Robust Point Set Registration Using Gaussian Mixture Models

Bing Jian, *Member*, *IEEE Computer Society*, and Baba C. Vemuri, *Fellow*, *IEEE*

Abstract—In this paper, we present a unified framework for the rigid and nonrigid point set registration problem in the presence of significant amounts of noise and outliers. The key idea of this registration framework is to represent the input point sets using Gaussian mixture models. Then, the problem of point set registration is reformulated as the problem of aligning two Gaussian mixtures such that a statistical discrepancy measure between the two corresponding mixtures is minimized. We show that the popular iterative closest point (ICP) method [1] and several existing point set registration methods [2], [3], [4], [5], [6], [7] in the field are closely related and can be reinterpreted meaningfully in our general framework. Our instantiation of this general framework is based on the the L2 distance between two Gaussian mixtures, which has the closed-form expression and in turn leads to a computationally efficient registration algorithm. The resulting registration algorithm exhibits inherent statistical robustness, has an intuitive interpretation, and is simple to implement. We also provide theoretical and experimental comparisons with other robust methods for point set registration.

[7]:CPD

感觉D2D NDT很明显借鉴了这个想法

Index Terms—Point set registration, nonrigid registration, Gaussian mixtures, robust matching.

1 Introduction

Point set representations frequently arise in a variety of applications of computer vision, medical image analysis, pattern recognition, and computer graphics. Many problems in these fields can be solved by algorithms operating on the point sets extracted from the input data. In this paper, we focus on the important problem of point set registration, which is often encountered in stereo correspondence, shape matching, feature-based image registration, model-based segmentation, and several other problems.

Mathematically, the problem can be described as follows: Let $\{\mathcal{M}, \mathcal{S}\}$ be two finite size point sets to be registered, where $\mathcal{M}$ is labeled as the moving "model" set and $\mathcal{S}$ the fixed "scene" set. Both $\mathcal{M}$ and $\mathcal{S}$ are assumed to be subsets of a finite-dimensional real vector space $\mathbb{R}^d$ and in general, they can be of different sizes. A common approach to registering these point sets is to estimate the mapping from $\mathbb{R}^d$ to $\mathbb{R}^d$, which yields the best alignment between the transformed "model" set and the target "scene" set. The mapping may represent either a rigid or nonrigid transformation. Note that for a general purpose point set registration algorithm, each input point set is considered as a collection of isolated unstructured points, meaning additional knowledge beyond the spatial information, such as the mesh structure, labels, features, may not be assumed.

如何让点云能够展现出相对结构化的表示?

Developing efficient algorithms for point pattern matching problems has been an active research topic in the computational geometry and pattern recognition communities. There is a large body of research work on point pattern matching from the view point of structural pattern recognition as well. For example, traditional probabilistic relaxation methods employ randomized algorithms along with the neighborhood information to find the correspondence between patterns in the model and target scene sets, respectively. For a more detailed discussion on probabilistic relaxation methods, we refer the reader to [8] and references therein.

One popular approach to the point pattern matching problem involves applying the algorithms for the more general graph matching problem [9]. Representative work reported in [10], [11] involved extracting the structural information of a point set via a neighborhood graph on the points and using inexact relational matching and a spectral graph method. The spectral-based method, however, is known to be vulnerable to perturbations of the underlying structural matrix built from the point set. In more recent work [12], the point pattern matching problem is formulated as a weighted graph matching problem, which is further reformulated as finding a maximum probability configuration in a probabilistic graphical model. However, this approach is currently limited to rigid registrations and the computational complexity is relatively high. In the context of nonrigid 2D shape matching, methods like matching the so-called shape context [13] or preserving the local neighborhood structures [14] have also been reported in literature. The applicability of these methods, however, is limited by the assumption that the corresponding points have similar neighborhood structures.

对于谱方法的评价，这个需要后续去看

这几个好像之前完全没听说过

The *iterative closest point* (ICP) algorithm [1] is one of the most well-known algorithms for point set registration. The idea of the ICP algorithm was motivated by the fact that there exist closed-form solutions for estimating 3D rigid body transformations, given a correspondence of point pairs [15]. Traditional ICP works as follows: For each point m_i in the model set $\mathcal{M}$, it finds its closest point, $\hat{s}_i$, in the scene set $\mathcal{S}$.

- *B. Jian is with Siemens Medical Solutions, 51 Valley Stream Parkway, E51, Malvern, PA 19355. E-mail: bing.jian@siemens.com.*
- *B.C. Vemuri is with the Department of Computer and Information Sciences, E324, University of Florida, Gainesville, FL 32611-6120. E-mail: vemuri@cise.ufl.edu.*

Manuscript received 6 Sept. 2009; revised 11 May 2010; accepted 11 Nov. 2010; published online 13 Dec. 2010.
Recommended for acceptance by C. Stewart.
For information on obtaining reprints of this article, please send e-mail to: tpami@computer.org, and reference IEEECS Log Number TPAMI-2009-09-0592.
Digital Object Identifier no. 10.1109/TPAMI.2010.223.

The rigid transformation T that best aligns the $\{m_i, \hat{s}_i\}$ pairs in a least-squares sense is then calculated using the closed-form solution. Then, all of the points in $\mathcal{M}$ are transformed by T. This establish-correspondence-then-register cycle is iterated until the specified stopping criterion is satisfied. The traditional ICP algorithm is intuitive and simple but has practical limitations due to its assumption that every closest point pair should correspond to each other. This assumption can easily fail when the two point sets are not coarsely aligned or the model set is not a proper subset of the scene due to possible occlusions in the scene. The occlusion problem can be alleviated by some statistical analysis on the distances of corresponding points in order to identify and reject outliers [16], [17], [18] or by trimming point sets through repeated random sampling [19]. However, a good initial transformation that places the two data sets in approximate registration is still required and it is indeed a nontrivial problem for ICP algorithm practitioners to solve.

Instead of assuming a one-to-one correspondence based on the nearest neighbor criterion, one-to-many relaxations have been proposed to allow for fuzzy correspondences. One of the most notable contributions along these lines is the work in [2], wherein a soft assignment technique and the deterministic annealing are combined to alternatively estimate the transformation and update the correspondence. The key idea of their work is to assume that each model point corresponds to a weighted sum of the scene points, instead of the closest scene point alone. These weights are taken from an affinity matrix whose entries are proportional to a Gaussian function of the pairwise distances between the moving model and the fixed scene. In other words, this affinity matrix assigns each model point to all points in the scene. The closer the scene point is to the model point, the larger the assigned weight is. Clearly, the closest point correspondence in ICP can be viewed as a binary version of this graduated assignment scheme. In [2], the registration problem is expressed as an joint optimization over the transformation parameter and correspondence matrix. However, in practice, it is implemented as an alternating update strategy: At each iteration, a transformation is estimated from the virtual correspondence between model points and the linear combinations of scene points, then the correspondence matrix gets modified as the model points move toward the scene and the scale parameter in the Gaussian function for computing the weights decreases. Recently, Myronenko et al. [6] proposed yet another robust nonrigid point set registration algorithm. Like in [2], Myronenko et al. [6] maintain the same Gaussian affinity matrix and also adopt a similar alternating update strategy interpreted in an expectation maximization (EM) [20] framework. The major difference between these two algorithms is that in [2], the nonrigid deformation is modeled by thin-plate splines (TPS), while in [6], they use Gaussian radial basis functions (GRBF). More recently, two interesting point set registration methods in the EM framework have been proposed [21], [22]. However, both methods are currently limited to rigid and articulated registration problems.

There also exist methods that attempt to align the given two point sets without establishing the explicit point correspondence, and thus are less sensitive to the missing correspondences and outliers. One popular approach is based on the concept of distance transform. For example, in [23], the point set registration is achieved by rigidly moving one point set such that a robust function on the distance field precomputed from the fixed point set is minimized. Another trick is to model each of the two point sets by a probability distribution and then a distance measure between the two distributions is minimized over the transformation space to yield the desired transformation. For instance, Tsin and Kanade [4] proposed a kernel correlation (KC)-based point set registration approach where the cost function is proportional to the correlation of two kernel density estimates. Glaunes et al. [24] formulated the problem of aligning two unlabeled point sets by first modeling the point sets as weighted sums of Dirac measures and then finding the optimal diffeomorphic transformation between the two discrete distributions.

In this paper, we present a method that belongs to the aforementioned class of approaches. A preliminary version of this work has been reported in a conference article [5]. The key contributions of our research presented here are: 1) We investigate the idea of using Gaussian mixture models (GMM) as a natural and simple way to represent the given point sets. Then, we treat the problem of point set registration as that of aligning two Gaussian mixtures by minimizing the discrepancy between two Gaussian mixtures. Interestingly, several existing point set registration methods, including [1], [2], [3], [4], [5], [6], [7], can be viewed as special cases in this generic framework. However, we do not claim that we are the first to use the GMM in the point set registration problem as GMM has already been applied to this problem either explicitly [3] or implicitly in the form of kernel density estimates [4]. 2) In order to align mixture densities, we leverage the *closed-form expression* for the L2 distance between two Gaussian mixtures, which in turn leads to a computationally efficient registration algorithm. Our registration algorithm has an intuitive interpretation, is simple to implement, and exhibits inherent statistical robustness. Additionally, we have made the C++, MATLAB, and Python implementations of our algorithm publicly available at http://gmmreg.googlecode.com.

The rest of this paper is organized as follows: We present the main idea of matching point sets using a mixture model in Section 2. A general robust framework involving the minimization of the L2 distance between Gaussian mixtures is developed in Section 3. We also explain the relationship between our method and several existing methods in Section 4. Some implementation details are addressed in Section 5. Section 6 contains several experiments on rigid and nonrigid point set registration using shape, image, range scan, and motion capture data sets. Finally, we present a discussion and conclusion in Section 7.

2 Mixture Models for Point Set Representation

One of the key ideas of our work is to represent discrete point sets by continuous density functions, namely, Gaussian mixture models. The motivation behind this idea is twofold. First, one can interpret the given point set as a statistical sample drawn from a continuous probability

distribution of random point locations. In some sense, this interpretation explicitly reflects the uncertainty of feature extraction, the preprocessing step that produces the point sets to be registered [25]. Second, by doing so, traditionally hard discrete optimization problems usually encountered in the point matching literature can be potentially converted to more tractable continuous optimization problems. Note that the general idea of this density-based registration approach favors the pair of point sets of similar sampling rates. In cases where the two point sets to be registered have quite different sampling rates, for example, point sets from range scans of a slanted surface from different viewing angles, the performance of the density-based registration may degrade and this degradation depends on the robustness of the adopted approach.

The probability density function of a general Gaussian mixture is defined as $p(\mathbf{x}) = \sum_{i=1}^{k} w_i \phi(\mathbf{x}|\mu_i, \Sigma_i)$, where

$$\phi(\mathbf{x}|\mu_i, \Sigma_i) = \frac{\exp\left[-\frac{1}{2}(\mathbf{x}-\mu_i)^T \Sigma_i^{-1}(\mathbf{x}-\mu_i)\right]}{\sqrt{(2\pi)^d |\det(\Sigma_i)|}}. \quad (1)$$

If the number of components is quite large, then almost any density may be well approximated by this model. Assuming no prior information, we choose to explicitly construct the Gaussian mixture model from the given point set in a simplified setting as follows: 1) The number of Gaussian components is the number of the points in the point set and all components are weighted equally, 2) for each component, the mean vector is given by the spatial location of each point, and 3) all components share the same spherical covariance matrix. The resulting density is an overparameterized Gaussian mixture which can be equivalently obtained by applying a fixed-bandwidth kernel density estimate [26], [27] with a Gaussian kernel. There are many references on selecting an appropriate bandwidth; see [26], [27]. Our experience has suggested that the final registration results are similar for most reasonable selections of the bandwidth and, in practice, we leave it as a free parameter tunable by the user of registration algorithm. Note that the original discrete point set can be thought of as the limiting case when the bandwidth approaches zero.

It is true that some additional information can be used to guide the mixture construction. For example, the confidence values, if available, intuitively indicate the weights of components in the resulting mixture. Furthermore, if the intensity values in the neighborhood of feature points can be made available, then for each feature point, one is able to estimate a covariance matrix that measures the spread of its potential position [28], [29], [30], though experimental results in [30] show that in most cases, it suffices to take the spherical covariance matrices other than the covariance matrices computed from image intensity information for image matching. It is also possible to explore the use of variable bandwidth kernel density estimators [31], [32].

When the input point set has a very large number of points, for example, a dense point cloud, a mixture model-based clustering or grouping may be performed as a preprocessing step to simplify the model, for example, by recursively collapsing statistically similar components [33], where the mean vectors and covariance matrices can be estimated as byproducts. The clustering methods that have been widely studied in the literature, however, will not be discussed in this paper. Instead, we leave the choice of clustering methods to the user when clustering is needed.

3 Similarity Measures between Gaussian Mixtures

Based on the Gaussian mixture model representation discussed in the last section, an intuitive reformulation of the point set registration problem is to solve an optimization problem such that a certain dissimilarity measure between the Gaussian mixtures constructed from the transformed model set and the fixed scene set is minimized.

In this paper, we choose the L2 distance for measuring similarity between two Gaussian mixtures, motivated by the following two reasons: *1) The L2 distance is strongly related to the inherently robust estimator L_2E [34] and 2) there exists a closed-form expression for the L2 distance between Gaussian mixtures, which in turn affords an efficient implementation of the registration algorithm.* Formally, given two point sets, the *model* set $\mathcal{M}$ and the *scene* set $\mathcal{S}$, our registration method finds the parameter θ of a parametrized spatial transformation family T which minimizes the following cost function:

$$d_{L_2}(\mathcal{S}, \mathcal{M}, \theta) = \int \left(gmm(\mathcal{S}) - gmm(T(\mathcal{M}, \theta))\right)^2 dx, \quad (2)$$

where $gmm(\mathcal{P})$ refers to the Gaussian mixture density constructed from a point set $\mathcal{P}$. Note that in practice, regularization terms have to be included in the cost function for imposing constraints on the transformations.

The L2 distance function can be treated as a special case of the *density power divergence* [35]:

$$d_\alpha(g, f) = \int \left\{ \frac{1}{\alpha} g^{1+\alpha} - \frac{1+\alpha}{\alpha} g f^\alpha + f^{1+\alpha} \right\} dx, \quad (3)$$

where f and g are two density functions. It is an interesting family of Bregman divergence [36] functions controlled by a single parameter α. The well-known Kullback-Leibler (KL) divergence can be obtained by letting α approach 0: $d_0(g, f) = \lim_{\alpha \to 0} d_\alpha(g, f) = \int g(x) \log\{g(x)/f(x) dx\}$. It is also known that in this case, the minimizer of KL divergence corresponds to maximum likelihood estimation (MLE). On the other hand, when $\alpha = 1$, the divergence $d_1(g, f) = \int \{f(x) - g(x)\}^2 dx$ becomes exactly the L2 distance between the densities and the corresponding estimator is called the L_2E estimator. For general $0 < \alpha < 1$, the class of density power divergences provides a smooth bridge between the KL divergence and the L2 distance. Furthermore, this parameter α controls the trade-off between robustness and asymptotic efficiency of the parameter estimators, which are the minimizers of this family of divergences.

The closed-form expression for the L2 distance between two Gaussian mixtures can be easily derived due to the following formula:

$$\int \phi(\mathbf{x}|\mu_1, \Sigma_1)\phi(\mathbf{x}|\mu_2, \Sigma_2) dx = \phi(0|\mu_1 - \mu_2, \Sigma_1 + \Sigma_2). \quad (4)$$

Note that (3) does not have a closed form for $0 < \alpha < 2$ except for the L2 distance at $\alpha = 1$ for Gaussians and Gaussian mixtures [34]. This fact offers an advantage to the

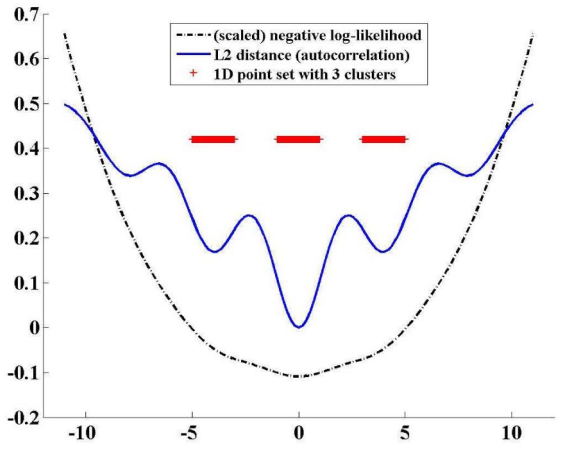


Fig. 1. An example of aligning two 1D point sets by minimizing the L2 distance and maximizing the log-likelihood function versus the translation parameter. The scale parameter used in GMM for both approaches is $\sigma = 1$.

L2 distance since the numerical integration or approximation can be a practical limitation not only in computation time but also for obtaining sufficient accuracy to perform the numerical optimization.

Another rationale for our choice of the L2 distance is its inherent relation to L_2E, a robust estimator minimizing the L2 distance between densities. L_2E has been well studied in statistics literature [34], [35]. The robustness of L_2E can be explained from the view point of M-estimators [37], [38]. It has also found successful applications such as image registration [39], segmentation [40], and blind source separation [41]. It is known that the maximum likelihood is well suited for the estimation problem in which the model is a good descriptor for the data and all of the data are indeed coming from the model. However, the MLE estimates can be badly biased if the model is not good enough or there is a small fraction of outliers. Here, we present two 1D toy examples to demonstrate different behaviors of the L2 distance between Gaussian mixtures representing two point sets and the log-likelihood function of one point set as data with respect to another as the model. In our first example, both of the point sets are clustered into three separated regions: $[-5, -3], [-1, 1], [3, 5]$, all uniformly sampled with the same sampling rate. As shown in Fig. 1, the log-likelihood function has only one optimum when the two point sets are perfectly aligned. The L2 distance, however, has one global minimum corresponding to the perfect alignment and also has a few local minima corresponding to the partial alignment due to the repeating pattern in the data. The existence of the local minima is appropriate since, in this setting, minimizing L2 distance is essentially equivalent to maximizing the autocorrelation function, which is known as a good mathematical tool for detecting periodic patterns in the data.

The second example, shown in Fig. 2, is a more realistic case where one point set has high density in segment $[-1, 1]$ and much lower density in $[4, 5]$, while another point set also has high density in segment $[-1, 1]$ but much lower density in $[-5, -4]$, as shown in the figure. In this scenario, it is reasonable to consider the lower density regions as outliers. Figs. 2b, 2c, 2d, 2e, 2f, 2g, and 2h show how the log-likelihood functions and the L2 distances at different scales behave when translating the first point set with respect to the second one. Not surprisingly, at a sufficiently large scale, both the log-likelihood and distance profiles have one optimum which corresponds to the shift between the centroids of the two point sets, indicating a center of mass alignment. It is interesting to observe that at small to intermediate scales, the L2 distance function is consistent in reaching the global minimum at the expected location, while this desirable property is not seen from the log-likelihood function which also has multiple local optima.

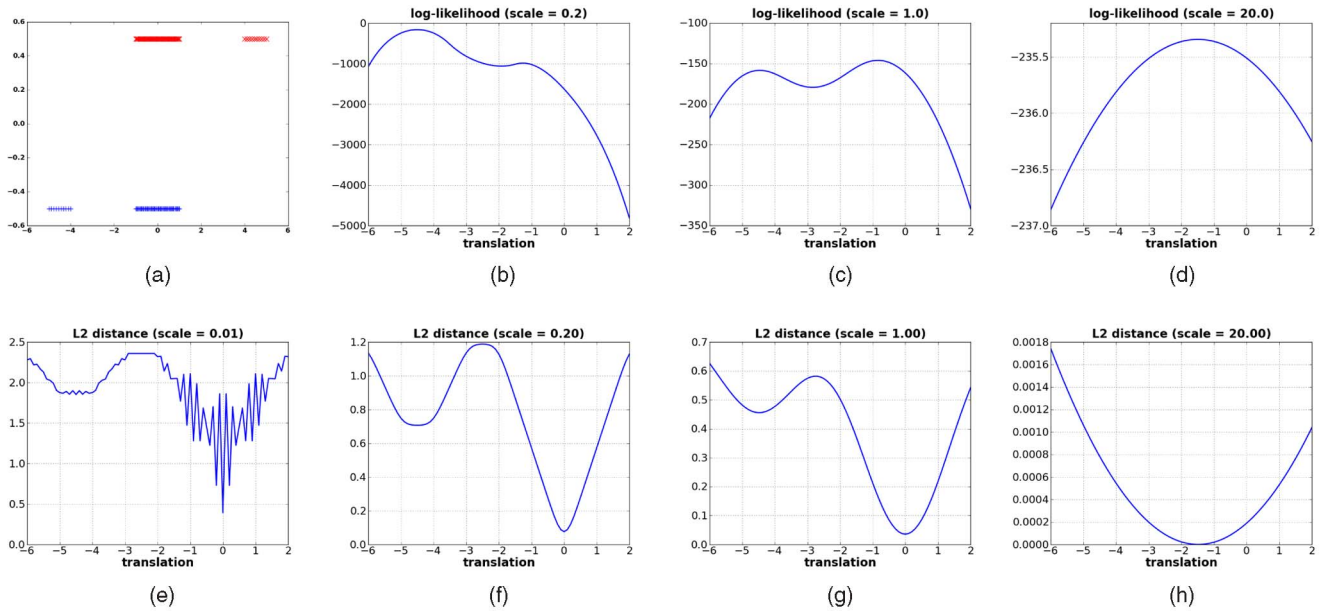


Fig. 2. (a) Two 1D point sets, each of which is clustered into two separate regions of different sampling rates. (b)-(h) The log-likelihood functions and the L2 distances at different scales versus translation of one point set.

这个图是什么意思？？

4 Related Work

这个地方挺有意思的...

In [5], we made an interesting observation that the traditional ICP method can be interpreted as minimizing the KL divergence between two Gaussian mixtures using the approximation formula suggested in [42]:

$$KL(f\|g) \approx \sum_{i=1}^{n} \alpha_i \min_j \left(KL(f_i\|g_j) + \log \frac{\alpha_i}{\beta_j} \right), \tag{5}$$

where $f = \sum_i \alpha_i f_i$ and $g = \sum_j \beta_j g_j$ are two Gaussian mixtures. Note that the KL divergence between two Gaussians $\phi(\mu_1, \Sigma_1)$ and $\phi(\mu_2, \Sigma_2)$ is

$$\frac{1}{2}\left(\log \frac{|\Sigma_2|}{|\Sigma_1|} + Tr(\Sigma_2^{-1}\Sigma_1) + (\mu_1 - \mu_2)^T \Sigma_2^{-1}(\mu_1 - \mu_2) \right). \tag{6}$$

In the simplified version of a Gaussian mixture representation of a given point set, each component is assumed to be the same spherical Gaussian centered at a point location with the same covariance matrix and weight, and the term $\min_j(KL(f_i\|g_j)$ corresponds to the minimum euclidean distance from the ith point in the set modeled by the mixture f to the point set modeled by mixture g. From this, one can easily see that the idea of minimizing the approximated KL divergence between two Gaussian mixtures bears much resemblance to the popular ICP registration method [1].

Assume that the fixed scene $\mathcal{S}$ is taken as a sample drawn from a Gaussian mixture p and let q be the mixture distribution representing the moving model set.[1] Since $KL(p\|q) = \int(p(x)\log p(x) - p(x)\log q(x))dx$ and p is fixed, minimizing the $KL(p\|q)$ is equivalent to maximizing the log-likelihood of $\mathcal{S}$ with respect to q by transforming $\mathcal{M}$ in an asymptotic sense. The idea of formulating the point set registration problem as a maximum likelihood estimation problem has been proposed in [3], wherein they choose one sparsely distributed point set as the template density modeled by a Gaussian mixture and treat another relatively dense point set as sample data. Note that the asymmetry here is due to the distinct role of the data and the model. Similarly to [2], [3], Myronenko et al. [6] also interpret the point set registration as maximizing the log-likelihood of scene with respect to the moving model, which is clearly seen from the energy function being minimized in [6]:

$$-\sum_{s\in\mathcal{S}} \log \sum_{m\in\mathcal{M}} \exp\left(-\frac{1}{2}\| \frac{s-m}{\sigma} \|^2 \right). \tag{7}$$

Since the MLE is sensitive to outliers, all of these algorithms [2], [3], [6] include an additional Gaussian component to account for outliers, which is not needed in our method.

The similarity measure used in [4] is the so-called *kernel correlation* between the moving model and the fixed scene: $\sum_{s\in\mathcal{S}}\sum_{m\in\mathcal{M}} KC(s, T(m))$, where the kernel correlation between two points $\mathbf{x}_i$ and $\mathbf{x}_j$ is defined as follows:

$$KC(\mathbf{x}_i, \mathbf{x}_j) = \int K(\mathbf{x}, \mathbf{x}_i) K(\mathbf{x}, \mathbf{x}_j) dx. \tag{8}$$

Clearly, when the kernel function $K(\mathbf{x}, \mathbf{x}_i)$ is chosen as a Gaussian, the sum of all pairwise kernel correlations between two point sets is proportional to the $\int fgdx$ term in L2 distance, derived from (4). For the nonrigid transformation, a normalization term is added to the cost function in [4], leading to the following similarity measure:

$$\mathrm{cor}(f, g) = \frac{\int fgdx}{\sqrt{\int f^2 dx \int g^2 dx}}, \tag{9}$$

which is commonly considered as a "correlation" between densities [33], [43] and also has closed-form expression for Gaussian mixtures. This correlation measure is also used in a recent point set registration algorithm [44] where only rigid registration results are reported as is the case for [4].

It is easy to see that when g is a fixed density, finding f to maximize (9) is equivalent to minimizing the following form:

$$\int g^2 dx - \frac{(\int fgdx)^2}{\int f^2 dx}, \tag{10}$$

加了个负号

which we note as a special case of a divergence family defined as

$$\frac{1}{\beta}\left(\int g^{1+\beta} dx - \frac{(\int f^\beta g dx)^{1+\beta}}{(\int f^{1+\beta} dx)^\beta} \right). \tag{11}$$

As discussed in [45], the corresponding minimum divergence estimator of this divergence family (11) is equivalent to a robust model fitting technique proposed by Windham [46]. A very interesting observation made in [45] is that the density power divergence (3) proposed in [35] and the divergence family (11) are closely related, yet different in general, and are both special cases of a larger family of divergences given by

$$\phi^{-1}\left(\int f^{1+\gamma} dx \right)^\phi - \frac{1+\gamma}{\gamma}\phi^{-1}\left(\int f^\gamma g dx \right)^\phi + \frac{1}{\gamma}\phi^{-1}\left(\int g^{1+\gamma} dx \right)^\phi. \tag{12}$$

Note that (12) gives the density power divergence (3) with $\phi = 1$ and $\gamma = \alpha$, while the divergence family (11) can be obtained by taking the limit $\phi \to 0$ and letting $\gamma = \beta$. It has been further shown in [45] that the cases $\phi = 0$ and $\phi = 1$ are actually statistically the most interesting ones due to the following arguments: 1) Both contain maximum likelihood as a limiting case ($\alpha \to 0$ in the density power divergence or $\beta \to 0$ in (11)); 2) for f close to g, both are close to $\frac{1}{2}(1+\gamma)\int g^{\gamma-1}(f-g)^2 dx$ for fixed $\gamma = \alpha = \beta$ if the integral is finite, which is not true for other choices of ϕ; and 3) the estimator associated with (12) is exactly unbiased and can be shown to be an M-estimator [37], [38], only for $\phi = 0$ or 1. A comparison of these two estimator classes is also made in [45]. The conclusion is that overall, these two classes appear to perform rather similarly, while some relatively small advantages of minimizing the density power divergence over the other are identified, especially in terms of asymptotic efficiency and the breakdown point. We summarize the relationship between these divergence functions in Fig. 3. For more details, see [45].

1. Note that in general, p is taken as a fixed empirical distribution of $\mathcal{S}$, not necessarily a Gaussian mixture.

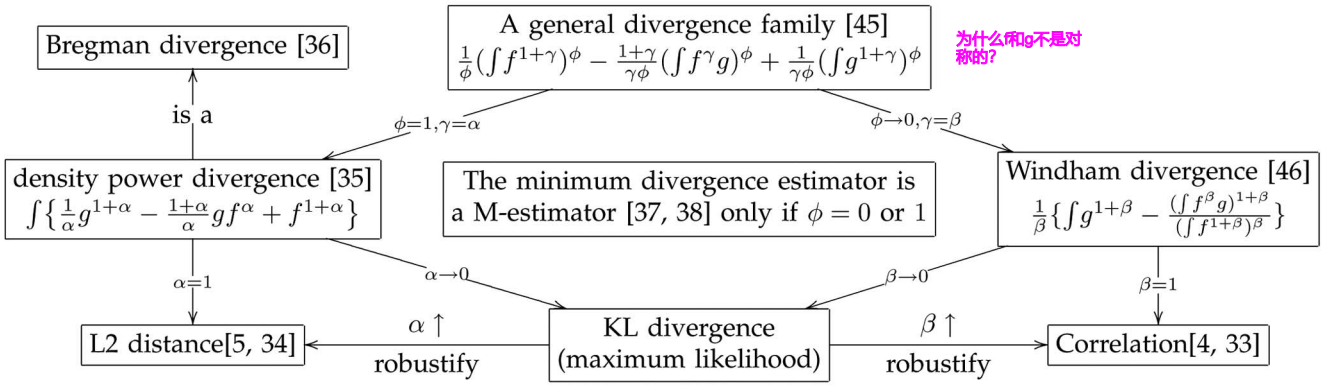


Fig. 3. The relationship between various divergence functions discussed in the text.

5 Implementation Details

Let $f = gmm(T(\mathcal{M}, \theta))$ and $g = gmm(\mathcal{S})$ be the two Gaussian mixtures representing the moving model set $\mathcal{M}$ under a transformation family $T(,\theta)$ and the fixed scene set $\mathcal{S}$, respectively, the L2 distance (2) is simply the sum of three terms: $\int f^2dx - 2\int fgdx + \int g^2dx$. Since g is fixed during the optimization, the evaluation of the cost function only involves the first two terms, namely, the L2 integral of the moving Gaussian mixture, which is the inner product between the moving Gaussian mixture and itself, and the inner product between the moving Gaussian mixture and the fixed Gaussian mixture. Furthermore, the first term is invariant under rigid transformations, which implies that maximizing the kernel correlation [4] and minimizing the L2 distance [5] are essentially the same for the rigid point set registration.

gmm如何获得?

The closed form of the inner product between two d-dimensional Gaussian mixtures f and g with m and n components, respectively, is just the sum of n Gaussian functions evaluated at m points in d-dimensional space. In the case where all of the components are spherical Gaussians with a uniform scale, our cost function can be expressed in the form of the so-called "discrete Gauss transform." Apparently, the cost of direct evaluations of discrete Gauss transforms grows as $O(mn)$. There exist fast approximation schemes, such as the fast Gauss transform method [47] and its improved version [48], which require only $O(n+m)$ work with a constant factor depending on the dimension d and desired precision. However, consistent with the results shown in a recent work on improved fast Gauss Transform [48], our observations show that when 2D or 3D point sets with both m and n are less than 1,000, direct evaluation is sufficiently fast to yield reasonably good results, while the gain observed in using the fast approximation schemes is insignificant. For example, a direct evaluation of Gauss Transform with $m = n = 800, d = 3$, as well as the gradient, only takes 30 ms on a 2 GHz computer. Actually, for rigid or affine transformation models with relatively few parameters, we found that a nonstochastic gradient-free optimization method, for example, Powell's method, is sufficiently fast to achieve quite accurate results in most cases. However, for the nonrigid registration with thousands of points or more, a gradient-based numerical optimization algorithm might be necessary and it is worth applying fast approximation schemes to accelerate the evaluation of Gauss transforms in computing both the cost function and the gradient.

Let $\mathbf{M}_0 = (\mathbf{x}_1, \ldots, \mathbf{x}_m)^T$ be the $m \times d$ matrix denoting the input model point set $\mathcal{M}$ and $\mathbf{M}$ be the $m \times d$ matrix denoting the moving model point set $T(\mathcal{M}, \theta)$. The gradient of the cost function $F(\mathbf{M}(\theta))$ with respect to the motion parameter θ can be explicitly evaluated using the chain rule: $\frac{\partial F}{\partial \theta} = \frac{\partial F}{\partial \mathbf{M}} \cdot \frac{\partial \mathbf{M}}{\partial \theta}$. For convenience, in the following we represent $\mathbf{G} = \frac{\partial F}{\partial \mathbf{M}}$ by an $m \times d$ matrix. Note that $\mathbf{G}$ can be obtained as a byproduct when evaluating the cost function $F(\mathbf{M})$ using Gauss transforms. Under the rigid transformation model, the moving model point set M and the original model point set $\mathbf{M}_0$ are related by the rigid motion $\mathbf{M} = \mathbf{M}_0\mathbf{R}^T + \mathbf{t}$, where $\mathbf{R}$ is the rotation matrix and $\mathbf{t}$ is the translation vector. Let $\mathbf{1}_n$ be the n-dimensional column vector of all ones, then the derivative of the cost function with respect to the motion parameter is given by

$$\begin{aligned} \frac{\partial F}{\partial \mathbf{t}} &= \mathbf{G}^T\mathbf{1}_m \\ \frac{\partial F}{\partial r_i} &= \mathbf{1}_d^T\left((\mathbf{G}^T\mathbf{M}_0) \otimes \left(\frac{\partial \mathbf{R}}{\partial r_i}\right)\right)\mathbf{1}_d, \end{aligned} \tag{13}$$

这个公式是如何推的?

where $(r_1, \ldots, r_i, \ldots)$ is the parametrization of the rotation matrix $\mathbf{R}$ and $\otimes$ denotes the element-wise multiplication. In our implementation, 3D rotations are parameterized by unit quaternions.

In our implementation of nonrigid point set registration, we investigate two commonly used nonrigid transformation models, namely, the thin-plate splines [49], [50], [51] and the Gaussian radial basis functions [52]. Note that other parameterized transform models, for example, B-splines, can also be easily incorporated into our registration framework. In the following, we take thin-plate splines as an example to show the computation of analytical derivatives.

Let $\mathbf{Q} = (\mathbf{q}_1, \ldots, \mathbf{q}_c)^T$ be a $c \times d$ matrix representing a set of p control points; one can compute the kernel matrix $\mathbf{K} = \{K_{ij}\}$, where $K_{ij} = K(|\mathbf{q}_i - \mathbf{q}_j|)$ and $|\mathbf{q}_i - \mathbf{q}_j|$ is the euclidean distance between two points. The radial basis function used in TPS is defined as $K(r) = r^2 \log r$ for 2D and $K(r) = -r$ for 3D. Note that this kernel matrix $\mathbf{K}$ describes the internal structure of the control point set.

The TPS transformation can be decomposed into a linear part modeled by an affine motion and a nonlinear part governed by TPS warping coefficients, $\mathbf{w}$, which are usually represented by a $c \times d$ matrix. The bending energy of the nonlinear transformation is given by $trace(\mathbf{w}^T\mathbf{K}\mathbf{w})$. Note

that $\mathbf{K}$ is conditionally positive definite [50]; in order to ensure $\mathbf{w}^T\mathbf{K}\mathbf{w} > 0$ and the nonlinear part of the deformation is zero at infinity, the boundary condition $\mathbf{w}^T[\mathbf{1}|\mathbf{Q}] = 0$ has to be satisfied. Therefore, a common approach to enforce this boundary condition is to introduce a new parameter $\mathbf{v}$ as a $(c-d-1)\times d$ matrix and let $\mathbf{w} = \mathbf{N}\mathbf{v}$, where $\mathbf{N}$ is the matrix representing the left null space of $[\mathbf{1}|\mathbf{Q}]$. Under this transformation, the moving model point set is related to the source model point set $\mathbf{M}_0$ by

$$\mathbf{M} - \mathbf{M}_0 = [\mathbf{1}|\mathbf{M}_0]\mathbf{A}^T + \mathbf{U}\mathbf{w} = [\mathbf{1}|\mathbf{M}_0]\mathbf{A}^T + \mathbf{U}\mathbf{N}\mathbf{v}, \quad (14)$$

where $\mathbf{A}$ is the affine part of TPS and the basis matrix is computed as $\mathbf{U} = \{U_{ij}\} = \{K(|\mathbf{x}_i - \mathbf{q}_j|)\}$.

To regularize the TPS nonrigid transform, a penalty term $\frac{\lambda}{2}trace(\mathbf{w}^T\mathbf{K}\mathbf{w})$ is usually added to the final cost function, where $\lambda > 0$ controls the strength of regularization and a very large value of λ yields a nearly pure affine transformation. Hence, the derivatives of the final cost function F with respect to the affine and TPS parameters can be obtained as

$$\begin{aligned} \frac{\partial F}{\partial \mathbf{A}} &= [\mathbf{1}|\mathbf{M}_0]^T\mathbf{G} \\ \frac{\partial F}{\partial \mathbf{v}} &= (\mathbf{U}\mathbf{N})^T\mathbf{G} + \lambda\mathbf{N}^T\mathbf{K}\mathbf{N}\mathbf{v}. \end{aligned} \quad (15)$$

Note that for nonrigid deformations modeled by the Gaussian radial basis functions, the analytical derivatives can also be derived in a similar manner.

Observing that our cost function defined above is convex in the neighborhood of the optimal position and always differentiable, one can employ efficient gradient-based numerical optimization techniques such as the quasi-Newton method and the nonlinear conjugate gradient method to solve the optimization problem. However, as shown in Fig. 2, there are still local minima which may trap the numerical optimization. This local minimum problem usually can be observed when the global motion is large, for example, a flip of a shape, or there exist symmetries or repeated patterns in the spatial structure, or the transformation model has high degrees of freedom, for example, very localized nonrigid deformation.

To overcome the local minimum problem, some heuristic methods can be used, together with an efficient numerical optimization algorithm. Since the cost function tends to be smoother with a larger scale (σ) than with a smaller scale, we recommend starting first with a relatively large scale and then performing the numerical optimization using a default initial setting of transformation parameters. At the local minimum point, we compute the number of correspondences by counting the nearest pairs between the transformed model set and the scene set. If the number of correspondences is less than a threshold value, multiple randomly chosen initializations are supplied to the optimization until a sufficient number of correspondences are obtained. If needed, a multiresolution approach can be adapted in a coarse to fine manner.

In practice, we have found that by using multiple scales, multiple starts, and maximizing correspondence, in most cases we are able to escape from the trap of local minima and reach the global minimum in a few cycles. Our algorithm is outlined in Algorithm 1.

Algorithm 1. A robust algorithm for point set registration using Gaussian mixtures

Input: The *model* set $\mathcal{M}$, the *scene* set $\mathcal{S}$, and a parametrized transformation model T.

Output: The optimal transformation parameter θ of model T that best aligns $\mathcal{M}$ and $\mathcal{S}$.

begin
  Estimate an initial scale σ from input point sets;
  Specify an initial parameter θ, e.g., from the identity transform;
  repeat 如何构造GMM?
    Set up the objective function $f(\theta)$ as the L2 distance between the Gaussian mixtures constructed from the transformed model $T(\mathcal{M}, \theta)$ and the scene $\mathcal{S}$ with a scale σ. A regularization term can be added depending on the transformation model;
    Optimize the objective function f using a numerical optimization engine (e.g., quasi-Newton algorithm when ∇f is available) with θ as the initial parameter;
    Update the parameter $\theta \leftarrow \arg\min_T f$;
    Decrease the scale σ accordingly as an annealing step;
    untill *some stopping criterion is satisfied*
end

The computational cost of this algorithm for nonrigid registration largely depends on the number of control points, which drives the numerical optimization. With the same amount of regularization strength, a dense spacing of control points allows more localized and flexible nonrigid deformations than a sparse spacing of control points. Hence, the positioning of control points is mainly an empirical choice and can be tuned according to the trade-off on accuracy versus computing time. Again, one may also implement a hierarchical approach in which the density of the control points is increased in a coarse to fine fashion.

6 Experiments

In this section, we present experimental results on the application of our registration algorithm to point sets obtained/extracted from various sources including 2D shapes, 3D laser range scans, 3D surfaces, multiview images, and motion capture data sets. Quantitative comparisons with competing registration methods are also presented. All of our experiments were performed on a PC with 2 GB of RAM and a 2.0 GHz Intel CPU.

6.1 Rigid Registration

First, we present the rigid registration experiments on several 2D shape data sets and 3D range scan data sets. On the 2D data sets, we provide comparisons with the CPD rigid registration [7]. On the 3D data sets, we compare with both the CPD and LM-ICP algorithm [23].

6.1.1 2D Rigid Registration

In the first experiment, we compare the convergence range of the proposed registration algorithm against the CPD rigid registration [7]. Fig. 4 shows four different 2D shape data sets used in this experiment. For each data set, we rotated the

采用的数值优化可能收敛到一个局部最优，可以通过指定一些随机位置去寻找一个相对的全局最优

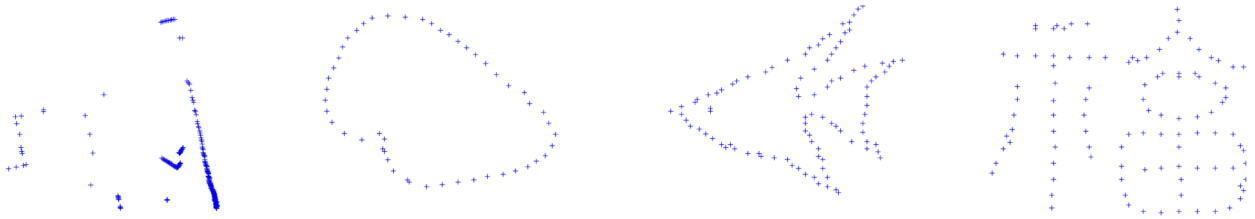


Fig. 4. Four 2D data sets used in the rigid registration experiment. The road data set was obtained from http://www.cs.cmu.edu/ytsin/KCReg/ [4] and the other data sets were obtained from http://www.cise.ufl.edu/anand/students/chui/research.html [2].

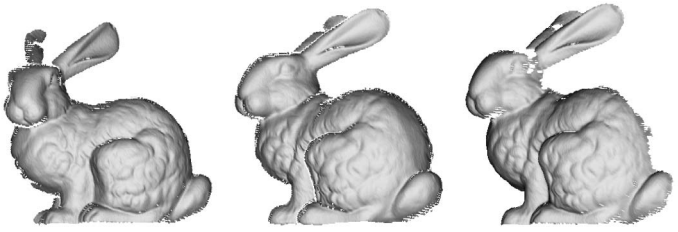


Fig. 5. The Stanford "bun000" model and "bun045" model are shown in the left and middle plates, respectively. On the right is the transformed "bun000" using the 3D rigid motion obtained from an application of our registration algorithm.

point set from $-\pi$ to π and ran both algorithms for aligning the original point set toward the rotated one. The range of angles for each algorithm to perform a successful alignment on each data set is reported in Table 1. The comparison shows that the proposed registration method has a wider convergence range than the CPD registration method.

6.1.2 Experiments with Range Data

We now report results of experimentation with two real-world 3D range scan data sets, the "bunny" and the "dragon," from the Stanford 3D scanning repository.[2] We first took two bunny models, "bun000" and "bun045," and downsampled both models from the original more than 40,000 points to 4,000-5,000 points. Then, we applied the proposed 3D rigid registration algorithm, the CPD rigid registration algorithm, and the LM-ICP algorithm to the downsampled point sets in order to align the "bun000" model to the "bun045" model. All three of these registration algorithms are able to achieve the alignment successfully in less than 80 s. The registration result using the proposed algorithm is shown in Fig. 5.

To test the performance of these algorithms with respect to the outliers in the point sets, we contaminated the model point set by adding random outlier points uniformly sampled inside the bounding box of the model point set and then applied the registration algorithms on the contaminated model. The observation is that all three of these algorithms perform quite accurately even when the percentage of random points is increased to 50 percent of the total number of points in the contaminated model.

In the experiments on the Stanford bunny data set, we found that all three of these algorithms failed to align those scans with large pose differences, for example, the pair of "bun000" and "bun090." In order to evaluate the performance of these algorithms with respect to the pose differences, we conducted experiments on the Stanford "dragon stand" data set, which contains 15 scans with in-plane rotation angles from 0 to 336 degrees evenly spaced by 24 degrees. We ran the registration algorithms on all 30 pairs with ± 24 degrees rotation and compared the registration results with the ground truth. Since, in this data set, the translation is insignificant, we only consider the accuracy of the 3D rotation part, which is evaluated by first representing rotations in unit quaternions and then taking the absolute value of the dot product between the two unit quaternions. The closer this value is to 1, the more accurate the 3D registration is. In our experiments, we say that a 3D registration is successful if this value is greater than 0.99 and count the numbers of successful registrations obtained by each of the three rigid registration algorithms. We also did the same experiments for the pairs with pose difference being ± 48, ± 72, and ± 96 degrees, respectively. The results on the success rates of the three algorithms on these pairs are summarized in Table 2. According to the results shown in Table 2, all of these algorithms are able to achieve good performance when the pose difference between the 3D scans is moderate; however, none of these algorithms is reliable for aligning 3D scans that exhibit a poor initialization.

6.2 Nonrigid Registration

To present a comparison of different nonrigid registration methods under a unified framework, we implemented three different cost functions, namely, minimizing the L2 distance, maximizing the kernel correlation, and maximizing the log-likelihood function, combined with two different nonrigid deformation models, namely, the thin-plate splines and the Gaussian radial basis functions.

TABLE 1
Convergence Range of the Proposed Registration Method and the CPD Registration Method [7] in Terms of 2D Rotation Angles (in Radians) on Four Data Sets Shown in Fig. 4

平面内旋转

dataset	Proposed method	CPD[7]
road	[-3,3]	[-2.02, 0.95]
contour	[-1.8, 1.8]	[-1.53, 1.53]
fish	[-2.0, 2.0]	[-1.25, 1.24]
chinese character	[-2.2, 2.2]	[-1.51, 1.54]

2. http://graphics.stanford.edu/data/3Dscanrep/.

TABLE 2
Success Rates of the Proposed Registration Algorithm, the CPD Algorithm, and the LM-ICP Algorithm on the Stanford "Dragon Stand" Data Set

Pose difference	Proposed method	CPD[7]	LM-ICP[23]
±24° [0.978]	29/30	26/30	28/30
±48° [0.913]	20/30	18/30	19/30
±72° [0.809]	13/30	14/30	13/30
±96° [0.669]	2/30	3/30	1/30

The numbers in the brackets are the accuracy scores of the initial alignments.

For minimizing the L2 distance and maximizing the kernel correlation, since both the cost functions and the gradients are available in closed-form expressions, we used the numerical optimization method L-BFGS-B, an efficient quasi-Newton method for large-scale bound-constrained or unconstrained optimization problems [53]. For maximizing the log-likelihood, the EM [20] method was used. It is interesting to note that the combination of EM and TPS is used in the TPS-RPM algorithm by Chui and Rangarajan [2], while the combination of EM and GRBF is used in the CPD algorithm by Myronenko et al. [6]. We also emphasize that Tsin and Kanade [4] only mentioned the possibility of extending the KC registration method to the nonrigid setting but did not present any experiments on this extension. In all of these approaches, we also allow the user to specify their own configuration of control points, which can significantly reduce the number of parameters and hence the execution time.

In the following, we first show several qualitative experiments on point sets extracted from 2D/3D shape data and 2D image data, then we present quantitative comparisons of different nonrigid point set registration methods on three data sets including 2D shape data, 2D image data, and 3D motion capture data.

6.2.1 Qualitative Experiments

Fig. 6 shows a 2D fish shape data and a 3D face surface data used in our experiments. For both data sets, if the input is free from outliers and occlusion, all of our implementations are able to produce an almost perfect alignment in less than 0.3 s for the 2D fish and in less than 4 s for the 3D face data on a 2 GHz computer. In each case, all implementations run with the same starting scale and configuration of control points. Overall, our perspective on these algorithms being investigated in this work is that they are efficient for most moderate nonrigid point set registration problems and can quickly provide a good initial alignment for more complicated problem-specific registration algorithms.

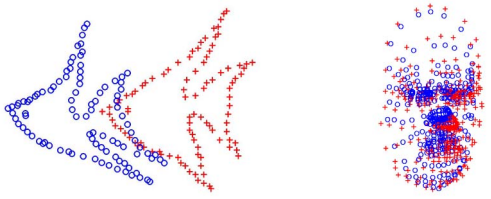


Fig. 6. Two data sets used in the nonrigid shape matching experiments. The 2D fish data were obtained from http://www.cise.ufl.edu/anand/students/chui/research.html [2] and each shape contains 98 points. The 3D face model with 392 points was taken from http://www.csee.ogi.edu/myron/matlab/cpd/ [6].

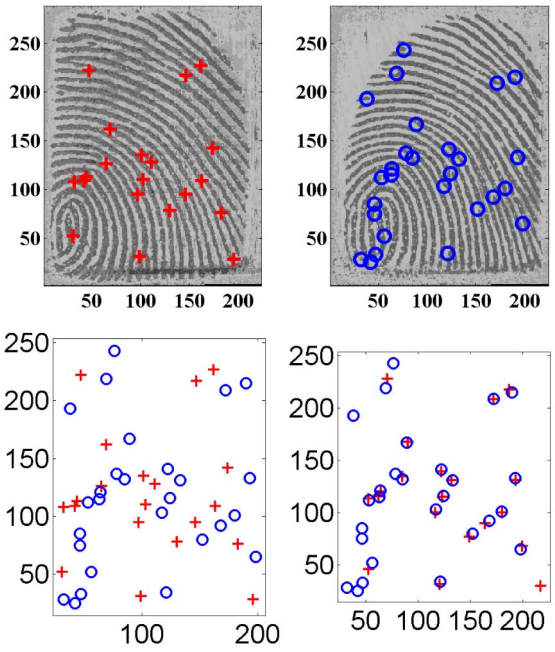


Fig. 7. Experiment on a pair of fingerprint images acquired using Infineon's FingerTip sensor. Top: Two different impressions from the same fingerprint with extracted minutiae superimposed on the images. Bottom: The spatial locations of minutiae before and after the nonrigid registration using our method. Nineteen corresponding minutiae pairs are found in 50 ms.

The top row of Fig. 7 shows a pair of fingerprint images from two impressions of the left index finger of the first author with automatically extracted minutiae superimposed. The elastic deformation between these two images and the corresponding minutiae is clearly seen from the figure. In the bottom row of Fig. 7, the minutiae extracted from the left image are successfully registered to the minutiae extracted from the right image, with 19 matches found in 50 iterations. The matching stage is quite fast and takes about 50 ms on a 2 GHz machine. Since the goal of the point set registration in the fingerprint matching problem is to establish the pairing of minutiae, it is sufficient to bring the corresponding minutiae roughly close instead of into perfect alignment. It is also important to note that our matching algorithm here is a general purpose algorithm and only requires the location information of fingerprint minutiae, while most fingerprint matching algorithms in real working systems usually take additional application-related information such as minutiae type (endpoint or bifurcation point), ridge orientation, and neighborhood structure to accelerate the matching process.

The second experiment is to show the application of our point set registration in a typical multiview computer vision problem. We took two images (frame 1 and frame 4) from the *Oxford corridor sequence*.[3] The corners extracted using the Harris detector [54] are superimposed on these two images, as shown in the top row of Fig. 8. We then ran our nonrigid registration algorithm on the detected corner sets and it took less than 10 s to produce the results of alignment shown in the middle row of Fig. 8. The alignment given by

3. http://www.robots.ox.ac.uk/~vgg/data/data-mview.html.

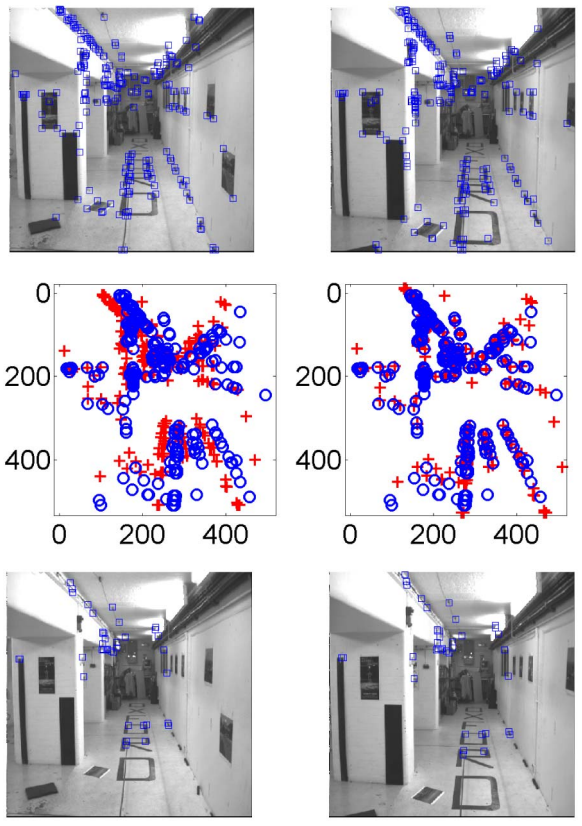


Fig. 8. Top: The two images from the *Oxford corridor sequence*. On the left is frame 1 and on the right is frame 4. The images are of size 512×512 pixels. There are 222 corners and 217 corners detected and superimposed on the two images, respectively. Middle Left: The corner point patterns extracted from frame 1 (red "+") and from frame 4 (blue "o") using the Harris detector. Middle Right: The alignment produced by the nonrigid point set registration where the spatial transform is applied on the "+" points. Bottom: 38 putative point correspondences between frame 1 and frame 4 in the *Oxford corridor sequence* are superimposed on the images. These correspondences are simply found based on the alignment produced by the nonrigid point set registration on the corner sets.

our point set registration can then be used to establish the putative point correspondences or line correspondences in order to estimate the camera motion between these two images. For example, the bottom row of Fig. 8 shows a number of point correspondences found by a simple nearest neighbor searching and distance thresholding. Note that the nonrigid deformation model (either the TPS or GRBF) is employed here only as a means to establish initial point matches but not to really model the underlying geometric transformation between the two images. By using a robust estimation algorithm, for instance, RANSAC [55], the putative point matches provided by the point set registration algorithm should be sufficient for estimating the fundamental matrix [56], even in the presence of mismatches (e.g., the corners attached on "R").

6.2.2 Quantitative Evaluation

We present three quantitative comparisons on the nonrigid point set registration algorithms. The first experiment was conducted on the 2D fish shape data in Fig. 6. To test the ability of these algorithms to handle outliers and occlusion, we delete increasing numbers of points from the head part of one fish and points from the tail part of another. Then, we apply the nonrigid registration algorithms on these pairs of incomplete fish shapes. To have a quantitative evaluation, we compute the *recall* as the metric used in [57]. The recall, or true-positive rate, is defined as the proportion of true-positive correspondences to the ground truth correspondences and a true-positive correspondence is counted when the pair falls within a given accuracy threshold in terms of pairwise distance. Fig. 9 plots the recall curves of different methods after 5, 10, 16, and 20 points were removed from one head and from another tail, respectively. We observed that when the number of deleted points is larger than 20, none of these algorithms is able to achieve a successful alignment. It is evident from Fig. 9 that, overall, the approach by minimizing the L2 distance and using the TPS nonrigid deformation achieves the best performance in this experiment. It also indicates that, in general, the approaches that minimize the L2 distance or those that maximize the kernel correlation outperform the approaches of maximizing the likelihood using EM in terms of robustness. Fig. 10 shows the results of three algorithms on the configuration with 20 points removed from one head and 20 points removed from another tail.

The second quantitative comparison was conducted on the *CMU house sequence*[4] which contains 111 images of a toy house captured from moving viewpoints. In each of the images, 30 landmark points were manually marked with known correspondences.[5] Two images in this sequence with landmarks superimposed are shown in Fig. 11. We selected all image pairs spaced by 70, 80, 90, and 100 frames and ran the different nonrigid registration algorithms on landmarks associated with these image pairs. Fig. 12 shows the results of these runs. The average value is taken over image pairs with the same spacing in the frame sequence. Since there are 111 frames, note that the number of image pairs spaced by 70, 80, 90, and 100 frames are, respectively, 41, 31, 21, and 11. The accuracy thresholds are equally spaced between 0 and 10 pixels. For reference, all the images in this sequence are of size 384×576 in pixels.

The third quantitative evaluation was conducted on a sample "swagger" motion capture data from the Advanced Computing Center for the Arts and Design at the Ohio State University.[6] This data set consists of 518 frames of a "swagger" motion and there are 42 markers whose XYZ positions were recorded in these 518 frames. Note that in most frames, there may be a few markers missing in the trajectory. Similarly to the previous experiment, we selected all time points in the motion sequence spaced by 150 and 250 frames, then we ran the nonrigid registration algorithms on corresponding markers. The performances of these algorithms in terms of recall-accuracy curves are shown in Fig. 13.

From the recall-accuracy curves shown in Figs. 12 and 13, we can see that the performances of approaches that minimize the L2 distance and those that maximize the kernel correlation are almost indistinguishable under the same deformation model and both are consistently better than those of the EM-based approaches.

4. http://vasc.ri.cmu.edu/idb/html/motion/house/index.html.
5. The authors thank Dr. Tibério S. Caetano for providing the landmarks with correspondences that were used in [12].
6. http://accad.osu.edu/research/mocap/mocap_data.htm.

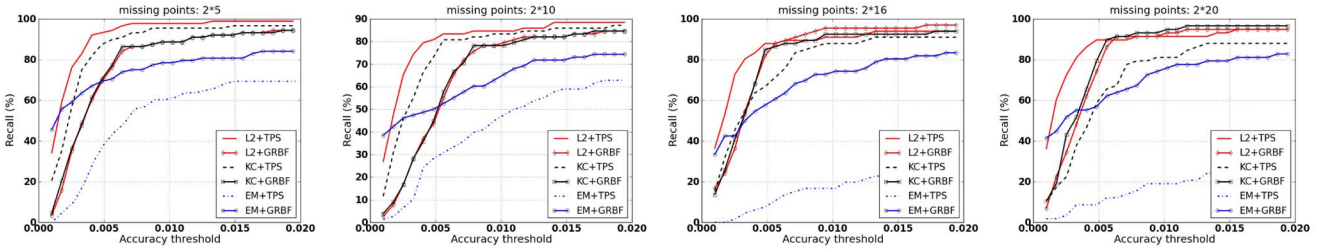


Fig. 9. Performances of nonrigid point set registration algorithms on the incomplete fish shape pairs.

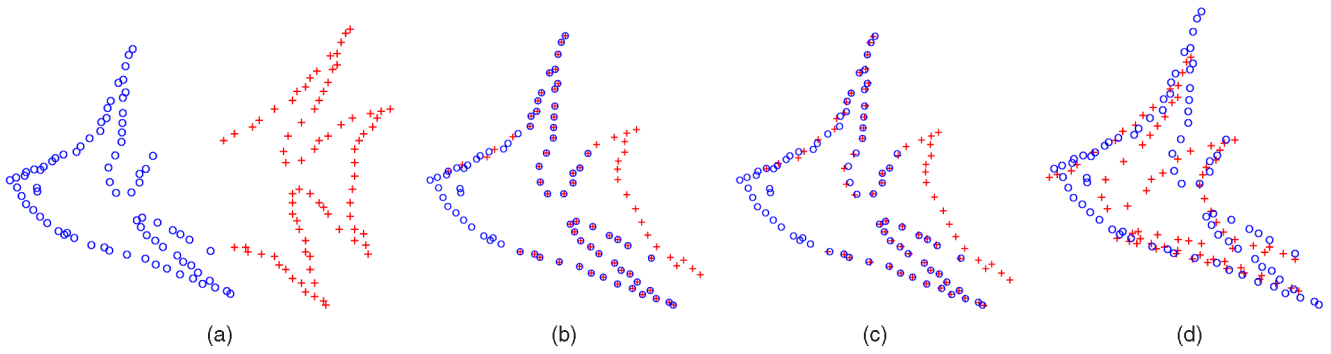


Fig. 10. An experiment on a pair of incomplete fish shapes. (a) The initial configuration with 20 points removed from the head part of one fish and 20 points removed from the tail part of another fish. (b), (c), and (d) The results of aligning the fish with no head to the one with no tail using the Jian and Vemuri [5] algorithm (L2+TPS) in (b), results from the Myronenko et al. [6] algorithm (EM+GRBF) in (c), and the Chui and Rangarajan [2] algorithm (EM+TPS) in (d), respectively.

7 Discussion and Conclusion

Point set registration is a problem of pivotal importance that continues to attract considerable interest in the computer vision community. In this work, we have presented a novel probabilistic modeling framework for rigid and nonrigid point set registration. The core idea of the proposed framework is to model each of the point sets using a Gaussian mixture and reformulate the point set registration problem as the problem of aligning one point set to another such that a statistical discrepancy measure between the two corresponding mixtures is minimized. It is interesting to see that while seemingly quite different, several existing algorithms [1], [2], [3], [4], [5], [6] in the field are closely related and can be reinterpreted meaningfully in this framework.

We observe that the registration algorithms presented in [2], [3], [6] all maximize the likelihood function of the fixed scene point set as the "data" with respect to a Gaussian mixture representing the moving model point set as the "model," which is equivalent to minimizing the KL divergence between the empirical distribution of the "data" and the Gaussian mixture representing the "model." We have also shown that the basic idea of the ICP algorithm can be reinterpreted as minimizing an approximation of the KL divergence between Gaussian mixtures, which may explain the nonrobustness of a naive ICP implementation. In order to achieve the robustness while maximizing the likelihood during the EM cycle, algorithms in [2], [3], [6] have to include an additional Gaussian density component in the mixture model to handle the outliers and hence require the tuning of additional parameters.

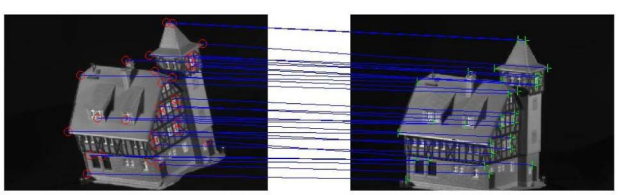


Fig. 11. Two images (left: frame 30; right: frame 80) from the *CMU house sequence*. Thirty landmarks points were manually marked in each of the images with known correspondences.

Our instantiation of the presented general framework is based on the L2 distance between two Gaussian mixtures. One obvious motivation of choosing the L2 distance is its closed-form expression for Gaussian mixtures, which in turn enables a fast implementation. Another rationale behind our choice is its inherent relation to L_2E, a robust estimator minimizing the L2 distance between densities. The robustness of L_2E has been carefully studied in statistics literature [34], [35] and one salient advantage of L_2E over some other M-estimators [37] is that it requires no specification of tuning parameters. Moreover, it has been recently shown in [58] that the pairwise L2 distance can be generalized to achieve groupwise registration [59], [60], which is very useful for image and shape atlas construction, whereas no such extensions have been shown for the other competing methods. Finally, we have demonstrated the merits of our method in rigid and nonrigid point set registration via experiments on synthetic and real data drawn from various application domains. Additionally, some quantitative evaluations were performed to depict the favorable performance of our nonrigid registration method compared to the competing nonrigid registration methods [2], [6], [7].

In summary, we have presented a unified framework in the form of a divergence family that allows for interpreting several of the existing methods as special cases of the family. Our experimental studies indicate that it is still not possible to declare any one algorithm an outright winner

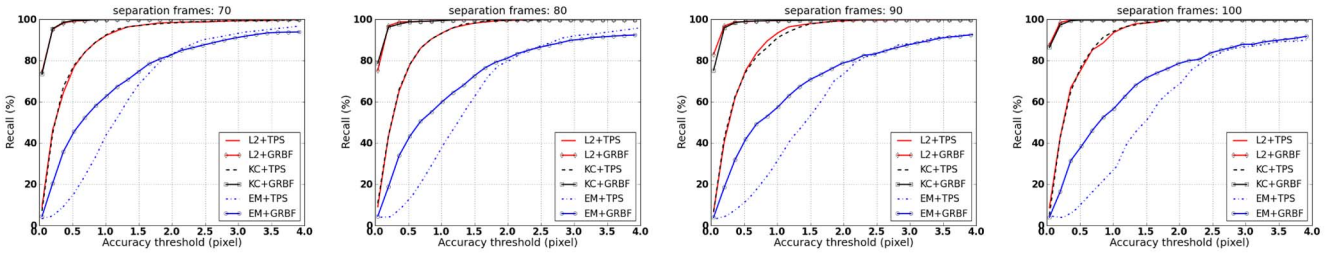


Fig. 12. Performance of the nonrigid point set registration algorithms on the CMU house sequence for different separation frames (70, 80, 90, and 100 frames). The average value is taken over image pairs with the same spacing in the sequence. Since there are 111 frames, the number of image pairs spaced by these numbers of frames are, respectively, 41, 31, 21, and 11.

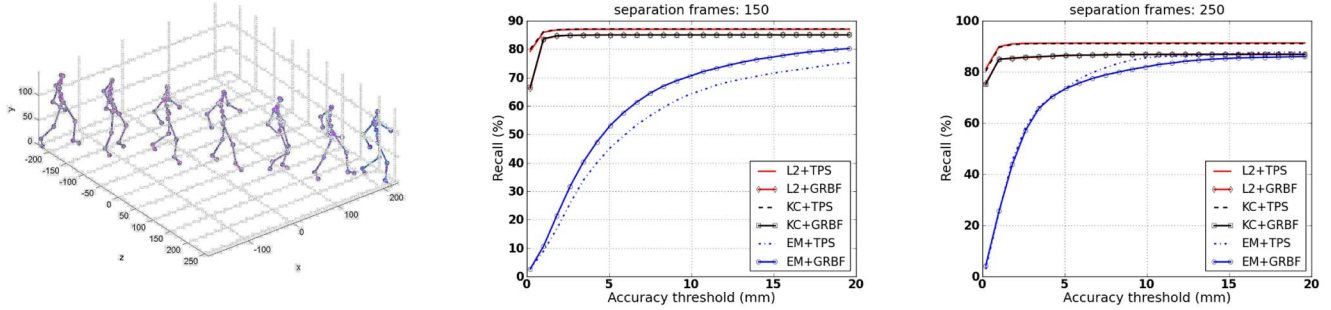


Fig. 13. Experiment on a sample "swagger" motion capture data. On the left are several selected frames from this data. Note that the unit on the axis is *cm*; In the middle and right are the recall-accuracy curves averaged on the matching frames separated by 150 and 250 frames, respectively.

不同的评价标准会有不同的结果....嗯.....

over other competing methods in terms of practical applicability due to the lack of diversified standard data sets, widely accepted criteria, and intrinsic difficulties in principled parameter tuning, all of which should be worthy for further investigation.

ACKNOWLEDGMENTS

This research was in part supported by NIH RO1 NS046812 and EB007082 to Baba C. Vemuri and was performed when Bing Jian was at the University of Florida.

REFERENCES

[1] P.J. Besl and N.D. McKay, "A Method for Registration of 3D Shapes," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 14, no. 2, pp. 239-256, Feb. 1992.

[2] H. Chui and A. Rangarajan, "A New Algorithm for Non-Rigid Point Matching," *Proc. IEEE Conf. Computer Vision and Pattern Recognition,* pp. 44-51, 2000.

[3] H. Chui and A. Rangarajan, "A Feature Registration Framework Using Mixture Models," *Proc. IEEE Workshop on Math. Methods in Biomedical Image Analysis,* pp. 190-197, 2000.

[4] Y. Tsin and T. Kanade, "A Correlation-Based Approach to Robust Point Set Registration," *Proc. European Conf. Computer Vision,* pp. 558-569, 2004.

[5] B. Jian and B.C. Vemuri, "A Robust Algorithm for Point Set Registration Using Mixture of Gaussians," *Proc. IEEE Int'l Conf. Computer Vision,* pp. 1246-1251, 2005.

[6] A. Myronenko, X.B. Song, and M.Á. Carreira-Perpiñán, "Non-Rigid Point Set Registration: Coherent Point Drift," *Advances in Neural Information Processing Systems,* pp. 1009-1016, MIT Press, 2006.

[7] A. Myronenko and X.B. Song, "Point-Set Registration: Coherent Point Drift," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 32, no. 12, pp. 2262-2275, Dec. 2010.

[8] P.B. van Wamelen, Z. Li, and S.S. Iyengar, "A Fast Expected Time Algorithm for the 2D Point Pattern Matching Problem," *Pattern Recognition,* vol. 37, no. 8, pp. 1699-1711, 2004.

[9] D. Conte, P. Foggia, C. Sansone, and M. Vento, "Thirty Years of Graph Matching in Pattern Recognition," *Int'l J. Pattern Recognition and Artificial Intelligence,* vol. 18, no. 3, pp. 265-298, 2004.

[10] B. Luo and E.R. Hancock, "A Unified Framework for Alignment and Correspondence," *Computer Vision and Image Understanding,* vol. 92, pp. 26-55, 2003.

[11] M. Carcassoni and E.R. Hancock, "Spectral Correspondence for Point Pattern Matching," *Pattern Recognition,* vol. 36, no. 1, pp. 193-204, 2003.

[12] T.S. Caetano, T. Caelli, D. Schuurmans, and D.A. Barone, "Graphical Models and Point Pattern Matching," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 28, no. 10, pp. 1646-1663, Oct. 2006.

[13] 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. 509-522, Apr. 2002.

[14] Y. Zheng and D.S. Doermann, "Robust Point Matching for Nonrigid Shapes by Preserving Local Neighborhood Structures," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 28, no. 4, pp. 643-649, Apr. 2006.

[15] K.S. Arun, T.S. Huang, and S.D. Blostein, "Least-Squares Fitting of Two 3d Point Sets," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 9, no. 5, pp. 698-700, Sept. 1987.

[16] Z. Zhang, "Iterative Point Matching for Registration of Free-Form Curves and Surfaces," *Int'l J. Computer Vision,* vol. 13, no. 2, pp. 119-152, 1994.

[17] 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.

[18] C.V. Stewart, C.-L. Tsai, and B. Roysam, "The Dual Bootstrap Iterative Closest Point Algorithm with Application to Retinal Image Registration," *IEEE Trans. Medical Imaging,* vol. 22, no. 11, pp. 1379-1394, Nov. 2003.

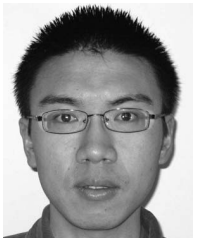
[19] D. Chetverikov, D. Stepanov, and P. Krsek, "Robust Euclidean Alignment of 3D Point Sets: The Trimmed Iterative Closest Point Algorithm," *Image and Vision Computing,* vol. 23, pp. 299-309, 2005.

[20] A.P. Dempster, N.M. Laird, and D.B. Rubin, "Maximum Likelihood from Incomplete Data via the EM Algorithm," *J. Royal Statistical Soc. B,* vol. 39, no. 1, pp. 1-38, 1977.

[21] M. Sofka, G. Yang, and C.V. Stewart, "Simultaneous Covariance Driven Correspondence (CDC) and Transformation Estimation in the Expectation Maximization," *Proc. IEEE Conf. Computer Vision and Pattern Recognition,* 2007.

[22] R.P. Horaud, F. Forbes, M. Yguel, G. Dewaele, and J. Zhang, "Rigid and Articulated Point Registration with Expectation Conditional Maximization," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. PP, no. 99, p. 1, Apr. 2010.

[23] A.W. Fitzgibbon, "Robust Registration of 2D and 3D Point Sets," *Image and Vision Computing,* vol. 21, nos. 13/14, pp. 1145-1153, 2003.
[24] J. Glaunes, A. Trouvé, and L. Younes, "Diffeomorphic Matching of Distributions: A New Approach for Unlabelled Point-Sets and Sub-Manifolds Matching," *Proc. IEEE CS Conf. Computer Vision and Pattern Recognition,* pp. 712-718, 2004.
[25] K. Kanatani, "Uncertainty Modeling and Model Selection for Geometric Inference," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 26, no. 10, pp. 1307-1319, Oct. 2004.
[26] D.W. Scott, *Multivariate Density Estimation: Theory, Practice, and Visualization.* John Wiley & Sons, 1992.
[27] M. Wand and M. Jones, *Kernel Smoothing.* Chapman & Hall, 1995.
[28] J. Shi and C. Tomasi, "Good Features to Track," *Proc. IEEE CS Conf. Computer Vision and Pattern Recognition,* pp. 593-600, 1994.
[29] D.D. Morris and T. Kanade, "A Unified Factorization Algorithm for Points, Line Segments and Planes with Uncertainty Models," *Proc. Int'l Conf. Computer Vision,* pp. 696-702, 1998.
[30] Y. Kanazawa and K. Kanatani, "Do We Really Have to Consider Covariance Matrices for Image Features?" *Proc. IEEE Int'l Conf. Computer Vision,* pp. 301-306, 2001.
[31] D. Comaniciu, V. Ramesh, and P. Meer, "Kernel-Based Object Tracking," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 25, no. 5, pp. 564-575, May 2003.
[32] D. Comaniciu, "An Algorithm for Data-Driven Bandwidth Selection," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 25, no. 2, pp. 281-288, Feb. 2003.
[33] D.W. Scott and W.F. Szewczyk, "From Kernels to Mixtures," *Technometrics,* vol. 43, no. 3, pp. 323-335, 2001.
[34] D. Scott, "Parametric Statistical Modeling by Minimum Integrated Square Error," *Technometrics,* vol. 43, no. 3, pp. 274-285, 2001.
[35] A. Basu, I.R. Harris, N.L. Hjort, and M.C. Jones, "Robust and Efficient Estimation by Minimising a Density Power Divergence," *Biometrika,* vol. 85, no. 3, pp. 549-559, 1998.
[36] L.M. Bregman, "The Relaxation Method of Finding the Common Points of Convex Sets and Its Application to the Solution of Problems in Convex Programming," *USSR Computational Math. and Math. Physics,* vol. 7, pp. 200-217, 1967.
[37] F. Hampel, E. Ronchetti, P. Rousseeuw, and W. Stahel, *Robust Statistics: The Approach Based on Influence Functions.* Wiley, 1986.
[38] C.V. Stewart, "Robust Parameter Estimation in Computer Vision," *SIAM Rev.,* vol. 41, no. 3, pp. 513-537, 1999.
[39] J. Liu, B.C. Vemuri, and J.L. Marroquín, "Local Frequency Representations for Robust Multimodal Image Registration," *IEEE Trans. Medical Imaging,* vol. 21, no. 5, pp. 462-469, May 2002.
[40] L. Yang, P. Meer, and D.J. Foran, "Unsupervised Segmentation Based on Robust Estimation and Color Active Contour Models," *IEEE Trans. Information Technology in Biomedicine,* vol. 9, no. 3, pp. 475-486, Sept. 2005.
[41] M. Mihoko and S. Eguchi, "Robust Blind Source Separation by Beta Divergence," *Neural Computation,* vol. 14, no. 8, pp. 1859-1886, 2002.
[42] J. Goldberger, S. Gordon, and H. Greenspan, "An Efficient Image Similarity Measure Based on Approximations of KL-Divergence between Two Gaussian Mixtures," *Proc. IEEE Int'l Conf. Computer Vision,* pp. 487-493, 2003.
[43] R. Jenssen, D. Erdogmus, J.C. Príncipe, and T. Eltoft, "The Laplacian PDF Distance: A Cost Function for Clustering in a Kernel Feature Space," *Advances in Neural Information Processing Systems,* MIT Press, 2004.
[44] R. Sandhu, S. Dambreville, and A. Tannenbaum, "Point Set Registration via Particle Filtering and Stochastic Dynamics," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 32, no. 8, pp. 1459-1473, Aug. 2010.
[45] M.C. Jones, N.L. Hjort, I.R. Harris, and A. Basu, "A Comparison of Related Density-Based Minimum Divergence Estimators," *Biometrika,* vol. 88, no. 3, pp. 865-873, 2001.
[46] M.P. Windham, "Robustifying Model Fitting," *J. Royal Statistical Soc. B,* vol. 57, pp. 599-609, 1995.
[47] L. Greengard and J. Strain, "The Fast Gauss Transform," *SIAM J. Scientific Computing,* vol. 12, no. 1, pp. 79-94, 1991.
[48] C. Yang, R. Duraiswami, N.A. Gumerov, and L.S. Davis, "Improved Fast Gauss Transform and Efficient Kernel Density Estimation," *Proc. IEEE Int'l Conf. Computer Vision,* pp. 464-471, 2003.
[49] F.L. Bookstein, "Principal Warps: Thin-Plate Splines and the Decomposition of Deformations," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 11, no. 6, pp. 567-585, June 1989.
[50] G. Wahba, *Spline Models for Observational Data.* SIAM, 1990.
[51] K. Rohr, *Landmark-Based Image Analysis: Using Geometric and Intensity Models.* Kluwer Academic Publishers, 2001.
[52] A.L. Yuille and N.M. Grzywacz, "A Mathematical Analysis of the Motion Coherence Theory," *Int'l J. Computer Vision,* vol. 3, no. 2, pp. 155-175, 1989.
[53] C. Zhu, R.H. Byrd, P. Lu, and J. Nocedal, "Algorithm 778: L-BFGS-B: Fortran Subroutines for Large-Scale Bound-Constrained Optimization," *ACM Trans. Math. Software,* vol. 23, no. 4, pp. 550-560, 1997.
[54] C. Harris and M. Stephens, "A Combined Corner and Edge Detector," *Proc. Fourth Alvey Vision Conf.,* pp. 147-152, 1988.
[55] 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, no. 6, pp. 381-395, 1981.
[56] R. Hartley and A. Zisserman, *Multiple View Geometry in Computer Vision.* Cambridge Univ. Press, 2003.
[57] J. Starck and A. Hilton, "Correspondence Labelling for Wide-Timeframe Free-Form Surface Matching," *Proc. IEEE Int'l Conf. Computer Vision,* 2007.
[58] F. Wang, T. Syeda-Mahmood, B.C. Vemuri, D. Beymer, and A. Rangarajan, "Closed-Form Jensen-Renyi Divergence for Mixture of Gaussians and Applications to Group-Wise Shape Registration," *Proc. Int'l Conf. Medical Image Computing and Computer Assisted Intervention,* pp. 648-655, 2009.
[59] F. Wang, B.C. Vemuri, A. Rangarajan, and S.J. 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.
[60] T. Chen, B.C. Vemuri, A. Rangarajan, and S.J. Eisenschenk, "Group-Wise Point-Set Registration Using a Novel CDF-Based Havrda-Charvát Divergence," *Int'l J. Computer Vision,* vol. 86, no. 1, pp. 111-124, 2010.



Bing Jian received the BS and MS degrees from the University of Science and Technology of China, Hefei, China, and the PhD degree from the University of Florida, Gainesville. He is currently with CAD R&D, Siemens Healthcare, USA, as a staff scientist. His research interests include computer vision, pattern recognition, and image processing, and analysis. He is a member of the IEEE Computer Society.

Baba C. Vemuri received the PhD degree in electrical and computer engineering from the University of Texas at Austin in 1987. He joined the Department of Computer and Information Sciences at the University of Florida, Gainesville, in 1987, and is currently a university research foundation professor of computer and information sciences and engineering. He was a coprogram chair of the 11th IEEE International Conference on Computer Vision (ICCV '07). He has been an area chair and a program committee member of several IEEE, MICCAI, and IPMI conferences. He was an associate editor for several journals, including the *IEEE Transactions on Pattern Analysis and Machine Intelligence* (from 1992 to 1996) and the *IEEE Transactions on Medical Imaging* (from 1997 to 2003). He is currently an associate editor for the *Journal of Medical Image Analysis* and the *Journal of Computer Vision and Image Understanding.* His research interests include medical image analysis, computational vision, information geometry, machine learning, and applied mathematics. For the last several years, his research work has primarily focused on information geometric methods applied to computer vision and medical image analysis. He has published more than 150 refereed articles in journals and conference proceedings and has received several best paper awards to date. He is a fellow of the ACM and the IEEE.