in *Proc. IEEE Int. Conf. Image Process.*, 2007, pp. 8–11.
- [20] S. Granger and X. Pennec, “Multi-scale EM-ICP: A fast and robust approach for surface registration,” in *Proc. Eur. Conf. Comput. Vis.*, 2002, pp. 418–432.
- [21] H. Chui and A. Rangarajan, “A new point matching algorithm for non-rigid registration,” *Comput. Vis. Understanding*, vol. 89, no. 2, pp. 114–141, Feb. 2002.
- [22] H. Chui, A. Rangarajan, J. Zhang, and C. M. Leonard, “Unsupervised learning of an Atlas from unlabeled point-sets,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 26, no. 2, pp. 160–172, Feb. 2004.
- [23] J. Ma, J. Jiang, Y. Gao, J. Chen, and C. Liu, “Robust image matching via feature guided Gaussian mixture model,” in *Proc. IEEE Int. Med. Imag.*, 2016, pp. 1–6.
- [24] J. Ma, J. Zhao, J. Jiang, and H. Zhou, “Non-rigid point set registration with robust transformation estimation under manifold regularization,” in *Proc. Assoc. Adv. Artif. Intell.*, 2017, pp. 4218–4224.
- [25] L. Bai, X. Yang, and H. Gao, “Nonrigid point set registration by preserving local connectivity,” *IEEE Trans. Cybern.*, vol. 48, no. 3, pp. 826–835, Mar. 2018.
- [26] G. D. Evangelidis and R. Horaud, “Joint alignment of multiple point sets with batch and incremental expectation-maximization,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 40, no. 6, pp. 1397–1410, Jun. 2018.
- [27] J. Ma, W. Qiu, J. Zhao, Y. Ma, A. L. Yuille, and Z. Tu, “Robust L_2E estimation of transformation for non-rigid registration,” *IEEE Trans. Signal Process.*, vol. 63, no. 5, pp. 1115–1129, Mar. 2015.
- [28] J. C. Principe, *Information Theoretic Learning: Renyi's Entropy and Kernel Perspectives*. New York, NY, USA: Springer, 2010.
- [29] B. Chen, Y. Zhu, J. Hu, and J. C. Principe, *System Parameter Identification: Information Criteria and Algorithms*. Waltham, MA, USA: Elsevier, 2013.
- [30] W. Liu, P. P. Pokharel, and J. C. Principe, “Correntropy: Properties and applications in non-Gaussian signal processing,” *IEEE Trans. Signal Process.*, vol. 55, no. 11, pp. 5286–5298, Nov. 2007.
- [31] B. Chen, X. Liu, H. Zhao, and J. C. Principe, “Maximum correntropy Kalman filter,” *Automatica*, vol. 76, pp. 70–77, Feb. 2017.
- [32] B. Chen, L. Xing, H. Zhao, N. Zheng, and J. C. Principe, “Generalized correntropy for robust adaptive filtering,” *IEEE Trans. Signal Process.*, vol. 64, no. 13, pp. 3376–3387, Jul. 2016.
- [33] X.-T. Yuan and B.-G. Hu, “Robust feature extraction via information theoretic learning,” in *Proc. Int. Conf. Mach. Learn.*, 2009, pp. 1193–1200.
- [34] A. Singh, R. Pokharel, and J. C. Principe, “The C-loss function for pattern classification,” *Pattern Recognit.*, vol. 47, no. 1, pp. 441–453, Jan. 2014.
- [35] L. Chen, H. Qu, J. Zhao, B. Chen, and J. C. Principe, “Efficient and robust deep learning with correntropy-induced loss function,” *Neural Comput. Appl.*, vol. 27, no. 4, pp. 1019–1031, May 2016.
- [36] E. Hasanbelliu, L. S. Giraldo, and J. C. Principe, “Information theoretic shape matching,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 36, no. 12, pp. 2436–2451, Dec. 2014.
- [37] S. Du, G. Xu, S. Zhang, X. Zhang, Y. Gao, and B. Chen, “Robust rigid registration algorithm based on pointwise correspondence and correntropy,” *Pattern Recognit. Lett.*, Jun. 2018. [Online]. Available: <https://doi.org/10.1016/j.patrec.2018.06.028>
- [38] Z. Wu, C. Chen, and S. Du, “Robust affine iterative closest point algorithm based on correntropy for 2D point set registration,” in *Proc. Int. Joint. Conf. Intell. Neural Netw.*, 2016, pp. 1415–1419.

- [39] B. Chen, L. Xing, X. Wang, J. Qin, and N. Zheng, "Robust learning with kernel mean p -power error loss," *IEEE Trans. Cybern.*, vol. 48, no. 7, pp. 2101–2113, Jul. 2018.
- [40] Y. Yang, W. Chen, B. Chen, S. Du, and L. Xiong, "Robust 2D point sets matching with kernel mean p -power error loss," in *Proc. IEEE Int. Conf. Syst. Man. Cybern.*, 2017, pp. 1898–1902.
- [41] J. Zhu, S. Du, Z. Yuan, Y. Liu, and L. Ma, "Robust affine iterative closest point algorithm with bidirectional distance," *IET Comput. Vis.*, vol. 6, no. 3, pp. 252–261, May 2012.
- [42] J. Dong, Z. Cai, and S. Du, "Improvement of affine iterative closest point algorithm for partial registration," *IET Comput. Vis.*, vol. 11, no. 2, pp. 135–144, Mar. 2017.
- [43] S. Ying, D. Li, B. Xiao, Y. Peng, S. Du, and M. Xu, "Nonlinear image registration with bidirectional metric and reciprocal regularization," *PLoS ONE*, vol. 12, no. 2, Feb. 2017, Art. no. e0172432.
- [44] K. S. Arun, T. S. Huang, and S. D. Blostein, "Least-squares fitting of two 3-D point sets," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. PAMI-9, no. 5, pp. 698–700, Sep. 1987.
- [45] L. J. Latecki, R. Lakamper, and T. Eckhardt, "Shape descriptors for non-rigid shapes with a single closed contour," in *Proc. IEEE Conf. Comput. Vis. Pattern. Recognit.*, vol. 1, 2000, pp. 424–429.
- [46] J. H. Zhu, N. N. Zheng, Z. J. Yuan, S. Y. Du, and L. Ma, "Robust scaling iterative closest point algorithm with bidirectional distance measurement," *Electron. Lett.*, vol. 46, no. 24, pp. 1604–1605, Nov. 2010.
- [47] S. Du *et al.*, "Robust iterative closest point algorithm based on global reference point for rotation invariant registration," *PLoS ONE*, vol. 12, no. 11, Nov. 2017, Art. no. e0188039.



Yang Yang (M'19) received the B.E. degree in information engineering from Xi'an Jiaotong University, Xi'an, China, in 2005, and the dual Ph.D. degrees in pattern recognition and intelligent systems from Xi'an Jiaotong University, and in systems innovation engineering from Tokushima University, Tokushima, Japan, in 2011.

She is currently an Associate Professor with the School of Electronic and Information Engineering, Xi'an Jiaotong University. Her current research interests include image processing, multimedia, and machine learning.



Dandan Fan received the B.E. degree in mechanical manufacturing and automation from Southwest Jiaotong University, Chengdu, China, in 2013, and the M.E. degree in software engineering from Xi'an Jiaotong University, Xi'an, China, in 2019.

Her current research interests include point set registration, and object detection and recognition.



Shaoyi Du received the dual bachelor's degrees in computational mathematics and in computer science, the M.S. degree in applied mathematics, and the Ph.D. degree in pattern recognition and intelligence systems from Xi'an Jiaotong University, Xi'an, China, in 2002, 2005, and 2009, respectively.

He is a Professor with Xi'an Jiaotong University. His current research interests include computer vision, machine learning, and pattern recognition.



Muyi Wang received the B.E. degree in electrical engineering and automation from the Changchun Institute of Technology, Changchun, China, in 2006, and the M.E. degree in power electronics and power drives from Northwestern Polytechnical University, Xi'an, China, in 2009. She is currently pursuing the Ph.D. degree in control science and engineering with Xi'an Jiaotong University, Xi'an.

Her current research interests include machine vision, pattern recognition, and big data processing.



Badong Chen (M'10–SM'13) received the B.S. and M.S. degrees in control theory and engineering from Chongqing University, Chongqing, China, in 1997 and 2003, respectively, and the Ph.D. degree in computer science and technology from Tsinghua University, Beijing, China, in 2008.

He was a Postdoctoral Researcher with Tsinghua University from 2008 to 2010; and a Postdoctoral Associate with the Computational NeuroEngineering Laboratory, University of Florida, Gainesville, FL, USA, from 2010 to 2012. In 2015, he visited the Nanyang Technological University, Singapore, as a Visiting Research Scientist. He also served as a Senior Research Fellow with Hong Kong Polytechnic University, Hong Kong, in 2017. He is currently a Professor with the Institute of Artificial Intelligence and Robotics, Xi'an Jiaotong University, Xi'an, China. He has published 2 books, 4 chapters, and over 200 papers in various journals and conference proceedings. His current research interests include signal processing, machine learning, artificial intelligence, and neural engineering and robotics.

Prof. Chen is an Associate Editor of the *IEEE TRANSACTIONS ON COGNITIVE AND DEVELOPMENTAL SYSTEMS*, the *IEEE TRANSACTIONS ON NEURAL NETWORKS AND LEARNING SYSTEMS*, and the *Journal of the Franklin Institute*, and has been on the editorial board of *Entropy*. He is a Technical Committee Member of IEEE SPS Machine Learning for Signal Processing and IEEE CIS Cognitive and Developmental Systems.



Yue Gao (SM'14) received the B.S. degree from the Harbin Institute of Technology, Harbin, China, and the M.E. and Ph.D. degrees from Tsinghua University, Beijing, China.

He is an Associate Professor with the School of Software, Tsinghua University. He has been with the School of Computing, National University of Singapore, Singapore, and with the Medicine School, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA.