

# Approximately Global Optimization for Robust Alignment of Generalized Shapes

Hongsheng Li, *Student Member, IEEE*, Tian Shen, *Student Member, IEEE*, and Xiaolei Huang, *Member, IEEE*

**Abstract**—In this paper, we introduce a novel method to solve shape alignment problems. We use gray-scale “images” to represent source shapes, and propose a novel two-component Gaussian Mixture (GM) distance map representation for target shapes. This asymmetric representation is a flexible image-based representation which is able to represent different kinds of shape data, including continuous contours, unstructured sparse point sets, edge maps, and even gray-scale gradient maps. Using this representation, a new energy function based on a novel two-component Gaussian Mixture distance model is proposed. The new energy function was empirically evaluated to be a more robust shape dissimilarity metric that can be computed efficiently. Such high efficiency is essential for global optimization methods. We adopt and modify one of them, the Particle Swarm Optimization (PSO), to effectively estimate the global optimum of the new energy function. Differently from the original PSO, several new strategies were employed to make the optimization more robust and prevent it from converging prematurely. The overall performance of the proposed framework as well as the properties of each algorithmic component were evaluated and compared with those of some state-of-the-art methods. Extensive experiments and comparison performed on generalized 2D and 3D shape data demonstrate the robustness and effectiveness of the method.

**Index Terms**—Shape alignment, point registration, matching, distance transform, particle swarm optimization.



## 1 INTRODUCTION

SHAPE registration is a crucial problem in computer vision and medical image analysis [46]. Shape alignment, also known as global shape registration, aims to recover a global transformation, such as rigid, similarity, or affine transformation, that brings the pose of a *source* (also known as *model* or *moving*) shape as close as possible to that of a *target* (also known as *template* or *fixed*) shape. The alignment has extensive uses in recognition, indexing and retrieval, and tracking. Shape alignment algorithms are usually different from each other in three aspects: shape representation, energy function, and optimization method.

In this paper, we introduce a novel shape alignment algorithm which has a robust shape dissimilarity measure defined on image-based representations and an efficient energy function whose optimum is found with a global optimizer. Inspired by [16], we propose a new asymmetric image-based representation which employs gray-scale images to represent source shapes and uses precomputed “distance” maps to represent target shapes. The gray-scale image representation for source shapes can represent shapes of arbitrary topology, even including generalized shapes such as gradient maps. Such flexibility enables directly using “raw” shapes, such as gradient maps of natural images, as source shapes and therefore avoids the need for shape

extraction (Section 4.5). For target shape “distance” maps, observing several drawbacks of the widely used nearest  $L_2$  distance model, we create a novel two-component univariate Gaussian Mixture (GM) distance model which achieves a high-peak-fat-tail effect. This model was empirically evaluated to be a more robust shape distance metric. It is different from the Gaussian Mixture models proposed in [5], [26], where an  $n$ -point set is modeled as an  $n$ -component bivariate (2D) or trivariate (3D) GM distribution if no prior knowledge is given and the problem of alignment is treated as fitting of GM distributions.

We also propose an energy function that can be computed efficiently by using precomputed target “distance” maps. Such efficiency enables us to use a global optimization method, the Particle Swarm Optimization (PSO), to find the globally optimal transformation parameters. The PSO, unlike the Particle Filtering (PF) in [37], is able to estimate the global optimum of a nonanalytic high-dimensional function without any a priori knowledge. In contrast, PF requires its proposal density to be properly approximated, which is of utmost importance in any Monte Carlo method. In addition to the original PSO, we detect “inactive particles” and use them to define a convergence criterion.

The main contribution of our algorithm is threefold: 1) A shape dissimilarity metric more robust than the widely used nearest  $L_2$  distance is created based on the new asymmetric shape representation and GM distance model. Moreover, the new representation can represent not only arbitrary shapes but also gradient maps, which are difficult for existing representations. 2) The high efficiency of the new energy function makes it feasible for the Particle Swarm Optimization, which has shown its effectiveness in various global optimization problems, to be used to recover the best transformation. 3) Based on the concept of “inactive

- The authors are with the Department of Computer Science and Engineering, Lehigh University, 19 Memorial Drive West, Bethlehem, PA 18015. E-mail: {h.li, tis207, xih206}@lehigh.edu.

Manuscript received 26 Nov. 2009; revised 5 July 2010; accepted 13 July 2010; published online 31 Aug. 2010.

Recommended for acceptance by K. Siddiqi.

For information on obtaining reprints of this article, please send e-mail to: [tpami@computer.org](mailto:tpami@computer.org), and reference IEEECS Log Number TPAMI-2009-11-0786.

Digital Object Identifier no. 10.1109/TPAMI.2010.169.

particles” from the Adaptive Particle Swarm Optimization (APSO) [52], we propose a convergence criterion for the PSO optimizer which reliably detects when the optimization process has found approximately the global optimum, while other global optimization frameworks—such as Simulated Annealing and Genetic Algorithm—usually lack convergence criteria.

We did extensive comparison with several state-of-the-art shape/point set registration algorithms. Experiments on a variety of shape data, which include continuous shapes, unstructured sparse point sets, edge maps, and gradient maps in both 2D and 3D, show the advantages of our algorithm.

## 2 BACKGROUND

Because of our interest in shape alignment, we mainly focus on the class of global transformation-based shape/point set matching methods. There has been a lot of previous research on this topic. We grouped the algorithms into the following four categories based on their shape representations.

### 2.1 Point-Based Representation

Point-based representations are widely used since they can represent shapes of arbitrary dimension and topology. The Iterative Closest Point (ICP) [3], [54] algorithm is one of the most widely used point sets alignment methods. The ICP algorithm iteratively associates points between the source and target point sets by the nearest neighbor criterion and transforms the source point set to minimize distances between these associated points. It terminates until the change in the sum of those distances falls below a given threshold. There are a large number of variants introduced on the basic ICP concept [36].

Although the standard ICP shows its effectiveness in various registration problems [36], it uses the  $L_2$  distance metric, which may not be a proper shape dissimilarity metric. It was observed in [45], [54] that longer “closest” distances tend to be between false correspondences, especially when outliers exist. To alleviate this problem, there are several mitigation measures proposed which can be categorized into two classes: rejection [11], [32], [45], [54] and weighting [17] (Section 3.3).

In ICP’s correspondence update step, for each point in the source shape it needs to find its nearest neighbor on the target shape. Therefore, it leads to an  $O(n^2)$  time complexity energy function. To accelerate the computation process of ICP, Fitzgibbon [16] suggested that the use of precomputed distance map could result in more efficient computation of the energy function. After the distance map of the target point set was calculated in linear time, the ICP energy function could also be calculated in linear time, which is asymptotically faster than the original ICP energy function’s quadratic time complexity. Most recently, Sandhu et al. [37] used the standard ICP energy function to match points, but applied PF in the iterative step to recover the best transformation. This algorithm expresses robust performance in partial matching and alignment in the presence of outliers.

Rangarajan et al. [35] proposed the Robust Point Matching (RPM) algorithm, in which an affine alignment

is interpreted as a mixed variable (binary and continuous) optimization problem. The correspondence problem is mapped into a linear assignment problem solved by softassign and deterministic annealing. Then, the transformation parameters are solved by least squares using obtained correspondences. Similarly to ICP, two such operations—correspondence assignment and transformation estimation—run alternatively until convergence. Chui and Rangarajan [6] further extended the PRM method to solve nonrigid point sets registration problems. The EM-ICP method [19] corresponds to an ICP with multiple matches weighted by normalized Gaussian weights in the case of Gaussian noises. It leads to an efficient matching algorithm based on Expectation-Maximization (EM) principles. Its experimental results demonstrated its improvements over ICP in terms of robustness and speed. Tsin and Kanade [44] proposed a kernel correlation-based point set registration approach where the cost function is proportional to the correlation of two kernel density estimates. This method can be considered as a robust, multiple-linked ICP. It has a built-in smoothing mechanism that makes it robust against noise and outlier corrupted data sets. Chui and Rangarajan [5] modeled the target point set by a Gaussian mixture and treated the source point set as sample data. Then, the point matching problem was treated as a mixture density estimation problem and solved by an EM-like algorithm. Another point sets registration method using Gaussian mixture models was proposed by Jian and Vemuri [26]. Unlike the work of [5], they treat two point sets symmetrically: Both point sets are modeled as Mixtures of Gaussian (MG) distributions.  $L_2$  distance is then used to measure dissimilarity between the distributions, which is minimized by Gradient Descent. Wang et al. [49], [48] employed the same MG distribution model, but extended it to groupwise point sets registration by using Jensen-Shannon (JS)/CDF-JS Divergence for atlas creation and distance measuring. A fully automatic approach for 3D point sets registration was proposed in [30]. Its crude alignment is based on the correlation of two Extended Gaussian Images (EGIs) in the Fourier domain and makes use of the spherical and rotational harmonic transform.

Another important family of point matching algorithms is RANSAC [15]. It randomly samples a minimal number of matches to estimate the geometric transformation between two point sets and then evaluates the estimated transformation using all points. But as the number of outliers increases, its computation time increases dramatically.

There are also lots of point matching algorithms proposed in the computational and digital geometry area [10], [14]. Most recently, Bhowmick et al. [4] proposed an approximate 2D point sets matching algorithm using a data structure called “Angular Tree,” which showed its effectiveness and efficiency on various test point sets.

### 2.2 Image-Based Representation

Image-based representations have recently gained increasing attention both in shape registration [22] and in statistical shape modeling [28]. They are obtained by embedding shapes into image planes. Most embedding methods are able to represent shapes of arbitrary dimension and topology. Through the embedding, an image-based representation

provides more constraints and supporting information from neighboring areas of the shape.

Distance transform is one of the most popular shape embedding methods. It is attractive in that it provides a generic distance function representation that naturally handles shapes of arbitrary dimension and topology. Because it represents shapes using their distance map “images,” it does not require explicit parameterization of the shapes. Paragios et al. [34] used signed distance functions whose zero level set represents original shapes. Shape dissimilarity is then calculated by computing the sum of squared differences (SSD) between two functions and minimized by Gradient Descent. Huang et al. [22] adopted the same level set shape representation and optimization method as those of [34]. The Mutual Information (MI) is employed to measure the dissimilarity between two distance functions. It is empirically evaluated and is shown to be more appropriate than SSD as a shape dissimilarity metric in estimating similarity and affine transformations between shapes. El Munim and Farag [13] kept a framework similar to [34], but euclidean distance functions are replaced by Vector Distance Functions (VDFs). Experimental results in [34], [22], [13] demonstrated the robustness of distance functions to shape perturbations and noise. In our experiments, however, we observed it is sometimes vulnerable to outliers.

Along this line, Tang and Hamarneh [42] computed various shape features, including geometric, appearance, and medial-axis-based shape features, which were then organized into vectors and assigned to the nearest pixels of corresponding shape parts to create feature images. However, the shapes this representation can represent are restricted by the feature descriptors it uses. For instance, all descriptors mentioned in [42] are for 2D closed contours. Moreover, the representation is based on the nearest neighbor transform and hence generates discontinuities in shape images, which may result in obstacles in the optimization.

### 2.3 Graph-Based Representation

Graph-based representations have a long history as shape representations. Methods based on this kind of representation usually directly recover point correspondences. Cross and Hancock [9] created graphs from point sets using Delaunay triangulation. The correspondence is recovered via inexact graph matching. In [51], given a 3D surface mesh, each vertex’s Gaussian curvature is calculated and projected into an extended boundary closed Markov Random Field (MRF). The correspondence between two meshes is then established by performing Gibbs sampling on the MRF. However, the Gaussian curvature is only locally isometric; hence, it cannot match meshes with significantly different scales. In [25], point sets are represented by a weighted undirected graph where pairwise distances specify weights between every two points. Then, point matching is obtained via a two-scale thermodynamics-based approach. Schmidt et al. [38] modeled the problem of 2D planar shape matching as finding the shortest path through a graph spanned by the two shapes, where nodes of the graph encode the local similarity of respective points on each contour. Huang et al. [21] proposed a Profile Hidden Markov Model (PHMM) for

2D planar shape modeling based on curvature descriptors. The special states and architecture in PHMMs can tolerate shape contour perturbations. This model shows effectiveness on planar shape registration and recognition.

### 2.4 Feature-Based Representation

Shape features can be further categorized as global and local features. On one hand, global features, similarly to other shape representations, describe an entire shape as an entity. Manay et al. [31] introduced a class of functionals which are invariant with respect to the euclidean group and are obtained by performing integral operations. Based on such integral invariants, a shape distance between matching parts is proposed which can be used for shape matching and recognition. Zhang and Fiume [53] proposed the normalized Fourier shape descriptor for 3D contours matching. It relies on normalizing the Fourier descriptors (FDs) of a 3D contour with respect to two FD coefficients corresponding to the lowest two frequencies. The remaining matching task only involves vertex shift and rotation about the z-axis. Medial axis has a long history as the shape representation for planar closed shapes. Liu and Greiger [29] used the A\* algorithm to match shape axis trees, which are defined by the locus of midpoints of optimally corresponding boundary points. Three local tree matching operations are introduced to yield optimally approximate matches. A variant of the medial axis is the shock graph, which is obtained by interpreting the medial axis as the converging points of wave propagation from shape boundaries [43]. One main advantage of the shock graph is that it no longer requires shapes to be closed contours. Sebastian et al. [39] presented a shape recognition framework which is based on matching shock graphs of 2D shape contours.

On the other hand, local features only have the ability to describe characteristics of a part of a shape. They are sometimes used as the basis for other kinds of representation [42], [38], [51], [21]. One most commonly used local feature is the curvature [21], [51], [38]. In [18], a shape is viewed as a set of line segments whose attributes are length and orientation. Matching these segments uniquely determines the similarity transformation between two shapes. An edit transformation which maps one shape to the other was proposed using dynamic programming. Belongie et al. [2] proposed the shape context descriptor. For every point, a log-polar space histogram recording the number of points in its neighborhood is calculated. Such histograms can then be used in shape matching and recognition.

### 2.5 Optimization Models and Methods in Shape Alignment

Various optimization models and methods are used in shape alignment. Similarly to other computer vision problems, one common way is to propose an energy function and then minimize it. Gradient-based local optimization methods such as Gradient Descent and Levenberg Marquardt are widely used [3], [13], [16], [22], [26], [34], [44]. Since their methods’ energy functions are usually highly nonconvex, those optimization methods can only find local optima.

Other optimization models and methods are also introduced. As mentioned before, RPM [6], [35] views point sets matching as a mixed binary and continuous variable

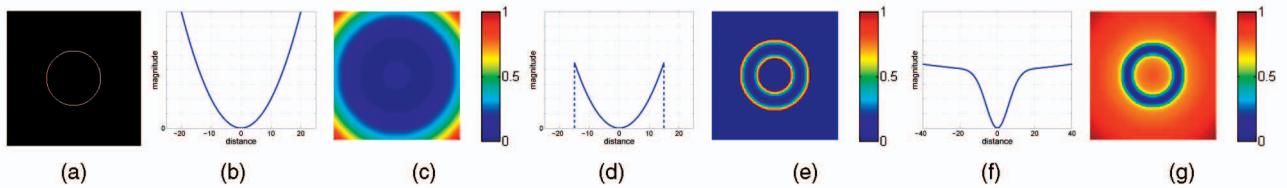


Fig. 1. (a) A circle ( $r = 40$ ). (b) The squared euclidean distance model. (c) Image embedding of (a) based on the model (b) (normalized). (d) The squared euclidean distance with threshold ( $D_{thres} = 15$ ) model. (e) Image embedding of (a) based on the model (d) (normalized). (f) The proposed two-component GM distance model. (g) Target shape representation of (a) based on the model (f) (normalized).

optimization problem, which is minimized by softassign and deterministic annealing. The 2D planar closed contour matching problem is modeled as finding a shortest path on a graph using graph cuts in [38]. In [42], the performance of a gradient-based optimizer is compared with that of a genetic algorithm known as the one-plus-one evolutionary algorithm for the registration of shape feature images. Particle filtering is used in ICP's iterative steps in [37]. More recently, Lempitsky and Boykov [27] presented a new regularization functional for the shape fitting problem, which maximizes the number of data points contained by a surface while tolerating some measurement errors. A touch-expand algorithm for minimum cuts is proposed to find the global minimum of its energy function.

### 3 METHODOLOGY

#### 3.1 The Standard ICP Energy Function

When point sets are used to represent shapes, let  $C_s$  and  $C_t$  be the source shape and target shape, respectively. Let  $\mathbf{x}_{i,s}$  ( $i = 1, \dots, N_s$ ) be the points on the source shape  $C_s$ , and let  $\mathbf{x}_{j,t}$  ( $j = 1, \dots, N_t$ ) be the points on the target shape  $C_t$ . The standard ICP energy function [3] based on squared euclidean distances is given by

$$\mathcal{F}(\Theta) = \frac{1}{N_s} \sum_{i=1}^{N_s} w_i d^2(\mathcal{T}(\mathbf{x}_{i,s}, \Theta), C_t), \quad (1)$$

where  $d(\mathbf{x}, C) = \min_{\mathbf{x}_c \in C} d(\mathbf{x}, \mathbf{x}_c)$  and  $d'(\mathbf{x}_1, \mathbf{x}_2)$  is the euclidean distance between the two points  $\mathbf{x}_1$  and  $\mathbf{x}_2$ .  $\Theta$  denotes the parameters of a chosen type of transformation, and  $\mathcal{T}(\mathbf{x}_{i,s}, \Theta)$  is the  $i$ th transformed source shape point according to  $\Theta$ .  $w_i$  denotes the weight of the  $i$ th "closest" distance, which is usually set to 1 in the absence of a priori knowledge.

#### 3.2 Previous Work on Computationally Efficient Shape Dissimilarity Metric

Our goal is to build a robust, generic, but computationally efficient shape dissimilarity metric because we want to employ global optimization methods other than commonly used gradient-based local optimization methods [13], [22], [34], [54], which have shown poor performance in overcoming local optima.

In [16], distance transform is applied to precompute a target shape's "closest" distances. Given a target shape  $C_t$ , the squared euclidean distance transform [33] of the target shape,  $\Phi_{C_t} : \Omega \rightarrow R^+$ , is

$$\Phi_{C_t}(x, y) = d^2((x, y), C_t), \quad (2)$$

where  $d((x, y), C_t)$  denotes the minimum  $L_2$  distance between the pixel at location  $(x, y)$  and the shape  $C_t$  (Figs. 1b and 1c). The time complexity of euclidean distance transform is  $O(n)$ , where  $n$  is the number of samples in the finite image domain. The ICP energy function based on the precomputed distance map  $\Phi_{C_t}$  [33] is then given by

$$E(\Theta) = \sum_{i=1}^{N_s} w_i \Phi_{C_t}(\mathcal{T}(\mathbf{x}_{i,s}, \Theta)). \quad (3)$$

Except for the normalization term, (3) has the same meaning as the standard ICP energy function (1). However, (3) is more computationally efficient because precomputed distance maps circumvent the need for searching for "closest" distances for every point  $\mathbf{x}_{i,s}$  in every iterative step. This technique is also related to Chamfer matching [1] and partial Hausdorff distance matching [24]. To alleviate the adverse effects of outliers,  $L_2$  distance is further replaced by the Huber norm in [16].

#### 3.3 New Asymmetric Shape Representation

Although the standard ICP energy function shows its effectiveness in various shape registration problems, as mentioned in Section 2.1, it uses the nearest  $L_2$  distance, which may not be a proper shape dissimilarity measure since longer "closest" distances are more likely to be between false correspondences. In agreement with this observation are our empirical experiments (Section 4.1.1) in which minimizing energy functions based on the nearest  $L_2$  distance sometimes led to wrong transformations.

Existing mitigation methods can be categorized into two classes: rejection [11], [32], [45], [54] and weighting [17]. One rejection approach [45], [54] is to set a maximum distance threshold,  $D_{thres}$ . If a "closest" distance is greater than  $D_{thres}$ , it would be rejected and not be used in the calculation of the energy function (Figs. 1d and 1e). However, this approach fails to properly measure the dissimilarity between two shapes that are significantly different in scale. For instance, a source shape which has all of its points' "closest" distances greater than  $D_{thres}$  would result in the rejection of all distances and thus cannot converge to the pose of the target shape. Another rejection policy, based on some multiple of the standard deviation of distances, is presented in [32]. But the deviation would be biased if strong outliers exist in the source shape. Furthermore, distance or point weights cannot be easily taken into consideration during the deviation computation. Rejecting distances that are not consistent with neighboring pairs was proposed in [11], although this rejection policy is  $O(n^2)$  complexity. All rejection-based methods cause

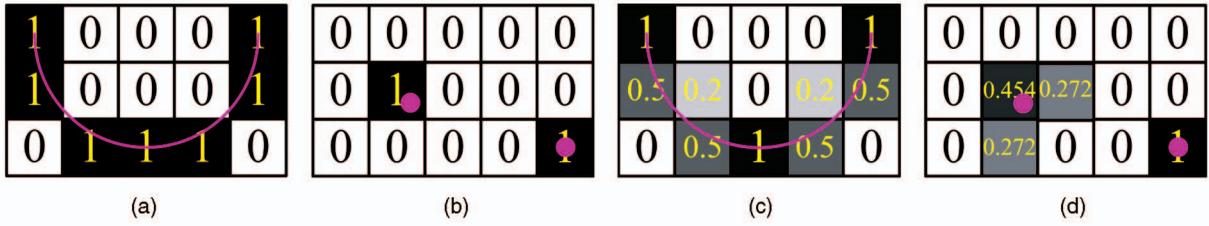


Fig. 2. Illustration of source shape representations of a parametric curve and a discrete point set. Source shape representations for the parametric curve (the purple curve) and discrete points (purple points) obtained through (a) and (b) discretization, and (c) and (d) antialiasing.

possible discontinuities in the energy function domain. A linear weighting policy which assigns longer “closest” distances lower weights,  $w_i = 1 - d_i/D_{\max}$ , is proposed in [17]. This scheme is sensitive to the  $D_{\max}$  value, however. It becomes similar to the  $L_2$  distance model when  $D_{\max}$  is large, and similar to the distance threshold model when  $D_{\max}$  is small.

### 3.3.1 Proposed Target Shape Representation

To address the possible false correspondence problem, we present a new target shape representation which replaces the squared euclidean distance model with a high-peak-fat-tail distance model. This model has an intuitive interpretation: The high peak indicates higher weight on short distances, and the fat tail decreases the weight for longer distances while keeping the energy function smooth and differentiable within the domain. In this paper, to achieve the high-peak-fat-tail effect, we choose a two-component GM model, leading to the following target shape representation:

$$\Psi_{C_t}(x, y) = -e^{-\Phi_{C_t}(x, y)/2\sigma_1^2} - \alpha e^{-\Phi_{C_t}(x, y)/2\sigma_2^2}, \quad (4)$$

where  $\Phi_{C_t}$  is the squared euclidean distance function of the target shape (2),  $\alpha \in (0, 1)$  is the GM weighting parameter,  $\sigma_1$  and  $\sigma_2$  ( $\sigma_1 < \sigma_2$ ) are the standard deviations of the two Gaussians (Figs. 1f and 1g). In practice, we set  $\alpha = 0.5$  and  $\sigma_1 = k_\sigma \sigma_2$ , where  $k_\sigma \in [5, 10]$ . Please note that the high-peak-fat-tail model is not limited to the proposed Gaussian Mixture model. Other models achieving the high-peak-fat-tail effect, such as the Pareto-Levy model, can also be used.

The Gaussian Mixture model is related to M-estimators in robust statistics [23]. Different objective functions of M-estimators can be used to alleviate the effects of outliers. However, our proposed GM model has its own advantages in the alignment framework. Compared with the Huber objective function, it more effectively suppresses the influence of outliers (Section 4.1.1). Compared with the bisquare objective function, it provides gradient toward the global minimum in fat-tail domains, while the bisquare model does not.

### 3.3.2 Proposed Source Shape Representation

For a source shape, we introduce a new gray-scale “image” representation that can represent generalized shapes including parametric curves/surfaces, sparse point sets, edge maps, and even gradient maps. Given a parametric shape or a discrete point set, discretization and antialiasing [20] (e.g., supersampling) techniques are employed to embed the

shape into the image plane. Fig. 2 illustrates how these techniques embed a parametric curve and a discrete point set into image planes using discretization and antialiasing. Clearly the antialiasing technique provides more accurate embedding results than the discretization does with limited image domain samples. Edge maps obtained by shape extraction algorithms, such as Canny edge detector, and even gradient maps can also be utilized directly as source shapes. Directly using gradient maps as source shapes in particular allows us to circumvent the shape extraction step and reduces the risk of misalignment because of improper shape extraction (Section 4.5).

Both the source and target shape representations can be easily extended to 3D by using 3D discretization (or antialiasing) and distance transform techniques, respectively.

## 3.4 Efficient Energy Function

Based on the source and target shape representations proposed in Section 3.3, we define an energy function measuring the dissimilarity between source and target shapes:

$$\mathcal{M}_\Psi(\Theta) = \frac{\iint \mathcal{T}(S, \Theta)(x, y) \cdot \Psi_{C_t}(x, y) dx dy}{\iint \mathcal{T}(S, \Theta)(x, y) dx dy}, \quad (5)$$

where  $\mathcal{T}(S, \Theta)$  is the transformed source shape image according to transformation parameters  $\Theta$ . The numerator of (5) calculates the sum of weighted “closest” GM distances between points on the transformed source shape image,  $\mathcal{T}(S, \Theta)$ , and the target shape,  $C_t$ . At a certain location  $(x, y)$ , according to the definition of the distance function (4),  $\Psi_{C_t}(x, y)$  is the closest GM distance between the point  $(x, y)$  and  $C_t$ . If  $\mathcal{T}(S, \Theta)(x, y)$  does not equal 0,  $\mathcal{T}(S, \Theta)(x, y) \cdot \Psi_{C_t}(x, y)$  equals the GM distance between the weighted point  $(x, y)$  on the  $\mathcal{T}(S, \Theta)$  and  $C_t$ . The weights are implicitly specified by the pixel values in  $\mathcal{T}(S, \Theta)$ . The denominator, the sum of point weights on the transformed source shape, is the normalization term necessary to handle scaling. Note that the new target shape representation  $\Psi_{C_t}$  based on the GM kernel is precomputed and most pixels in the source representation  $\mathcal{T}(S, \Theta)$  have zero values. Therefore, computation of the energy function (5) is highly efficient.

The time complexity of our energy function (5) can be explained as follows: Given two  $n$ -point sets, let  $k$  be the number of samples in the finite image domain. A one-time overhead to run the linear time distance transform algorithm is  $O(k)$  [33]. Then, it takes  $O(n)$  time to evaluate the energy function (5). In practice,  $k$  grows asymptotically slower than  $n$  does. Therefore, the overall time complexity can be further simplified from  $O(k + n)$  to  $O(n)$ . In contrast, the standard

ICP energy function takes  $O(n^2)$  by the brutal-force way or  $O(n \log n)$  by the k-d tree approach to evaluate [36].

### 3.5 Globally Optimal Transformation Estimation using the Particle Swarm Optimization

In previous research, local optimization methods such as Gradient Descent were widely used [13], [22], [34]. To these local methods, good initial parameter estimates are essential. To avoid this problem, some other optimization methods using particle filtering or genetic algorithms were also introduced [37], [42]. In our algorithm, the PSO is adopted to estimate the global optimum of the energy function (5). The PSO, which belongs to the class of swarm intelligence methods, is an effective optimization method for high-dimensional optimization problems, originally developed by Eberhart and Kennedy [12].

Let  $f(\mathbf{x})$  denote the function to be minimized. The basic particle swarm model consists of a swarm of  $m$  particles moving in an  $n$ -dimensional real valued variable space, each position of which potentially gives the global optimum of  $f(\mathbf{x})$  over a given domain. Let  $\mathbf{x}_i(t) = [x_1(t), x_2(t), \dots, x_n(t)]^T$  and  $\mathbf{v}_i(t) = [v_1(t), v_2(t), \dots, v_n(t)]^T$  be the  $i$ th particle's position and velocity at time  $t$ . Each particle knows the best position it has been at so far,  $pbest_i$ , and the overall best position obtained so far,  $gbest$ , by any particle in the swarm. Each particle updates its position and velocity according to the following equations:

$$\mathbf{v}_i(t+1) = \omega \mathbf{v}_i(t) + c_1 \cdot rand() \cdot (pbest_i - \mathbf{x}_i(t)) + c_2 \cdot rand() \cdot (gbest - \mathbf{x}_i(t)), \quad (6)$$

$$\mathbf{x}_i(t+1) = \mathbf{x}_i(t) + \mathbf{v}_i(t), \quad (7)$$

where  $\omega$  is the inertia weight representing the degree of the momentum of the particles.  $c_1$  and  $c_2$  are "cognitive" and "social" parameters modulating attraction terms that move the  $i$ th particle toward  $pbest_i$  and  $gbest$ , respectively.  $rand()$  generates pseudorandom numbers drawn from a uniform distribution in the range  $[0, 1]$ .  $\|\mathbf{v}_i\|$  is usually limited to be within a range  $[0, v_{\max}]$ , where  $v_{\max}$  is the maximum velocity.

Before the PSO computation begins, positions and speeds of all particles are randomly generated in the given continuous domain. Then, they are iteratively renewed according to (6) and (7) until a minimum error criterion or a predefined maximum iteration is attained.

Because the objective is to find the global optimum, we choose not to use the original PSO's maximum iteration stop criterion. In our modified PSO, to decide whether the value of the energy function has globally converged or not, first we use a relative error function,

$$\Delta f_i(t) = \frac{|f(\mathbf{x}_i(t)) - f(gbest)|}{|\min(f(\mathbf{x}_i(t)), f(gbest))|}, \quad (8)$$

proposed in [52], to determine inactive particles. If a particle's  $\Delta f_i(t)$  is less than a threshold for more than  $T_c$  time steps, we consider this particle to be inactive. If during a certain period of time, the number of inactive particles exceeds a threshold  $N_c$ , we consider the global optimum found. During the computation and before convergence, any detected inactive particles are randomly

relocated in the domain and are given random initial speeds so that they can keep searching the space.

The PSO has a strong ability to find the global optimum without any initial parameter estimate. In order to accelerate its convergence speed and to prevent premature termination, many variants have been introduced. Shi and Eberhart [41] presented a strategy of time-varying inertia weight:  $\omega$  linearly decreases as the number of iteration increases. The modified PSO in our algorithm also uses the strategy of time-varying inertia weight [41]. The difference from [41] is that each particle has its own inertia weight instead of all particles sharing the same one. If a particle is relocated, its inertia weight gets reinitialized to the maximum inertia weight. This enables a newly relocated particle to search more areas instead of directly traveling back to  $gbest$  and becoming inactive again.

For 2D shape alignment, the search space for particles has either four (similarity transformation) or six (affine transformation) dimensions. In 3D, we consider similarity transformation, which has an eight-dimensional search space. The combination of the new energy function (5) and the modified PSO allows us to solve shape alignment problems robustly and efficiently, as we will demonstrate in our experiments.

### 3.6 From 2D Alignment to 3D Alignment and Handling Local Deformations

The proposed alignment framework can easily be extended from 2D to 3D. For the source shape representation, existing 3D discretization or antialiasing techniques can be used to embed shapes into 3D image volumes. For the target shape representation, we create 3D image volumes using the results of 3D GM distance transform of target shapes. The PSO's particles now search in an eight-dimensional variable space for 3D cases instead of a four-dimensional one. In 3D, the eight-dimensional variable consists of translation (three-dimensional), rotation axis (three-dimensional), rotation angle (one-dimensional), and scaling factor (one-dimensional). Usually more particles are used in 3D cases than 2D cases because of the higher dimension of its variable space.

Similarly to [26], [13], [22], [42], our method can be extended to handle local deformations. After the global alignment is performed, the source shape can be fit to the target locally by minimizing the energy function (5) using local deformation models such as thin-plate spline (TPS) [26] and Free-Form Deformation (FFD) [22]. The optimization scheme remains the same, only using local deformation parameters. To curtail the computational cost, we can adopt a coarse-to-fine mechanism so that fewer parameters are to be optimized using PSO at a coarser deformation level, and then additional parameters that capture finer deformation details can be further solved using gradient descent.

## 4 EXPERIMENTS

For all experiments, we set  $\alpha = 0.5$ ,  $\sigma_1 = 5$ , and  $\sigma_2 = 50$  in (4), and let  $c_1 = c_2 = 2$  in (6).  $\omega$  in (6) linearly decreases from 1.0 to 0.2 in each particle's first 40 iterations. We used 100 particles for 2D cases and 3,000 particles for 3D. Most cases converged within 300 iterations. On a PC workstation with an Intel E6850

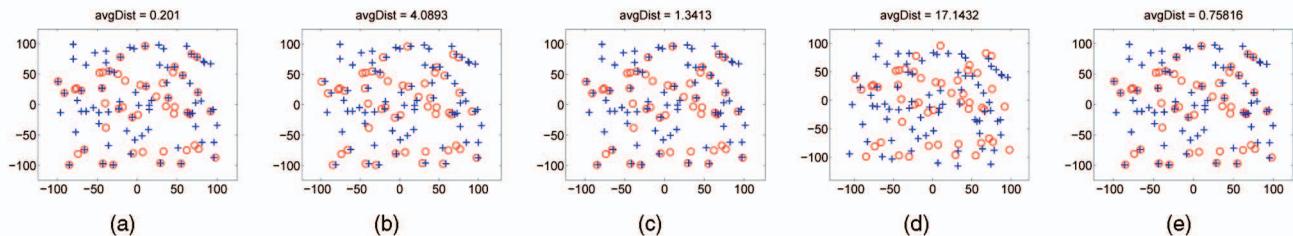


Fig. 3. Comparison between different dissimilarity metrics by aligning unstructured sparse point sets with strong outliers (source as blue “+” versus target as red “o”). (a) The result of the proposed GM kernel. (b) The result of nearest  $L_2$  distance model [3]. (c) The result of nearest Huber norm ( $\sigma = 10$ ) [16]. (d) The result of  $L_2$  distance between Mixtures of Gaussian [26]. (e) The result of Kernel Correlation [44].

CPU, single thread, a 2D case took about 1-5 s; a 3D case took around 60-200 s.

#### 4.1 Empirical Evaluation of the Proposed Method

To solve a problem by minimizing an energy function, it is important to validate that 1) the global minimum of the energy function corresponds to the desired solution, and 2) the minimization of the energy function can recover the global minimum point. In this section, we empirically evaluate the proposed shape dissimilarity metric and the modified PSO optimizer based on such two criteria and compare them with those of some state-of-the-art methods.

##### 4.1.1 The New Distance Model

We compared our new high-peak-fat-tail GM kernel<sup>1</sup> with other dissimilarity metrics including nearest  $L_2$  distance [3], nearest Huber norm [16],  $L_2$  distance between Mixtures of Gaussian [26]<sup>2</sup>, and Kernel Correlation [44]<sup>3</sup>. To eliminate the possibility that a certain distance kernel would prefer a specific shape or point pattern, we used challenging cases of aligning random 2D point sets with strong outliers.

Given a random 2D point set, we generate another point set and quantify its outlier and noise strength following the setup in [26]. The following procedures are used to generate source point sets from a target point set: For a target set with  $n$  points randomly spread in the range  $[-D, D] \times [-D, D]$  ( $n = 50$  and  $D = 100$  for 2D point sets), we generate a source set and control its degree of corruption by

1. discarding a subset of  $(1 - \rho)n$  points from the target set,
2. adding uniformly distributed noise of  $[-\epsilon, \epsilon] \times [-\epsilon, \epsilon]$  to all points in the target set,
3. applying a similarity transformation ( $s, \theta, t_x$ , and  $t_y$ ) to the target set (in this section, we use rigid transformation, i.e.,  $s = 1$ ), and
4. adding  $(\tau - \rho)n$  spurious, uniformly distributed points to the target set.

Therefore, after the corruption, a source set would have a total number of  $\tau n$  points, in which only  $\rho n$  have corresponding points in the target set. We quantify the strength of outliers and noise as  $S_o = (\tau - \rho)/\rho$  and  $S_n = \epsilon/D$ , respectively. The average  $L_2$  distance between known correspondences is

computed as the error measure and displayed on top of each registered frame.

In Fig. 3, the source point set with outlier strength  $S_o = 2$  and noise strength  $S_n = 0$  is aligned to the target point set with 50 points using rigid transformation (no scaling). To compare only the performance of different shape dissimilarity metrics and validate whether the global minimum of each dissimilarity metric corresponds to the desired solution, we initialize the source set’s pose to the known ground truth pose so that it is initially correctly aligned with the target set. We then register the point sets under those metrics using a common local minimization method.<sup>4</sup> The hypothesis is that, if the global optimum of a dissimilarity metric corresponds to the desired ground truth solution, we would observe little deviation in the converged pose after registration under that metric from the initial correct pose. We randomly generated over 1,000 pairs of random point sets under rigid transformations with outlier strength  $S_o = 2$  and noise strength  $S_n = 0$ , and evaluated different distance kernels’ final alignment errors based on those pairs. Results show that our GM kernel is the most robust to outliers and consistently leads to the smallest average distance error between corresponding points among all these metrics.

##### 4.1.2 The Modified PSO Optimizer

To evaluate the performance of the systematically “random” PSO, we compared it with 1) a “pure” random restart local optimizer and with 2) a local optimizer with a shape moment-based initialization using above-mentioned random pairs of 2D point sets. We chose the Levenberg Marquardt (LM) optimizer from LM-ICP [16]. The gradients of LM-ICP’s energy function with respect to transformation parameters can be well approximated by the chain rule and discrete gradients of the distance maps. Since we have demonstrated that the GM kernel generates a smaller alignment error, we abandoned the nearest Huber norm in LM-ICP and used the GM kernel for all optimizers. We tried to align random generated pairs of point sets under rigid transformations using all optimizers and recorded the number of successful alignments by each optimizer. We measured the average distance between corresponding points in the target and the aligned source sets, and judged one alignment as a successful one if its resulting average distance is less than some epsilon (we set it to 1 in this experiment). Each optimizer’s “success rate” can then be obtained by the number of successful alignments divided by the number of total alignments.

1. We set  $\alpha = 0.5$ ,  $\sigma_1 = 5$ , and  $\sigma_2 = 50$  for the GM model.

2. Since we followed the experimental setup in [26], we keep the parameter setting in the MG method’s original code.

3. We empirically set the bandwidth of the KC method to 10 based on experimental performance.

4. The `fminsearch` function in MATLAB is used.

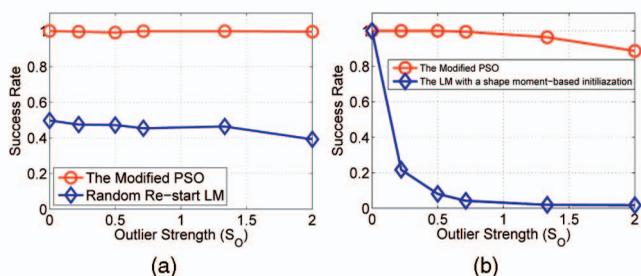


Fig. 4. Comparison between the success rates of the modified PSO and those of (a) the random restart Levenberg Marquardt and of (b) the LM with the shape moment-based initialization, using cases of rigid alignment of random point sets described in Section 4.1.2.

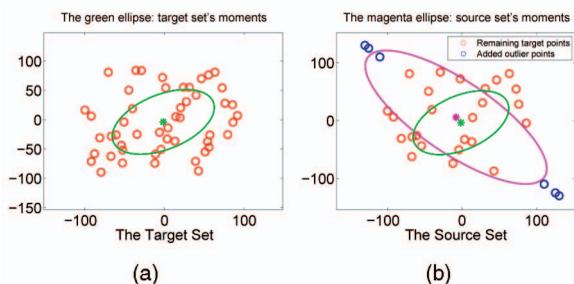


Fig. 5. Illustration of the sensitivity of the covariance to outliers. A point set's covariance matrix is illustrated by an ellipse. The center of the ellipse denotes the point set's mean, the directions of the ellipse's axes denote its covariance matrix's eigenvectors, and the lengths of the two axes are proportional to its covariance matrix's eigenvalues. (a) The target set (shown as red "o"). (b) The source set (obtained by removing 50 percent of the target points and adding six outlier points; the remaining target points and the added outlier points are shown in red and blue, respectively).

For the "pure" random restart LM, we generated 1,000 pairs of random point sets under every outlier strength level and tried to align them using both optimizers. However, the random restart LM does not have a convergence criterion. To make a fair comparison, for each alignment case, we record the processing time of the modified PSO once it terminates according to its convergence criterion and simply let the random restart LM run for the same amount of time. Statistics (Fig. 4a) shows that even when the outlier strength level is low, for the same processing time, random restart LM only has a success rate around 50 percent. As the outlier

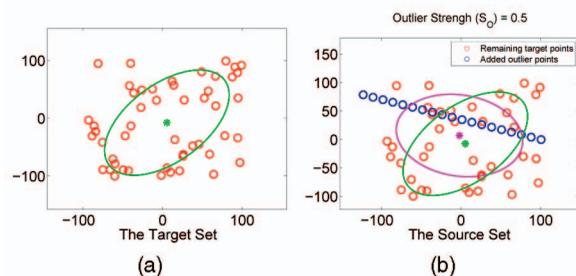


Fig. 6. One example of a pair of random point sets with line-form outliers. (a) The target point set. (b) The source point set (remaining points and added outlier points are shown in red and blue, respectively).

strength increases, the performance of random restart LM further degenerates. In contrast, our modified PSO consistently keeps a high success rate above 95 percent. This comparison demonstrated that the systematically randomized PSO performs more robustly than the "pure" random LM optimizer within the same running time.

For the LM with a shape moment-based initialization, we find the initialization by calculating the 2D point set's first moment (mean) for initial translation and its second-order moment (covariance) for initial rotation (and scale if the similarity transformation model is used). Without any outliers and noise, this strategy directly initializes the source point set to the correct pose. However, initial rotation and scale obtained from covariance are sensitive to even a small number of outliers. In Fig. 5, we show one example of how covariance fails to initialize the correct rotation and scale. In Fig. 5a, a target point set is shown as red "o" and its mean and covariance matrix are illustrated by a green ellipse. The center of the ellipse denotes the point set's mean, the directions of the ellipse's axes denote its covariance matrix's eigenvectors, and the lengths of the two axes are proportional to its covariance matrix's eigenvalues. In Fig. 5b, the source point set is obtained by 1) removing 50 percent of target points (the remaining points are shown as red "o") and 2) adding six outlier points (shown as blue "o"). The magenta ellipse illustrates the covariance matrix of the source set and is largely different from the green ellipse of the target set in rotation and scale. We did a further experiment comparing the PSO with random initialization and the LM with the shape moment-based initialization. At each outlier strength level, we

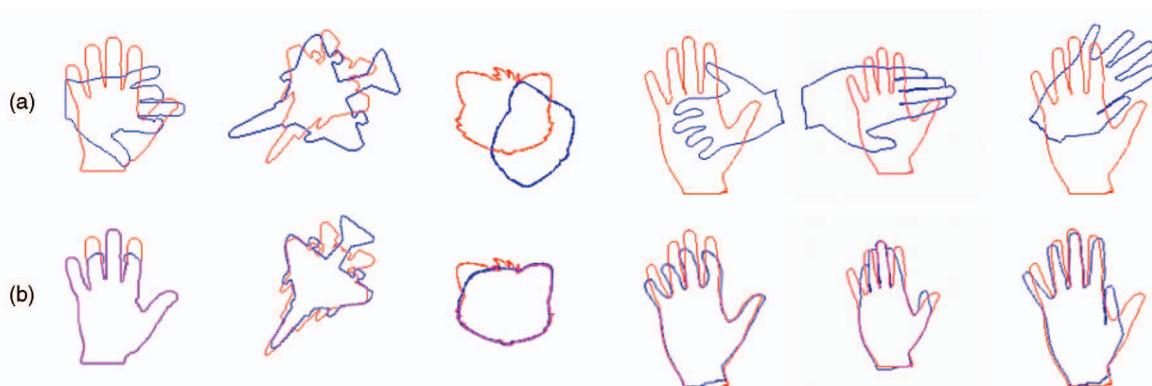


Fig. 7. Similarity alignment results. (a) Initial poses (source in blue versus target in red). (b) Alignment results.

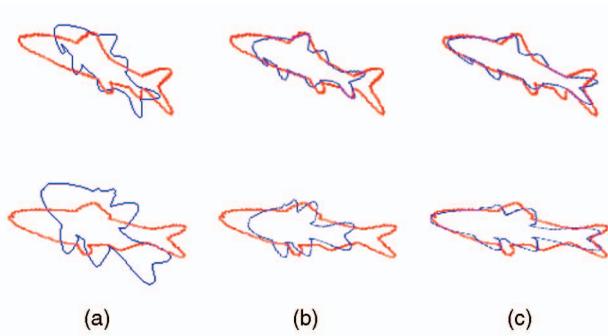


Fig. 8. Comparison between similarity and affine transformation. (a) Initial poses. (b) Results of similarity alignment. (c) Results of affine alignment.

generated 1,000 pairs of random point sets with line-form outliers and used them to perform the comparison. One pair of this kind of random point sets is shown in Fig. 6. This type of outliers is possible in real scenarios (e.g., range data in Section 4.6.2). Statistics of this comparison shown in Fig. 4b demonstrate that, as the outlier strength level rises, the success rate of LM with the shape moment-based initialization quickly drops to be less than 5 percent, while our PSO optimizer keeps a success rate around 90 percent.

## 4.2 Similarity and Affine Alignment of 2D Shapes

We used some shapes from the SIID shape database [40] to perform experiments of 2D similarity and affine shape alignment (Figs. 7 and 8). Source and target shapes in Fig. 7 differ in parts, and initial poses of the sources shapes vary in a broad range. Despite such difficulties, the results show that the energy function always converges to the global optimum through our modified PSO. Fig. 8 illustrates the difference between alignment results using similarity and affine transformations.

## 4.3 Alignment of Shapes with Strong Outliers and Comparison with State-of-the-Art Algorithms

In practical applications, shapes as clear and sharp as those in Fig. 7 are usually difficult to obtain because of complex backgrounds surrounding interested objects. Thus, extracted shapes may contain spurious contours or missing parts, which can significantly affect alignment results. Some state-of-the-art algorithms [13], [22], [34], [54] have difficulties in handling such shape registration problems with strong outliers. Because our method integrates outlier-resistant mechanisms based on the high-peak-fat-tail GM model and looks for the global optimum, it is well suited to solve this challenging problem. In our experiment, we used “clear” shapes as source shapes and aligned them to shapes with strong artificial outliers. In Fig. 9, artificial outliers “airplane”

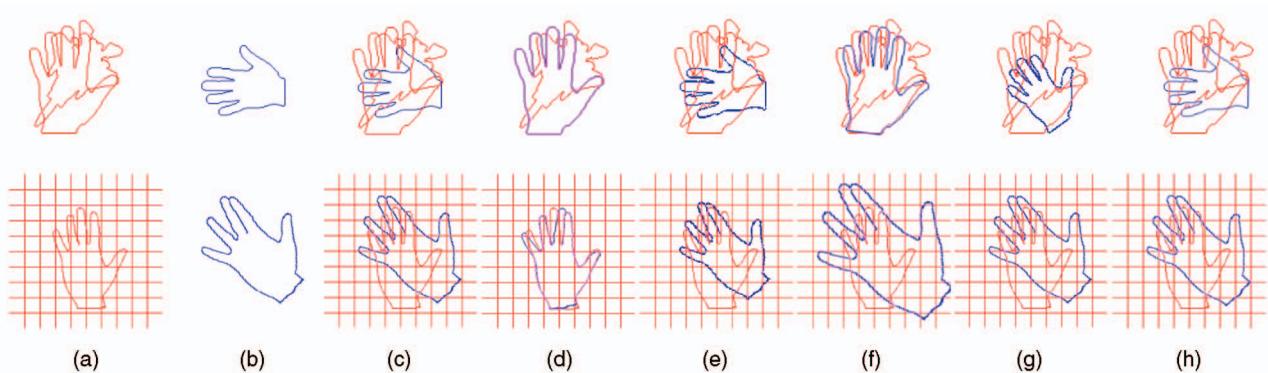


Fig. 9. Alignment results of “hand” shapes with strong outliers. (a) Target shapes. (b) Source shapes. (c) Initial poses (source in blue and target in red). (d) Alignment results by the proposed method. (e) Alignment results by ICP with distance threshold [54]. (f) Alignment results by Distance Functions (DF)-based method [34]. (g) Alignment results by MI-based method [22]. (h) Alignment results by VDF-based method [13].

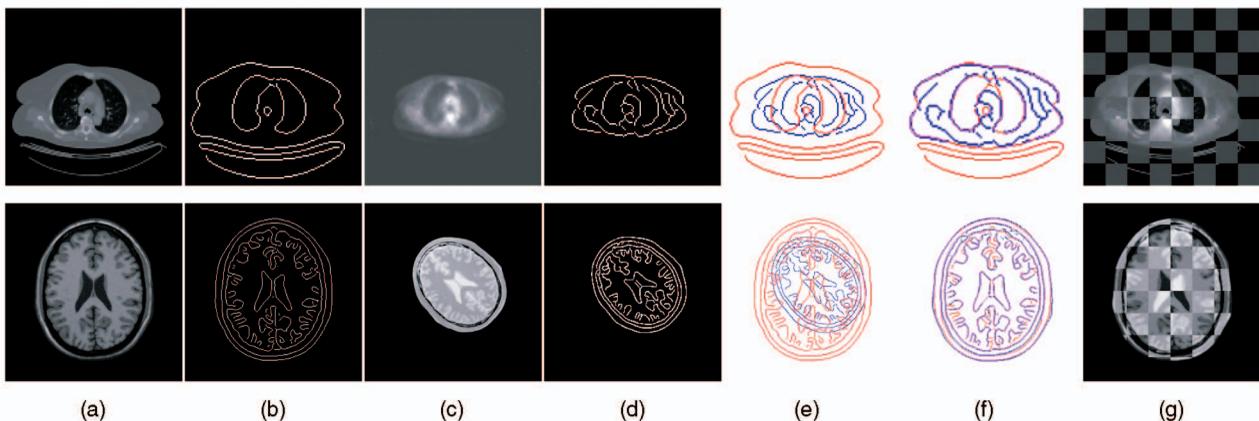


Fig. 10. Alignment of pairs of images through aligning their edge maps. (First row) CT/PET image alignment, and (second row) synthetic MRI image [8] alignment. (a) The target image. (b) Edges extracted from (a). (c) The source image. (d) Edges extracted from (c). (e) Initial poses. (f) Edge shape alignment results. (g) Source-target alternate checkerboard display based on the results in (f).

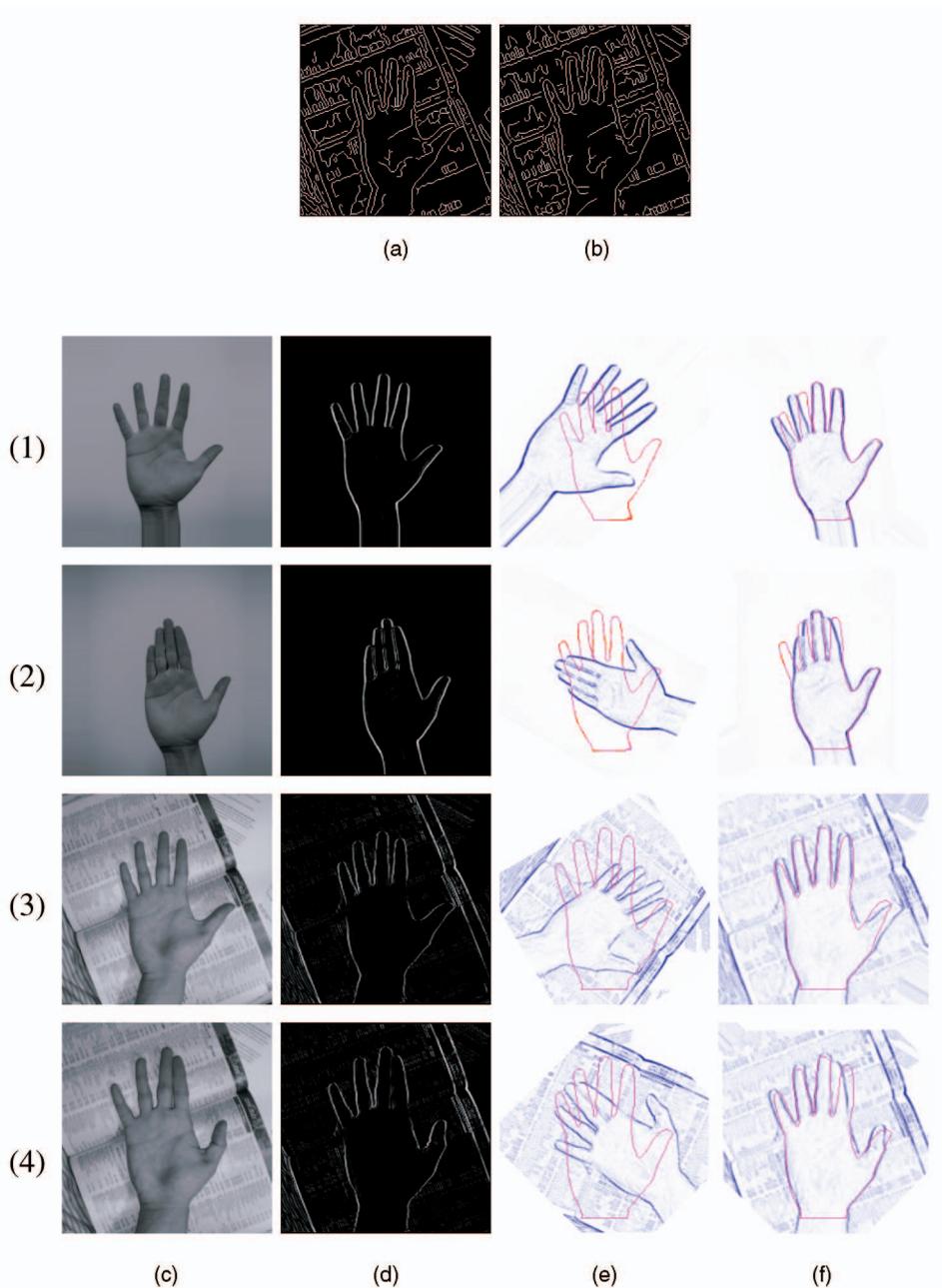


Fig. 11. Alignment between contour shapes and gradient maps. (a) Edges of (3)c extracted by Canny detector using MATLAB default parameters. (b) Edges of (4)c extracted by Canny detector using MATLAB default parameters. (c) Original images. (d) Gradient maps of (c) as source shapes. (e) Initial poses (source in blue versus target in red). (f) Alignment results.

and “grids” are added to “hand” shapes to create challenging target shapes. For this “hand” shape alignment problem, we compared the registration performance of our method with those in [13], [22], [34], [54]<sup>5</sup>. In the cases of Fig. 9, energy functions would have multiple local optima no matter which shape registration method is used. Finding the global optimum is impossible for methods using local optimizers unless an initial value close enough to the global optimum is given. In contrast, even with initial poses far away from global optima, our method successfully found the best transformation parameters because of PSO’s strong ability to systematically and efficiently search in a broad space to

find the global optimum. Other algorithms [13], [22], [34], [54] often got stuck at local optima (Figs. 9e, 9f, 9g, and 9h).

#### 4.4 Alignment of Edge Maps from Images

Shapes represented by edges were also used to evaluate the robustness of our algorithm. Many other shape representations [31], [38], [42] are only able to represent closed shape contours and therefore would have difficulty correctly representing edge shapes because of the existence of crossings and outliers. In our method, an edge map can directly be used as the “gray-scale” image representation of the source shape, and the GM distance transform is applied to the target shape. Fig. 10 shows two examples of aligning images through aligning edge maps extracted from them

5. We implemented the methods according to the original papers.

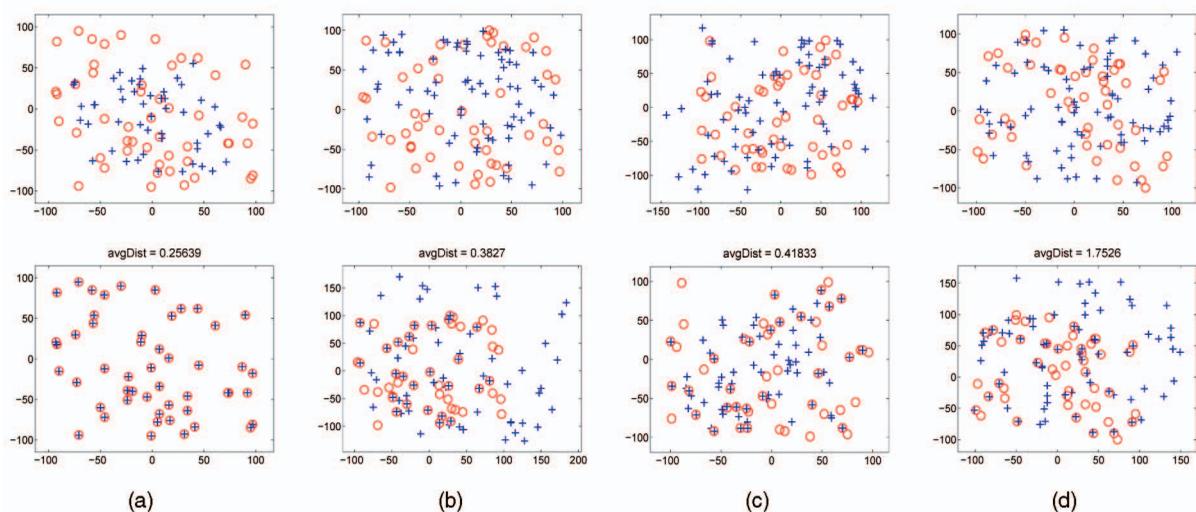


Fig. 12. Similarity alignment results of 2D unstructured point sets (source as blue “+”s versus target as red “o”s). (First row) Initial poses and (second row) alignment results. (a) Alignment of point sets without outliers and noise ( $S_o = 0$ ,  $S_n = 0$ ). (b) and (c) Alignment of point sets with outliers ( $\rho = 0.5$ ,  $\tau = 1.5$ ,  $S_o = 2.0$ , and  $S_n = 0$ ). (d) Alignment of point sets with outliers and noise ( $\rho = 0.5$ ,  $\tau = 1.5$ ,  $S_o = 2.0$ , and  $S_n = 2\%$ ).

TABLE 1

Empirical Convergence Ranges of the Proposed and MG [26] Methods in the 2D Unstructured Sparse Point Set Registration Tasks

Method	Outlier and noise level	Empirical convergence range
		(scale $\times$ rotation angle $\times$ translation in $x$ and $y$ directions)
MG method [26]	$S_o = 0, S_n = 0$	$[1, 1] \times [-120^\circ, 120^\circ] \times [-40, 40] \times [-40, 40]$ [26]
Proposed method	$S_o = 0, S_n = 0$	$[0.667, 1.5] \times [-180^\circ, 180^\circ] \times [-80, 80] \times [-80, 80]$
	$S_o = 2, S_n = 0$	$[0.75, 1.333] \times [-180^\circ, 180^\circ] \times [-40, 40] \times [-40, 40]$
	$S_o = 2, S_n = 2\%$	$[0.8, 1.25] \times [-180^\circ, 180^\circ] \times [-40, 40] \times [-40, 40]$

using our shape alignment algorithm. The first example (Fig. 10, first row) aligns a pair of CT/PET images. The second example (Fig. 10, second row) aligns a synthetic T2 MRI image with a T1 MRI image (from image source [8]). Our alignment method generated satisfactory results which indicate the potential extension of our method to image registration.

#### 4.5 Alignment of Generalized Shapes

Extracting edges from images is a difficult task because the choice of extraction algorithm parameters has a large effect on extracted shapes. In Figs. 11a and 11b, we show the edge maps of Figs. 11(3)c and 11(4)c, respectively, using Canny edge detector with default parameter settings in MATLAB’s implementation. Using such edge maps as shapes with very strong outliers, our algorithm (Section 4.4) failed to correctly align them to a template hand shape. To circumvent the difficulties caused by inappropriate edge extraction parameters or algorithms, our method is capable of directly using gradient maps as source shape images because of the natural characteristics of its gray-scale image-based shape representation. In Figs. 11(1) and 11(2), we demonstrate simple alignment cases where hands are on white backgrounds, gradient maps are used as

source shapes, and the GM distance transform of a hand template as the target shape. Cases in Figs. 11(3) and 11(4) are more challenging and closer to real applications where the background is a phone book with complex texture; our method successfully aligned the hand template shape to hands in these noisy images. So far we have only tested gradient maps as source shape images. However, other gray-scale images such as discretized line drawings or pencil drawings may also be considered as candidate shape images for alignment.

#### 4.6 Alignment of 2D Point Sets

##### 4.6.1 Unstructured Sparse Point Sets

To validate our algorithm on registering unstructured sparse 2D point sets, we followed the experimental setup of [26] and compared the performance of our algorithm with that of two other point registration methods, the Mixtures of Gaussian [26] and Kernel Correlation (KC) [44] methods. The code for the MG<sup>6</sup> and KC<sup>7</sup> methods is available at their authors’ websites.

6. <http://gmmreg.googlecode.com/svn/trunk/>.

7. <http://www.cs.cmu.edu/~ytsin/KCReg/>.

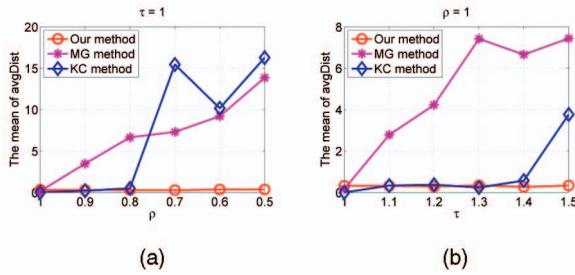


Fig. 13. Comparison between the proposed, MG [26] and KC [44] methods. The mean of 50 average distances upon convergence at each outlier level versus varying  $\rho$  or  $\tau$ . (a)  $\tau = 1, \rho \in [0.5, 1]$ . (b)  $\rho = 1, \tau \in [1, 1.5]$ .

First, we performed similarity alignment experiments on pairs of random point sets without noise and outliers. The source point set is a randomly transformed version of the target point set. Fig. 12a shows one example registration result. The target point set contains 50 points (red “o”s) randomly spread in a region  $[-100, 100] \times [-100, 100]$ . And the source point set (blue “+”s) is a transformed version of the target, after a 180 degree rotation and 1.5 scaling. Next, we also did experiments in the presence of noise and outliers following the setup mentioned in Section 4.1.1. Again,  $S_o = (\tau - \rho)/\rho$  and  $S_n = \epsilon/D$  quantify the strength of outliers and noise, respectively, and the average  $L_2$  distance between known correspondences is computed as the error measure and displayed on top of each registered frame (Figs. 12b, 12c, and 12d). We did extensive experiments to

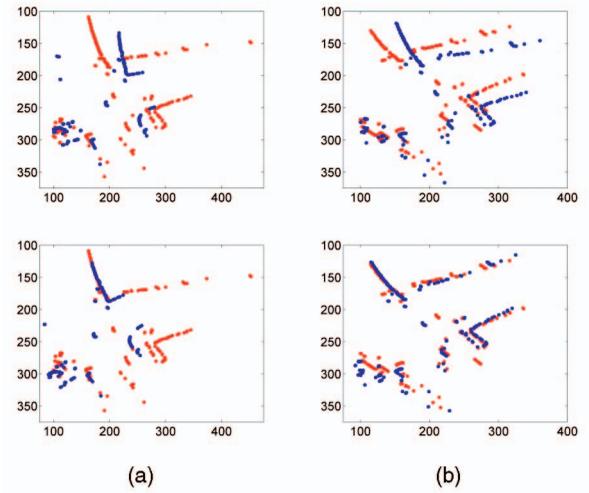


Fig. 15. Two examples of rigid alignment of 2D range scan point sets. (First row) Initial poses. (Second row) Alignment results. (a) Example 1. (b) Example 2.

obtain empirical convergence ranges for 2D unstructured sparse point set registration (Table 1).

We compared our method with two state-of-the-art point registration methods, the MG [26] and KC [44] methods with parameter settings mentioned in Section 4.1.1. The comparison was done using rigid transformation (no scaling). At each outlier or noise strength level, we generated 50 pairs of target and corrupted source sets. For

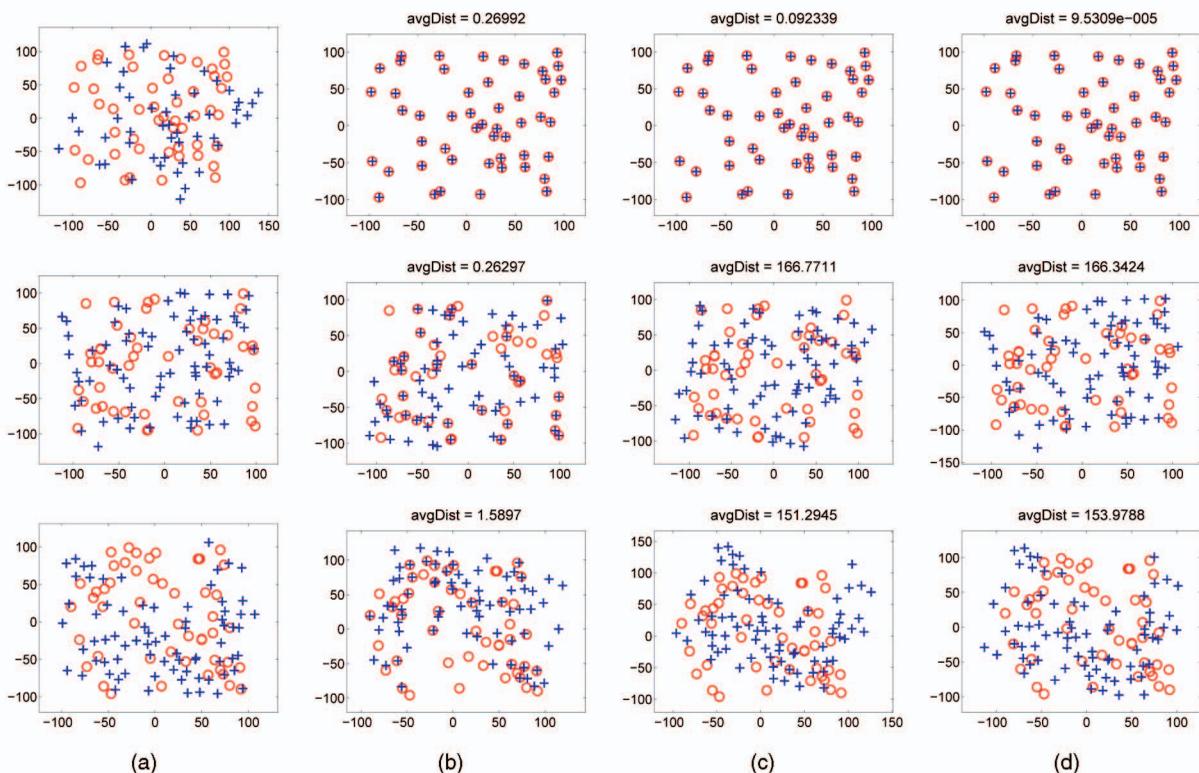


Fig. 14. Examples of alignment of 2D random point sets using the proposed, MG [26], and KC [44] methods under rigid transformation (source as blue “+”s versus target as red “o”s). (First row) Alignment of point sets without outliers and noise ( $S_o = 0$  and  $S_n = 0$ ). (Second row) Alignment of point sets with outliers ( $\rho = 0.5$ ,  $\tau = 1.5$ ,  $S_o = 2.0$ , and  $S_n = 0$ ). (Third row) Alignment of point sets with outliers and noise ( $\rho = 0.5$ ,  $\tau = 1.5$ , and  $S_o = 2.0$ ,  $S_n = 2\%$ ). (a) Initial poses. (b) Results by the proposed method. (c) Results by the MG [26] method. (d) Results by the KC [44] method.

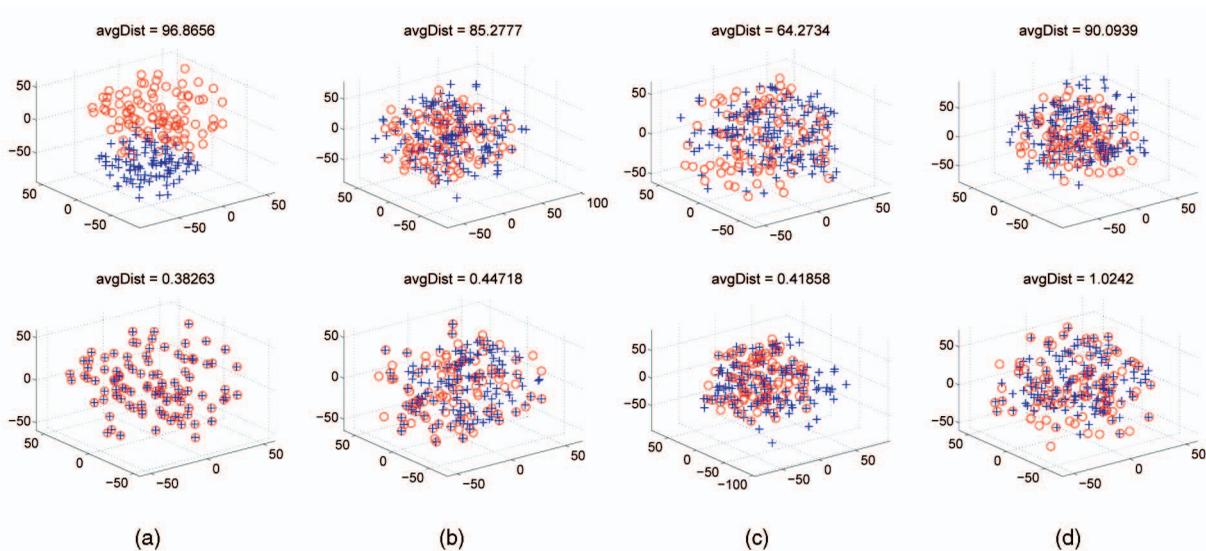


Fig. 16. Similarity alignment results of 3D unstructured sparse point sets. (First row) Initial poses (source as blue “+”s versus target as red “o”s). (Second row) Registration results. (a) Registration of point sets without outliers and noise ( $S_o = 0$  and  $S_n = 0$ ). (b) and (c) Registration of point sets with outliers ( $\rho = 0.5$ ,  $\tau = 1.5$ ,  $S_o = 2.0$ , and  $S_n = 0$ ). (d) Registration of point sets with outliers and noise ( $\rho = 0.5$ ,  $\tau = 1.5$ ,  $S_o = 2.0$ , and  $S_n = 2\%$ ).

each pair, we used the proposed, MG and KC methods for registration and computed the resulting average distance between known correspondences. Results show that when outlier and noise levels are low, both the MG and KC methods register two sets accurately. But as the outlier and noise levels rose, our method performed more robustly than the MG and KC methods (Figs. 13 and 14).

#### 4.6.2 Real Laser Scan Point Sets

Following the experimental setup in [44], we also used a set of 2D range data acquired by a SICK LMS 221 laser scanner [47] for performance evaluation. This data set consists of 3,715 scans acquired during an 18 min vehicle travel. We used every scan and the scan 30 time steps after it as the target and source point sets, respectively. Therefore, we were able to obtain 3,685 pairs of point sets for alignment. Alignment results generated by our method on those pairs were manually evaluated as either satisfactory or unsatisfactory. All except 358 alignments were evaluated as satisfactory. Two examples of the alignment experiments are shown in Fig. 15.

## 4.7 Similarity Alignment of 3D Point Sets

### 4.7.1 Unstructured Sparse Point Sets

Experiments similar to those with 2D unstructured sparse point sets were done on 3D point sets. For 3D point sets, the number of points  $n = 100$  and the range of points  $D = 50$ . Four examples of 3D unstructured point set alignment under different  $S_o$  and  $S_n$  settings are shown in Fig. 16. Empirical convergence ranges show the robustness of our algorithm for registration of 3D point sets (Table 2).

### 4.7.2 Real Laser Scan Point Sets

For performance evaluation, we followed the experimental setup in [37]. One thousand points were randomly chosen from the 3D “bunny” model [45]. We then generated translation  $\vec{t} = [t_x, t_y, t_z]$  from a normal distribution with each component having a standard deviation of 70; this value of standard deviation is chosen based on the range of the target set  $([-126, 127] \times [-125, 125] \times [-98, 98])$ . The rotation angle  $\theta$  is generated randomly along the  $z$ -axis, from a normal distribution  $\mathcal{N}(0, (\pi/3)^2)$ . The scaling factor is chosen randomly from a uniform distribution  $\mathcal{U}(0.7, 1.3)$ . Before the transformation is applied to the target set, a

TABLE 2  
Empirical Convergence Ranges of the Proposed Method in the 3D Unstructured Sparse Point Set Registration Tasks

Outlier and noise level	Empirical convergence range (scale $\times$ rotation angle $\times$ rotation axis $\times$ translation in $x$ , $y$ , and $z$ directions)
$S_o = 0, S_n = 0$	$[0.667, 1.5] \times [-180^\circ, 180^\circ] \times \text{all possible axes} \times [-40, 40] \times [-40, 40] \times [-40, 40]$
$S_o = 2, S_n = 0$	$[0.8, 1.25] \times [-180^\circ, 180^\circ] \times \text{all possible axes} \times [-20, 20] \times [-20, 20] \times [-20, 20]$
$S_o = 2, S_n = 2\%$	$[0.8, 1.25] \times [-180^\circ, 180^\circ] \times \text{all possible axes} \times [-20, 20] \times [-20, 20] \times [-20, 20]$

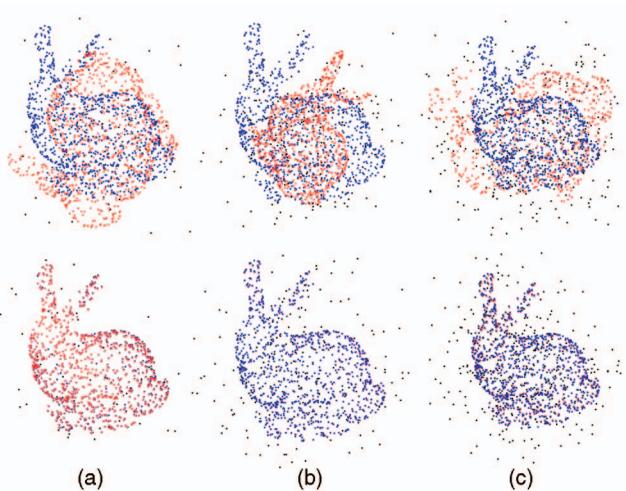


Fig. 17. Testing the robustness of the proposed method to noise and initialization with the 3D “bunny” model (source in red, target in blue, and noise points in black). Note that not all noise points are shown in the figure because of viewport cropping. (First row) Initial poses, and (second row) alignment results. (a) 5 percent Gaussian zero mean noise points. (b) 20 percent Gaussian zero mean noise points. (c) 35 percent Gaussian zero mean noise points.

certain percentage of points is replaced with zero-mean Gaussian noise  $\mathcal{N}(0, 60^2)$ . We consider three noise levels, with 5, 20, and 35 percent of the points replaced, respectively. We performed 100 experiments at each noise level. Three example alignments are shown in Fig. 17. The statistics of errors with respect to ground truth are given in Table 3.

TABLE 3  
3D Laser Scan Point Set Alignment Results: Mean ( $\mu$ ), Standard Deviation ( $\sigma$ ), and Maximum (Max) of Errors (Compared to Ground Truth) in Recovering the Scaling Factor  $s$ , Rotation Axis  $\vec{R}$ , Rotation Angle  $\theta$ , and Translation  $\vec{t}$  at Each Noise Level (with 100 Experiments per Noise Level)

Noise Level	5%	20%	35%
$\Delta s$	$\mu = 0.00165$ $\sigma = 0.00112$ max = 0.00382	$\mu = 0.00431$ $\sigma = 0.00251$ max = 0.00836	$\mu = 0.0122$ $\sigma = 0.0132$ max = 0.0798
$\ \Delta \vec{R}\ $	$\mu = 0.00258$ $\sigma = 0.00127$ max = 0.00549	$\mu = 0.00900$ $\sigma = 0.0113$ max = 0.0580	$\mu = 0.00893$ $\sigma = 0.00845$ max = 0.0340
$\Delta \theta$	$\mu = 0.124^\circ$ $\sigma = 0.0986^\circ$ max = $0.324^\circ$	$\mu = 0.182^\circ$ $\sigma = 0.155^\circ$ max = $0.669^\circ$	$\mu = 0.394^\circ$ $\sigma = 1.10^\circ$ max = $6.10^\circ$
$\ \Delta \vec{t}\ $	$\mu = 0.670$ $\sigma = 0.274$ max = 1.25	$\mu = 0.951$ $\sigma = 0.710$ max = 2.92	$\mu = 1.20$ $\sigma = 0.705$ max = 2.70

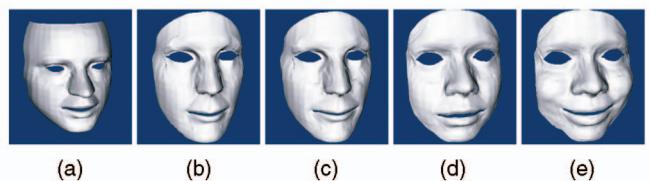


Fig. 18. Real people’s face models [50] used in 3D surface registration. (a) The male target face model. (b) and (c) Faces of another male subject (as source surface). (d) and (e) Faces of a female subject (as source surface).

#### 4.8 Similarity Alignment of 3D Surfaces

We also did experiments on 3D surface registration. A male’s face (Fig. 18a) is used as the target surface and two other people’s faces with different facial expressions (smiling and neutral) are aligned to it (Figs. 18b, 18c, 18d, and 18e) [50]. One can see that the source surfaces are greatly different from the target surface model. Fig. 19 shows four examples of this experiment. Our method successfully recovered the optimal transformation even when the initialization was far from the correct pose.

### 5 CONCLUSIONS AND DISCUSSIONS

In this paper, we present a novel shape alignment method. Flexible gray-scale “images” and GM “distance” maps were proposed to represent source and target shapes, respectively. Gray-scale “images” for source shapes can represent

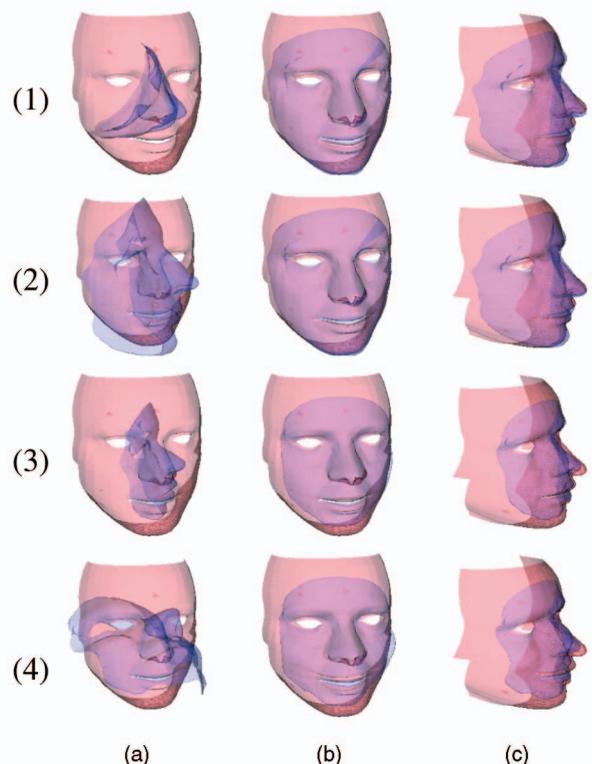


Fig. 19. 3D surface alignment experiments. (1) Alignment of models in Figs. 18a and 18b. (2) Alignment of models in Figs. 18a and 18c. (3) Alignment of models in Figs. 18a and 18d. (4) Alignment of models in Fig. 18a and Fig. 18e. (a) Initial poses (source in blue versus target in red). (b) Alignment results (front view). (c) Alignment results (side view).

shapes with arbitrary dimension and topology, including continuous contours, unstructured sparse point sets, edge maps, and even gradient maps. GM “distance” maps are precomputed by the novel two-component GM distance transform, which provides a more robust dissimilarity metric for shapes. Because of the high efficiency of our new energy function, a global optimum estimation algorithm, the Particle Swarm Optimization, was introduced to optimize the energy function. Several additional strategies were used to modify the original PSO so as to provide a convergence criterion and prevent the optimization from stopping prematurely. The proposed method showed its robustness and effectiveness in solving various challenging registration cases using generalized shape data.

In the end, some limitations of the proposed method should be discussed. Similarly to the MG [26] and the KC [44] methods, the setting of parameters (i.e.,  $\sigma_1$  and  $\sigma_2$ ) for our two-component Gaussian Mixture distance model is still an open issue and may affect the performance of the alignment, although in our experiments it is observed that the performance of our method is less sensitive than the MG and KC methods to parameter settings. Our current modified PSO does not utilize gradient information of the energy function even if the outlier level is low. Some PSO variants incorporated gradient information into the original PSO framework and reported acceleration on the convergence speed of the optimization process. The gradients of our energy function can be well approximated by the chain rule and discrete gradients of distance maps. Although the convergence and stability of the standard PSO have been theoretically studied in [7], further analysis of the convergence properties of our modified PSO (theoretical convergence guarantee, convergence with respect to the number of swarm particles, etc.) is necessary since the modified PSO has more complex behaviors because it reinitializes inactive particles.

## ACKNOWLEDGMENTS

This work was partially supported by the US National Science Foundation (NSF) Grant IIS-0812120, and US National Institutes of Health (NIH) Grant R21GM083928.

## REFERENCES

- [1] H.G. Barrow, J.M. Tenenbaum, R.C. Bolles, and H.C. Wolf, “Parametric Correspondence and Chamfer Matching: Two Techniques for Image Matching,” *Proc. Int’l Joint Conf. Artificial Intelligence*, pp. 1175-1177, 1977.
- [2] S. Belongie, J. Malik, and J. Puzicha, “Shape Matching and Object Recognition Using Shape Contexts,” *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 24, no. 4, pp. 509-521, Apr. 2002.
- [3] P. Besl and H. McKay, “A Method for Registration of 3-D Shapes,” *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 14, no. 2, pp. 239-256, Feb. 1992.
- [4] P. Bhowmick, R.K. Pradhan, and B.B. Bhattacharya, “Approximate Matching of Digital Point Sets Using a Novel Angular Tree,” *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 31, no. 5, pp. 769-782, May 2009.
- [5] H. Chui and A. Rangarajan, “A Feature Registration Framework Using Mixture Models,” *Proc. IEEE Workshop Math. Methods in Biomedical Image Analysis*, pp. 190-197, 2000.
- [6] H. Chui and A. Rangarajan, “A New Point Matching Algorithm for Non-Rigid Registration,” *Computer Vision and Image Understanding*, vol. 89, pp. 114-141, 2003.
- [7] M. Clerc and J. Kennedy, “The Particle Swarm—Explosion, Stability, and Convergence in a Multidimensional Complex Space,” *IEEE Trans. Evolutionary Computation*, vol. 6, no. 1, pp. 58-73, Feb. 2002.
- [8] C. Cocosco, V. Kollokian, R. Kwan, and A. Evans, “BrainWeb: Online Interface to a 3D MRI Simulated Brain Database,” *Proc. Int’l Conf. Functional Mapping of the Human Brain*, 1997.
- [9] A.D.J. Cross and E.R. Hancock, “Graph Matching with a Dual-Step EM Algorithm,” *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 20, no. 11, pp. 1236-1253, Nov. 1998.
- [10] P.J. de Rezende and D.T. Lee, “Point Matching in d-Dimensions,” *Algorithms*, vol. 13, pp. 387-404, 1995.
- [11] C. Dorai, J. Weng, and A. Jain, “Optimal Registration of Object Views Using Range Data,” *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 19, no. 10, pp. 1131-1138, Oct. 1997.
- [12] R. Eberhart and J. Kennedy, “A New Optimizer Using Particle Swarm Theory,” *Proc. Int’l Symp. Micro Machine and Human Science*, pp. 39-43, 1995.
- [13] H. El Munim and A. Farag, “Shape Representation and Registration Using Vector Distance Functions,” *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, pp. 1-8, 2007.
- [14] A. Erafat and A. Itai, “Improvements on Bottleneck Matching and Related Problems Using Geometry,” *Proc. Ann. ACM Symp. Computational Geometry*, pp. 301-310, 1996.
- [15] M.A. Fischler and R.C. Bolles, “Random Sample Consensus: A Paradigm for Model Fitting with Applications to Image Analysis and Automated Cartography,” *Comm. ACM*, vol. 24, pp. 381-395, 1981.
- [16] A. Fitzgibbon, “Robust Registration of 2D and 3D Point Sets,” *Image and Vision Computing*, vol. 21, pp. 1145-1153, 2003.
- [17] G. Godin, M. Rioux, and R. Baribeau, “Three-Dimensional Registration Using Range and Intensity Information,” *Proc. SPIE: Videometrics III*, 1994.
- [18] S. Gold and A. Rangarajan, “A Graduated Assignment Algorithm for Graph Matching,” *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 18, no. 4, pp. 377-388, Apr. 1996.
- [19] S. Granger and X. Pennec, “Multi-Scale EM-ICP: A Fast and Robust Approach for Surface Registration,” *Proc. European Conf. Computer Vision*, pp. 418-432, 2002.
- [20] F.S. Hill and S.M. Kelley, *Computer Graphics Using OpenGL*. Prentice Hall, 2007.
- [21] R. Huang, V. Pavlovic, and D.N. Metaxas, “Embedded Profile Hidden Markov Models for Shape Analysis,” *Proc. IEEE Int’l Conf. Computer Vision*, pp. 1-8, 2007.
- [22] X. Huang, N. Paragios, and D.N. Metaxas, “Shape Registration in Implicit Spaces Using Information Theory and Free Form Deformations,” *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 28, no. 8, pp. 1303-1318, Aug. 2006.
- [23] P.J. Huber, *Robust Statistics*. Wiley, 1981.
- [24] D.P. Huttenlocher, G.A. Klanderman, and W.J. Rucklidge, “Comparing Images Using the Hausdorff Distance,” *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 15, no. 9, pp. 850-863, Sept. 1993.
- [25] A. Jagannathan and E. Miller, “Unstructured Point Cloud Matching within Graph-Theoretic and Thermodynamic Frameworks,” *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, vol. 2, pp. 1008-1015, 2005.
- [26] B. Jian and B. Vemuri, “A Robust Algorithm for Point Set Registration Using Mixture of Gaussians,” *Proc. IEEE Int’l Conf. Computer Vision*, vol. 2, pp. 1246-1251, 2005.
- [27] V. Lempitsky and Y. Boykov, “Global Optimization for Shape Fitting,” *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, pp. 1-8, 2007.
- [28] M.E. Leventon, W.E.L. Grimson, and O. Faugeras, “Statistical Shape Influence in Geodesic Active Contours,” *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, vol. 1, pp. 316-323, 2000.
- [29] T. Liu and D. Greiger, “Approximate Tree Matching and Shape Similarity,” *Proc. IEEE Int’l Conf. Computer Vision*, pp. 456-462, 1999.
- [30] A. Makadia, A. Patterson, and K. Daniilidis, “Fully Automatic Registration of 3D Point Clouds,” *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, vol. 1, pp. 1297-1304, 2006.
- [31] S. Manay, D. Cremers, B.W. Hong, A. Yezzi, and S. Soatto, “Integral Invariants for Shape Matching,” *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 28, no. 10, pp. 1602-1618, Oct. 2006.

- [32] T. Masuda, K. Sakaue, and N. Yokoya, "Registration and Integration of Multiple Range Images for 3-D Model Construction," *Proc. Int'l Conf. Pattern Recognition*, vol. 1, pp. 879-883, 1996.
- [33] C.R. Maurer, R. Qi, and V. Raghavan, "A Linear Time Algorithm for Computing Exact Euclidean Distance Transforms of Binary Images in Arbitrary Dimensions," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 25, no. 2, pp. 265-270, Feb. 2003.
- [34] N. Paragios, M. Rousson, and V. Ramesh, "Non-Rigid Registration Using Distance Functions," *Computer Vision and Image Understanding*, vol. 89, pp. 142-165, 2003.
- [35] A. Rangarajan, H. Chui, E. Mjolsness, S. Pappu, L. Davachi, P. Goldman-Rakic, and J. Duncan, "A Robust Point Matching Algorithm for Autoradiograph Alignment," *Medical Image Analysis*, vol. 4, pp. 379-398, 1997.
- [36] S. Rusinkiewicz and M. Levoy, "Efficient Variants of the ICP Algorithm," *Proc. Int'l Conf. 3D Digital Imaging and Modeling*, pp. 145-152, 2001.
- [37] R. Sandhu, S. Dambreville, and A. Tannenbaum, "Particle Filtering for Registration of 2D and 3D Point Sets with Stochastic Dynamics," *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, pp. 1-8, 2008.
- [38] F. Schmidt, D. Farin, and D. Cremers, "Fast Matching of Planar Shapes in Sub-Cubic Runtime," *Proc. IEEE Int'l Conf. Computer Vision*, pp. 1-6, 2007.
- [39] T. Sebastian, P. Klein, and B. Kimia, "Recognition of Shapes by Editing Shock Graphs," *Proc. IEEE Int'l Conf. Computer Vision*, vol. 1, pp. 755-782, 2001.
- [40] D. Sharvit, J. Chan, H. Tek, and B. Kimia, "Symmetry-Based Indexing of Image Databases," *Proc. IEEE Workshop Content-Based Access of Image and Video Libraries*, pp. 56-62, 1998.
- [41] Y. Shi and R. Eberhart, "A Modified Particle Swarm Optimizer," *Proc. IEEE World Congress on Computational Intelligence*, pp. 69-73, 1998.
- [42] L. Tang and G. Hamarneh, "SMRFI: Shape Matching via Registration of Vector-Valued Feature Images," *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, pp. 1-8, 2008.
- [43] H. Tek and B.B. Kimia, "Symmetry Maps of Free-Form Curve Segments via Wave Propagation," *Proc. IEEE Int'l Conf. Computer Vision*, pp. 362-369, 1999.
- [44] Y. Tsin and T. Kanade, "A Correlation-Based Approach to Robust Point Set Registration," *Proc. European Conf. Computer Vision*, vol. 3, pp. 558-569, 2004.
- [45] G. Turk and M. Levoy, "Zippered Polygon Meshes from Range Images," *Proc. ACM SIGGRAPH*, pp. 311-318, 1994.
- [46] R.C. Veltkamp and M. Hagedoorn, "State-of-the-Art in Shape Matching," *Principles of Visual Information Retrieval*, pp. 87-119, Springer, 1999.
- [47] C. Wang, D. Duggins, J. Gowdy, J. Kozar, R. MacLachlan, C. Mertz, A. Suppe, and C. Thorpe, "Navlab SLAMMOT Data Sets," [http://www.cs.cmu.edu/~bobwang/data sets.html](http://www.cs.cmu.edu/~bobwang/data%20sets.html), 2010.
- [48] F. Wang, B. Vemuri, A. Rangarajan, and S. Eisenschenk, "Simultaneous Nonrigid Registration of Multiple Point Sets and Atlas Construction," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 30, no. 11, pp. 2011-2022, Nov. 2008.
- [49] F. Wang, B. Vemuri, and A. Rangarajan, "Groupwise Point Pattern Registration Using a Novel CDF-Based Jensen-Shannon Divergence," *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, vol. 1, pp. 1283-1288, 2006.
- [50] Y. Wang, X. Huang, C.S. Lee, S. Zhang, Z. Li, D. Samaras, D.N. Metaxas, A. Elgammal, and P. Huang, "High Resolution Acquisition, Learning and Transfer of Dynamic 3-D Facial Expressions," *Computer Graphics Forum*, vol. 23, pp. 677-686, 2004.
- [51] P. Xiao, N. Barnes, T. Caetano, and P. Lieby, "An MRF and Gaussian Curvature Based Shape Representation for Shape Matching," *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, pp. 1-7, 2007.
- [52] X. Xie, W. Zhang, and Z. Yang, "Adaptive Particle Swarm Optimization on Individual Level," *Proc. Int'l Conf. Signal Processing*, vol. 2, pp. 1215-1218, 2002.
- [53] H. Zhang and E. Fiume, "Shape Matching of 3D Contours Using Normalized Fourier Descriptors," *Proc. IEEE Int'l Conf. Shape Modeling and Applications*, pp. 261-268, 2002.
- [54] Z. Zhang, "Iterative Point Matching for Registration of Free-Form Curves and Surfaces," *Int'l J. Computer Vision*, vol. 13, pp. 119-152, 1994.



**Hongsheng Li** received the bachelor's degree in automation from East China University of Science and Technology in 2006, and the master's degree in computer science from Lehigh University, Bethlehem, Pennsylvania, in 2010. Currently, he is a PhD student in the Department of Computer Science and Engineering at Lehigh University. His research interests include computer vision and medical image analysis, with focus on shape analysis, feature matching, and active contour models. He has published several papers on these topics in CVPR and MICCAI. He is a student member of the IEEE and the IEEE Computer Society.



**Tian Shen** received the BEng and MEng degrees from Xidian University, China, in 2004 and 2007, respectively. Currently, he is working toward the PhD degree in the Department of Computer Science and Engineering at Lehigh University, Bethlehem, Pennsylvania. His research interests are in the fields of image processing and computer vision. He focuses on the problems of shape modeling, curve and surface evolution theory, and deformable models. He is a student member of the IEEE and the IEEE Computer Society.



**Xiaolei Huang** received the bachelor's degree in computer science and engineering from Tsinghua University, China, in 1999, and the master's and doctorate degrees in computer science from Rutgers, The State University at New Brunswick, New Jersey, in 2001 and 2006, respectively. She is currently a P.C. Rossin assistant professor in the Computer Science and Engineering Department at Lehigh University, Bethlehem, Pennsylvania. Her research interests are in the areas of computer vision, biomedical image analysis, and computer graphics. She focuses on developing novel and robust algorithms for segmentation, registration, deformable object modeling, and recognition. She serves on the program committees of several international conferences on computer vision and biomedical image computing, and she is a reviewer for journals including the *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, the *International Journal on Computer Vision (IJCV)*, the *IEEE Transactions on Image Processing (TIP)*, *IEEE Transactions on Biomedical Engineering (TBME)*, and *Computer Vision and Image Understanding (CVIU)*. She is a member of the IEEE and the IEEE Computer Society.

► For more information on this or any other computing topic, please visit our Digital Library at [www.computer.org/publications/dlib](http://www.computer.org/publications/dlib).