Robust Estimation of Nonrigid Transformation for Point Set Registration

Jiayi Ma[1,2], Ji Zhao[3], Jinwen Tian[1], Zhuowen Tu[4], and Alan L. Yuille[2]
[1]Huazhong University of Science and Technology, [2]Department of Statistics, UCLA
[3]The Robotics Institute, CMU, [4]Lab of Neuro Imaging, UCLA
{jyma2010, zhaoji84, zhuowen.tu}@gmail.com, jwtian@hust.edu.cn, yuille@stat.ucla.edu

Abstract

We present a new point matching algorithm for robust nonrigid registration. The method iteratively recovers the point correspondence and estimates the transformation between two point sets. In the first step of the iteration, feature descriptors such as shape context are used to establish rough correspondence. In the second step, we estimate the transformation using a robust estimator called L_2E. This is the main novelty of our approach and it enables us to deal with the noise and outliers which arise in the correspondence step. The transformation is specified in a functional space, more specifically a reproducing kernel Hilbert space. We apply our method to nonrigid sparse image feature correspondence on 2D images and 3D surfaces. Our results quantitatively show that our approach outperforms state-of-the-art methods, particularly when there are a large number of outliers. Moreover, our method of robustly estimating transformations from correspondences is general and has many other applications.

1. Introduction

Point set registration is a fundamental problem which frequently arises in computer vision, medical image analysis, and pattern recognition [5, 4, 6]. Many tasks in these fields – such as stereo matching, shape matching, image registration and content-based image retrieval – can be formulated as a point matching problem because point representations are general and easy to extract [5]. The points in these tasks are typically the locations of interest points extracted from an image, or the edge points sampled from a shape contour. The registration problem then reduces to determining the correct correspondence and to find the underlying spatial transformation between two point sets extracted from the input data.

The registration problem can be categorized into rigid or nonrigid registration depending on the application and the form of the data. Rigid registration, which only involves a small number of parameters, is relatively easy and has been widely studied [5, 4, 7, 19, 14]. By contrast, nonrigid registration is more difficult because the underlying nonrigid transformations are often unknown, complex, and hard to model [6]. But nonrigid registration is very important because it is required for many real world tasks including hand-written character recognition, shape recognition, deformable motion tracking and medical image registration.

In this paper, we focus on the nonrigid case and present a robust algorithm for nonrigid point set registration. There are two unknown variables we have to solve for in this problem: the correspondence and the transformation. Although solving for either variable without information regarding the other is difficult, an iterated estimation framework can be used [4, 3, 6]. In this iterative process, the estimate of the correspondence is used to refine the estimate of the transformation, and vice versa. But a problem arises if there are errors in the correspondence which occurs in many applications particularly if the transformation is large and/or there are outliers in the data (e.g., data points that are not undergoing the non-rigid transformation). In this situation, the estimate of the transformation will degrade badly unless it is performed robustly. The main contribution of our approach is to robustly estimate the transformations from the correspondences using a robust estimator named the L_2-Minimizing Estimate (L_2E) [20, 2].

More precisely, our approach iteratively recovers the point correspondences and estimates the transformation between two point sets. In the first step of the iteration, feature descriptors such as shape context are used to establish correspondence. In the second step, we estimate the transformation using the robust estimator L_2E. This estimator enable us to deal with the noise and outliers in the correspondences. The nonrigid transformation is modeled in a functional space, called the reproducing kernel Hilbert space (RKHS) [1], in which the transformation function has an explicit kernel representation.

1.1. Related Work

The iterated closest point (ICP) algorithm [4] is one of the best known point registration approaches. It uses

nearest-neighbor relationships to assign a binary correspondence, and then uses estimated correspondence to refine the transformation. Belongie *et al*. [3] introduced a method for registration based on the shape context descriptor, which incorporates the neighborhood structure of the point set and thus helps establish correspondence between the point sets. But these methods ignore robustness when they recover the transformation from the correspondence. In related work, Chui and Rangarajan [6] established a general framework for estimating correspondence and transformations for nonrigid point matching. They modeled the transformation as a thin-plate spline and did robust point matching by an algorithm (TRS-RPM) which involved deterministic annealing and soft-assignment. Alternatively, the coherence point drift (CPD) algorithm [17] uses Gaussian radial basis functions instead of thin-plate splines. Another interesting point matching approach is the kernel correlation (KC) based method [22], which was later improved in [9]. Zheng and Doermann [27] introduced the notion of a neighborhood structure for the general point matching problem, and proposed a matching method, the robust point matching-preserving local neighborhood structures (RPM-LNS) algorithm. Other related work includes the relaxation labeling method, generalized in [11], and the graph matching approach for establishing feature correspondences [21].

The main contributions of our work include: (i) we propose a new robust algorithm to estimate a spatial transformation/mapping from correspondences with noise and outliers; (ii) we apply the robust algorithm to nonrigid point set registration and also to sparse image feature correspondence.

2. Estimating Transformation from Correspondences by L_2E

Given a set of point correspondences $S = \{(x_i, y_i)\}_{i=1}^n$, which are typically perturbed by noise and by outlier points which undergo different transformations, the goal is to estimate a transformation $f : y_i = f(x_i)$ and fit the inliers.

In this paper we make the assumption that the noise on the inliers is Gaussian on each component with zero mean and uniform standard deviation σ (our approach can be directly applied to other noise models). More precisely, an inlier point correspondence (x_i, y_i) satisfies $y_i - f(x_i) \sim N(\mathbf{0}, \sigma^2 I)$, where I is an identity matrix of size $d \times d$, with d being the dimension of the point. The data $\{y_i - f(x_i)\}_{i=1}^n$ can then be thought of as a sample set from a multivariate normal density $N(\mathbf{0}, \sigma^2 I)$ which is contaminated by outliers. The main idea of our approach is then, to find the largest portion of the sample set (e.g., the underlying inlier set) that "matches" the normal density model, and hence estimate the transformation f for the inlier set. Next, we introduce a robust estimator named L_2-minimizing estimate (L_2E) which we use to estimate the transformation f.

2.1. Problem Formulation Using L_2E: Robust Estimation

Parametric estimation is typically done using maximum likelihood estimation (MLE). It can be shown that MLE is the optimal estimator if it is applied to the correct probability model for the data (or to a good approximation). But MLE can be badly biased if the model is not a sufficiently good approximation or, in particular, there are a significant fraction of outliers. In many point matching problems, it is desirable to have a robust estimator of the transformation f because the point correspondence set S usually contains outliers. There are two choices: (i) to build a more complex model that includes the outliers – which is complex since it involves modeling the outlier process using extra (hidden) variables which enable us to identify and reject outliers, or (ii) to use an estimator which is different from MLE but less sensitive to outliers, as described in Huber's robust statistics [8]. In this paper, we use the second method and adopt the L_2E estimator [20, 2], a robust estimator which minimizes the L_2 distance between densities, and is particularly appropriate for analyzing massive data sets where data cleaning (to remove outliers) is impractical. More formally, L_2E estimator for model $f(x|\theta)$ recommends estimating the parameter θ by minimizing the criterion:

$$L_2E(\theta) = \int f(x|\theta)^2 dx - \frac{2}{n}\sum_{i=1}^{n} f(x_i|\theta). \qquad (1)$$

To get some intuition for why L_2E is robust, observe that its penalty for a low probability point x_i is $-f(x_i|\theta)$ which is much less than the penalty of $-\log f(x_i|\theta)$ given by MLE (which becomes infinite as $f(x_i|\theta)$ tends to 0). Hence MLE is reluctant to assign low probability to any points, including outliers, and hence tends to be biased by outliers. By contrast, L_2E can assign low probabilities to many points, hopefully to the outliers, without paying too high a penalty. To demonstrate the robustness of L_2E, we present a line-fitting example which contrasts the behavior of MLE and L_2E, see Fig. 1. The goal is to fit a linear regression model, $y = \alpha x + \epsilon$, with residual $\epsilon \sim N(0, 1)$, by estimating α using MLE and L_2E. This gives, respectively, $\hat{\alpha}_{MLE} = \arg\max_\alpha \sum_{i=1}^n \log \phi(y_i - \alpha x_i|0, 1)$ and $\hat{\alpha}_{L_2E} = \arg\min_\alpha \left[\frac{1}{2\sqrt{\pi}} - \frac{2}{n}\sum_{i=1}^n \phi(y_i - \alpha x_i|0, 1)\right]$, where $\phi(x|\mu, \Sigma)$ denotes the normal density.

As shown in Fig. 1, L_2E is very resistant when we contaminate the data by outliers, but MLE does not show this desirable property. L_2E always has a global minimum at approximately 0.5 (the correct value for α) but MLE's estimates become steadily worse as the amount of outliers increases. Observe, in the bottom right figure, that L_2E also has a local minimum near $\alpha = 2$, which becomes deeper as the number n of outliers increases so that the two minima become approximately equal when $n = 200$. This is ap-

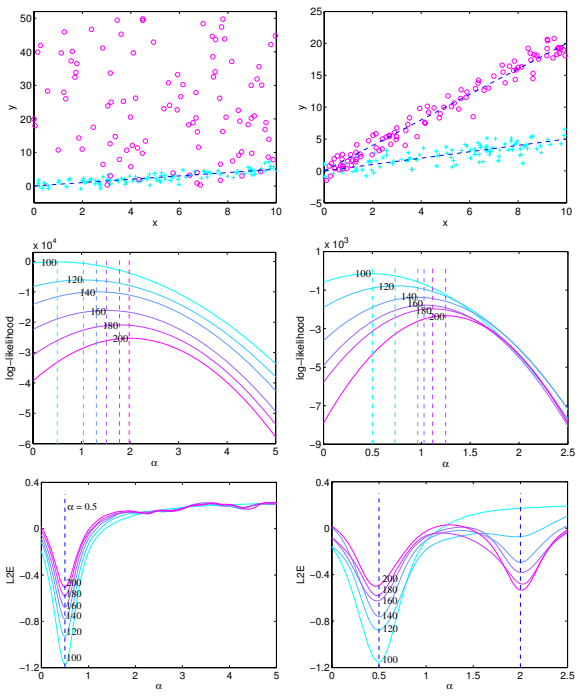

Figure 1. Comparison between L_2E and MLE for linear regression as the number of outliers varies. Top row: data samples, where the inliers are shown by cyan pluses, and the outliers by magenta circles. The goal is to estimate the slope α of the line model $y = \alpha x$. We vary the number of data samples $n = 100, 120, \cdots, 200$, by always adding 20 new outliers (retaining the previous samples). In the second column the outliers are generated from another line model $y = 2x$. Middle and bottom rows: the curves of MLE and L_2E respectively. The MLE estimates are correct for $n = 100$ but rapidly degrade as we add outliers, see how the peak of the log-likelihood changes in the second row. By contrast, (see third row) L_2E estimates α correctly even when half the data is outliers and also develops a local minimum to fit the outliers when appropriate (third row, right column). Best viewed in color.

propriate because, in this case, the contaminated data also comes from the same linear parametric model with slope $\alpha = 2$, e.g., $y = 2x$.

We now apply the L_2E formulation in (1) to the point matching problem, assuming that the noise of the inliers is given by a normal distribution, and obtain the following functional criterion:

$$L_2E(f, \sigma^2) = \frac{1}{2^d(\pi\sigma)^{d/2}} - \frac{2}{n}\sum_{i=1}^{n} \phi\left(y_i - f(x_i)|\mathbf{0}, \sigma^2 I\right). \tag{2}$$

We model the nonrigid transformation f by requiring it to lie within a specific functional space, namely a reproducing kernel Hilbert space (RKHS) [1, 24, 16]. Note that other parameterized transformation models, for example, thin-plate splines (TPS) [23, 15], can also be easily incorporated into our formulation.

We define an RKHS $\mathcal{H}$ by a positive definite matrix-valued kernel $\Gamma : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}^{d\times d}$. The optimal transformation f which minimizes the L_2E functional (2) then takes the form $f(x) = \sum_{i=1}^{n} \Gamma(x, x_i)c_i$ [16, 26], where the coefficient c_i is a $d \times 1$ dimensional vector (to be determined). Hence, the minimization over the infinite dimensional Hilbert space reduces to finding a finite set of n coefficients c_i. But in point correspondence problem the point set typically contains hundreds or thousands of points, which causes significant complexity problems (in time and space). Consequently, we adopt a sparse approximation, and randomly pick only a subset of size m input points $\{\tilde{x}_i\}_{i=1}^{m}$ to have nonzero coefficients in the expansion of the solution. This follows [18] who found that this approximation works well and that simply selecting a random subset of the input points in this manner, performs no worse than more sophisticated and time-consuming methods. Therefore, we seek a solution of form

$$f(x) = \sum_{i=1}^{m} \Gamma(x, \tilde{x}_i)c_i. \tag{3}$$

The chosen point set $\{\tilde{x}_i\}_{i=1}^{m}$ are somewhat analogous to "control points" [5]. By including a regularization term for imposing smooth constraint on the transformation, the L_2E functional (2) becomes:

$$L_2E(f, \sigma^2) = \frac{1}{2^d(\pi\sigma)^{d/2}} - \frac{2}{n}\sum_{i=1}^{n} \frac{1}{(2\pi\sigma^2)^{d/2}} e^{-\frac{\left\|y_i - \sum_{j=1}^{m}\Gamma(x_i,\tilde{x}_j)c_j\right\|^2}{2\sigma^2}} + \lambda\|f\|_\Gamma^2, \tag{4}$$

where $\lambda > 0$ controls the strength of regularization, and the stabilizer $\|f\|_\Gamma^2$ is defined by an inner product, e.g., $\|f\|_\Gamma^2 = \langle f, f\rangle_\Gamma$. By choosing a diagonal decomposable kernel [26]: $\Gamma(x_i, x_j) = e^{-\beta\|x_i - x_j\|^2} I$ with β determining the width of the range of interaction between samples (i.e. neighborhood size), the L_2E functional (4) may be conveniently expressed in the following matrix form:

$$L_2E(C, \sigma^2) = \frac{1}{2^d(\pi\sigma)^{d/2}} - \frac{2}{n}\sum_{i=1}^{n} \frac{1}{(2\pi\sigma^2)^{d/2}} e^{-\frac{\left\|y_i^{\mathrm{T}} - U_{i,\cdot}C\right\|^2}{2\sigma^2}} + \lambda\,\mathrm{tr}(C^{\mathrm{T}}\mathbf{\Gamma}C), \tag{5}$$

where kernel matrix $\mathbf{\Gamma} \in \mathbb{R}^{m\times m}$ is called the Gram matrix with $\mathbf{\Gamma}_{ij} = \Gamma(\tilde{x}_i, \tilde{x}_j) = e^{-\beta\|\tilde{x}_i - \tilde{x}_j\|^2}$, $U \in \mathbb{R}^{n\times m}$ with $U_{ij} = \Gamma(x_i, \tilde{x}_j) = e^{-\beta\|x_i - \tilde{x}_j\|^2}$, $U_{i,\cdot}$ denotes the i-th row of the matrix U, $C = (c_1, \cdots, c_m)^{\mathrm{T}}$ is the coefficient matrix of size $m \times d$, and $\mathrm{tr}(\cdot)$ denotes the trace.

2.2. Estimation of the Transformation

Estimating the transformation requires taking the derivative of the L_2E cost function, see equation (5), with respect

Algorithm 1: Estimation of Transformation from Correspondences

Input: Correspondence set $S = \{(x_i, y_i)\}_{i=1}^n$, parameters γ, β, λ

Output: Optimal transformation f

1 Construct Gram matrix $\mathbf{\Gamma}$ and matrix U;
2 Initialize parameter σ^2 and C;
3 Deterministic annealing:
4 Using the gradient (6), optimize the objective function (5) by a numerical technique (e.g., the quasi-Newton algorithm with C as the old value);
5 Update the parameter $C \leftarrow \arg\min_C L_2E(C, \sigma^2)$;
6 Anneal $\sigma^2 = \gamma\sigma^2$;
7 The transformation f is determined by equation (3).

to the coefficient matrix C, which is given by:

$$\frac{\partial L_2 E}{\partial C} = \frac{2U^{\mathrm{T}}[V \circ (W \otimes \mathbf{1}_{1\times d})]}{n\sigma^2(2\pi\sigma^2)^{d/2}} + 2\lambda\mathbf{\Gamma}C, \qquad (6)$$

where $V = UC - Y$ and $Y = (y_1, \cdots, y_n)^{\mathrm{T}}$ are matrices of size $n \times d$, $W = \exp\{\mathrm{diag}(VV^{\mathrm{T}})/2\sigma^2\}$ is an $n \times 1$ dimensional vector, $\mathrm{diag}(\cdot)$ is the diagonal of a matrix, $\mathbf{1}_{1\times d}$ is an $1 \times d$ dimensional row vector of all ones, $\circ$ denotes the Hadamard product, and $\otimes$ denotes the Kronecker product.

By using the derivative in equation (6), we 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. But the cost function (5) is convex only in the neighborhood of the optimal solution. Hence to improve convergence we use a coarse-to-fine strategy by applying deterministic annealing on the inlier noise parameter σ^2. This starts with a large initial value for σ^2 which is gradually reduced by $\sigma^2 \mapsto \gamma\sigma^2$, where γ is the annealing rate. Our algorithm is outlined in algorithm 1.

Computational complexity. By examining equations (5) and (6), we see that the costs of updating the objective function and gradient are both $O(dm^2 + dmn)$. For the numerical optimization method, we choose the Matlab Optimization toolbox, which implicitly uses the BFGS Quasi-Newton method with a mixed quadratic and cubic line search procedure. Thus the total complexity is approximately $O(dm^2 + dm^3 + dmn)$. In our implementation, the number m of the control points required to construct the transformation f in equation (3) is in general not large, and so use $m = 15$ for all the results in this paper (increasing m only gave small changes to the results). The dimension d of the data in feature point matching for vision applications is typically 2 or 3. Therefore, the complexity of our method can be simply expressed as $O(n)$, which is about linear in the number of correspondences. This is important since it enables our method to be applied to large scale data.

2.3. Implementation Details

The performance of point matching algorithms depends, typically, on the coordinate system in which points are expressed. We use data normalization to control for this. More specifically, we perform a linear re-scaling of the correspondences so that the points in the two sets both have zero mean and unit variance.

We define the transformation f as the initial position plus a displacement function v: $f(x) = x + v(x)$ [17], and solve for v instead of f. This can be achieved simply by setting the output y_i to be $y_i - x_i$.

Parameter settings. There are three main parameters in this algorithm: γ, β and λ. The parameter γ controls the annealing rate. The parameters β and λ control the influence of the smoothness constraint on the transformation f. In general, we found our method was very robust to parameter changes. We set $\gamma = 0.5$, $\beta = 0.8$ and $\lambda = 0.1$ throughout this paper. Finally, the parameter σ^2 and C in line 2 of algorithm 1 were initialized to 0.05 and 0 respectively.

3. Nonrigid Point Set Registration

Point set registration aims to align two point sets $\{x_i\}_{i=1}^n$ (the model point set) and $\{y_j\}_{j=1}^l$ (the target point set). Typically, in the nonrigid case, it requires estimating a nonrigid transformation f which warps the model point set to the target point set. We have shown above that once we have established the correspondence between the two point sets even with noise and outliers, we are able to estimate the underlying transformation between them. Next, we discuss how to find correspondences between two point sets.

3.1. Establishment of Point Correspondence

Recall that our method described above does not jointly solve the transformation and point correspondence. In order to use algorithm 1 to solve the transformation between two point sets, we need initial correspondences.

In general, if the two point sets have similar shapes, the corresponding points have similar neighborhood structures which could be incorporated into a feature descriptor. Thus finding correspondences between two point sets is equivalent to finding for each point in one point set (e.g., the model) the point in the other point set (e.g., the target) that has the most similar feature descriptor. Fortunately, the initial correspondences need not be very accurate, since our method is robust to noise and outliers. Inspired by these facts, we use shape context [3] as the feature descriptor in the 2D case, using the Hungarian method for matching with the χ^2 test statistic as the cost measure. In the 3D case, the spin image [10] can be used as a feature descriptor, where the local similarity is measured by an improved correlation

Algorithm 2: Nonrigid Point Set Registration

Input: Two point sets $\{x_i\}_{i=1}^n$, $\{y_j\}_{j=1}^l$
Output: Aligned model point set $\{\hat{x}_i\}_{i=1}^n$

Compute feature descriptors for the target point set $\{y_j\}_{j=1}^l$;
repeat
  Compute feature descriptors for the model point set $\{x_i\}_{i=1}^n$;
  Estimate the initial correspondences based on the feature descriptors of two point sets;
  Solve the transformation f warping the model point set to the target point set using algorithm 1;
  Update model point set $\{x_i\}_{i=1}^n \leftarrow \{f(x_i)\}_{i=1}^n$;
until *reach the maximum iteration number*;
The aligned model point set $\{\hat{x}_i\}_{i=1}^n$ is given by $\{f(x_i)\}_{i=1}^n$ in the last iteration.

coefficient. Then the matching is performed by a method which encourages geometrically consistent groups.

The two steps of estimating correspondences and transformations are iterated to obtain a reliable result. In this paper, we use a fixed number of iterations, typically 10 but more when the noise is big or when there are a large percentage of outliers contained in the original point sets. We summarize our point set registration method in algorithm 2.

3.2. Application to Image Feature Correspondence

The image feature correspondence task aims to find visual correspondences between two sets of sparse feature points $\{x_i\}_{i=1}^n$ and $\{y_j\}_{j=1}^l$ with corresponding feature descriptors extracted from two input images. In this paper, we assume that the underlying relationship between the input images is nonrigid. Our method for this task is to estimate correspondences by matching feature descriptors using a smooth spatial mapping f. More specifically, we first estimate the initial correspondences based on the feature descriptors, and then use the correspondences to learn a spatial mapping f fitting the inliers by algorithm 1.

Once we have obtained the spatial mapping f, we then have to establish accurate correspondences. We predefine a threshold τ and judge a correspondence (x_i, y_j) to be an inlier provided it satisfies the following condition: $e^{-\|y_j - f(x_i)\|^2/2\sigma^2} > \tau$. We set $\tau = 0.5$ in this paper.

Note that the feature descriptors in the point set registration problem are calculated based on the point sets themselves, and are recalculated in each iteration. However, the descriptors of the feature points here are fixed and calculated from images in advance. Hence the iterative technique for recovering correspondences, estimating spatial mapping, and re-estimating correspondences can not be used here. In practice, we find that our method works well without iteration, since we focus on determining the right correspondences which does not need precise recovery of the underlying transformation, and our approach then plays a role of rejecting outliers.

4. Experimental Results

In order to evaluate the performance of our algorithm, we conducted two types of experiments: i) nonrigid point set registration for 2D shapes; ii) sparse image feature correspondence on 2D images and 3D surfaces.

4.1. Results on Nonrigid Point Set Registration

We tested our method on the same synthesized data as in [6] and [27]. The data consists of two different shape models: a *fish* and a *Chinese character*. For each model, there are five sets of data designed to measure the robustness of registration algorithms under deformation, occlusion, rotation, noise and outliers. In each test, one of the above distortions is applied to a model set to create a target set, and 100 samples are generated for each degradation level. We use the shape context as the feature descriptor to establish initial correspondences. It is easy to make shape context translation and scale invariant, and in some applications, rotation invariance is also required. We use the rotation invariant shape context as in [27].

Fig. 2 shows the registration results of our method on solving different degrees of deformations and occlusions. As shown in the figure, we see that for both datasets with moderate degradation, our method is able to produce an almost perfect alignment. Moreover, the matching performance degrades gradually and gracefully as the degree of degradation in the data increases. Consider the results on the occlusion test in the fifth column, it is interesting that even when the occlusion ratio is 50 percent our method can still achieve a satisfactory registration result. Therefore our method can be used to provide a good initial alignment for more complicated problem-specific registration algorithms.

To provide a quantitative comparison, we report the results of four state-of-the-art algorithms such as shape context [3], TPS-RPM [6], RPM-LNS [27], and CPD [17] which are implemented using publicly available codes. The registration error on a pair of shapes is quantified as the average Euclidean distance between a point in the warped model and the corresponding point in the target. Then the registration performance of each algorithm is compared by the mean and standard deviation of the registration error of all the 100 samples in each distortion level. The statistical results, error means, and standard deviations for each setting are summarized in the last column of Fig. 2. In the deformation test results (e.g., 1st and 3rd rows), five algorithms achieve similar registration performance in both *fish* and *Chinese character* at low deformation levels, and our method generally gives better performance as the degree of

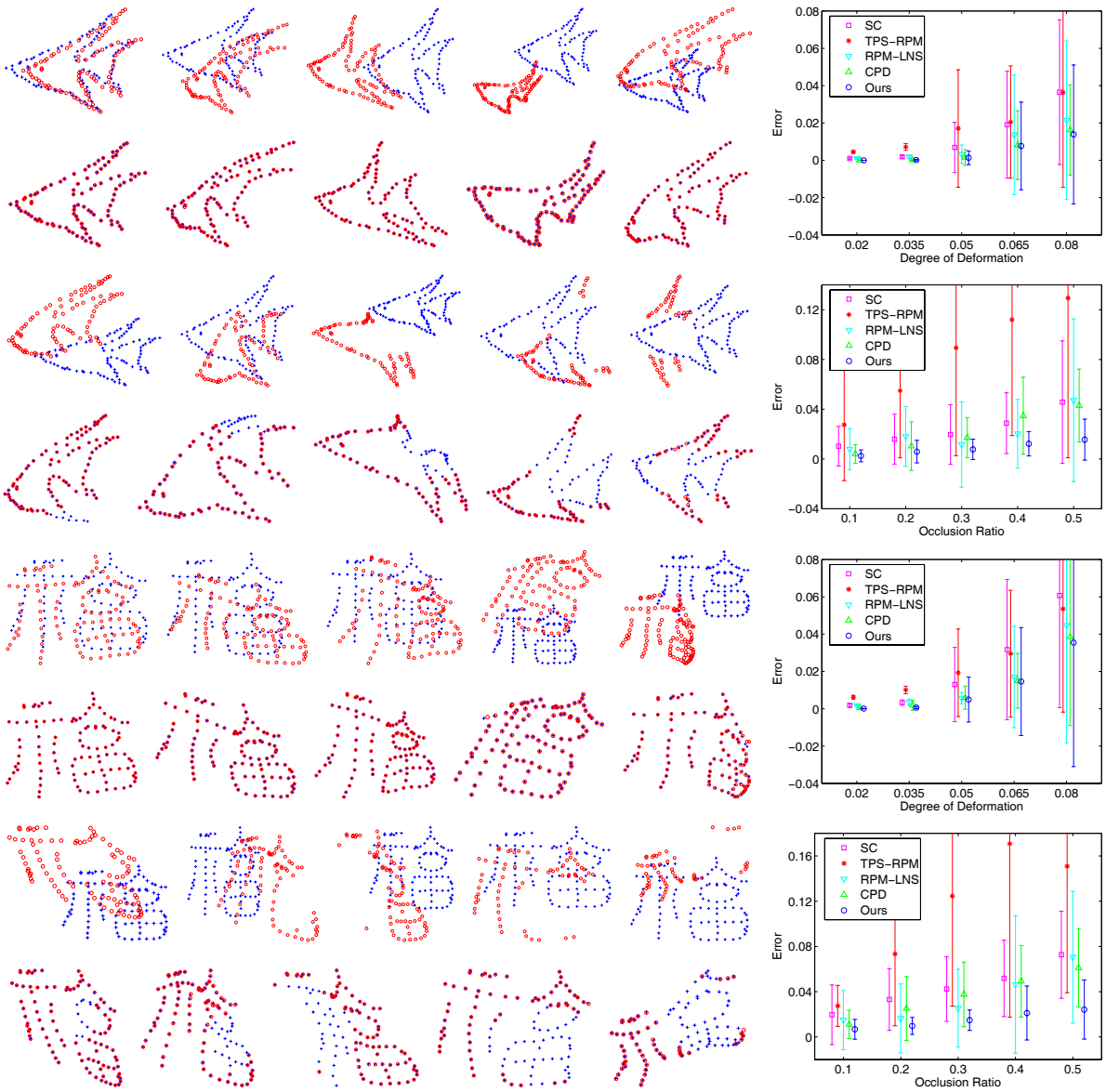

Figure 2. Point set registration results of our method on the *fish* (top) and *Chinese character* (bottom) shapes [6, 27], with deformation and occlusion presented in every two rows. The goal is to align the model point set (blue pluses) onto the target point set (red circles). For each group of experiments, the upper figure is the model and target point sets, and the lower figure is the registration result. From left to right, increasing degree of degradation. The rightmost figures are comparisons of the registration performance of our method with shape context (SC) [3], TPS-RPM [6], RPM-LNS [27] and CPD [17] on the corresponding datasets. The error bars indicate the registration error means and standard deviations over 100 trials.

deformation increases. In the occlusion test results (e.g., 2nd and 4th rows), we observe that our method shows much more robustness compared with the other four algorithms.

More experiments on rotation, noise and outliers are also performed on the two shape models, as shown in Fig. 3. From the results, we again see that our method is able to generate good alignment when the degradation is moderate, and the registration performance degrades gradually and is still acceptable as the amount of degradation increases. Note that our method is not affected by rotation which is not surprising because we use the rotation invariant shape context as the feature descriptor. We also performed experiments on 3D data and got similar results.

In conclusion, our method is efficient for most non-rigid point set registration problems with moderate, and in some cases severe, distortions. It can also be used to provide a good initial alignment for more complicated problem-specific registration algorithms.

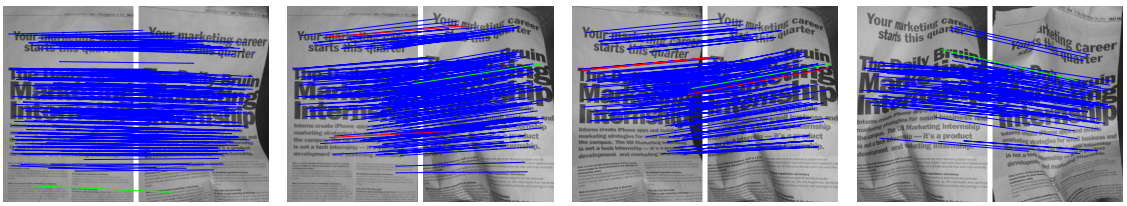

Figure 4. Results of image feature correspondence on 2D image pairs of deformable objects. From left to right, increasing degree of deformation. The inlier percentages in the initial correspondences are 79.61%, 56.57%, 51.84% and 45.71% respectively, and the corresponding precision-recall pairs are (100.00%, 99.73%), (99.06%, 99.53%), (99.09%, 99.35%) and (100.00%, 98.96%) respectively. The lines indicate matching results (blue = true positive, green = false negative, red = false positive). Best viewed in color.

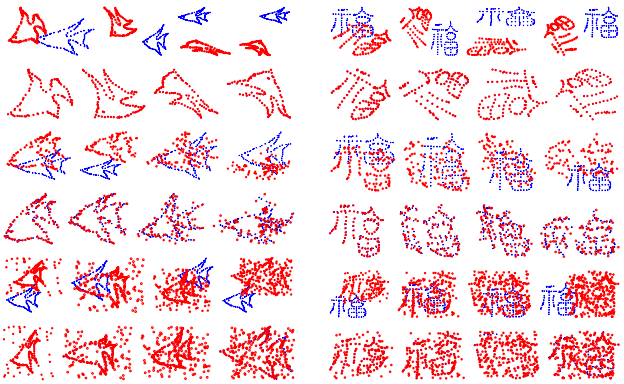

Figure 3. From top to bottom, results on rotation, noise and outliers presented in every two rows. For each group of experiments, the upper figure is the data, and the lower figure is the registration result. From left to right, increasing degree of degradation.

4.2. Results on Image Feature Correspondence

In this section, we perform experiments on real images, and test the performance of our method for sparse image feature correspondence. These images contain deformable objects and consequently the underlying relationships between the images are nonrigid.

Fig. 4 contains a newspaper with different amounts of spatial warps. We aim to establish correspondences between sparse image features in each image pair. In our evaluation, we first extract SIFT [13] feature points in each input image, and estimate the initial correspondences based on the corresponding SIFT descriptors. Our goal is then to reject the outliers contained in the initial correspondences and, at the same time, to keep as many inliers as possible. Performance is characterized by precision and recall.

The results of our method are presented in Fig. 4. For the leftmost pair, the deformation of the newspaper is relatively slight. There are 466 initial correspondences with 95 outliers, and the inlier percentage is about 79.61%. After using our method to establish accurate correspondences, 370 out of the 371 inliers are preserved, and simultaneously all the 95 outliers are rejected. The precision-recall pair is about (100.00%, 99.73%). On the rightmost pair, the deformation is relatively large and the inlier percentage in the initial correspondences is only about 45.71%. In this case, our method still obtains a good precision-recall pair (100.00%, 98.96%). Note that there are still a few false positives and false negatives in the results since we could not precisely estimate the true warp functions between the image pairs in this framework. The average run time of our method on these image pairs is about 0.5 seconds on an Intel Pentium 2.0 GHz PC with Matlab code.

Table 1. Performance comparison on the image pairs in Fig. 4. The values in the first row are the inlier percentages (%), and the pairs are the precision-recall pairs (%).

Inlier	79.61	56.57	51.84	45.71
ICF [12]	(96.05, 98.38)	(83.95, 86.73)	(80.43, 95.48)	(75.42, 92.71)
VFC [26]	(100.00, 97.04)	(98.59, 99.53)	(98.09, 99.35)	(98.94, 97.91)
Ours	(100.00, 99.73)	(99.06, 99.53)	(98.09, 99.35)	(100.00, 98.96)

In addition, we also compared our method to two state-of-the-art methods, such as identifying point correspondences by correspondence function (ICF) [12] and vector field consensus (VFC) [26]. The ICF uses support vector regression to learn a correspondence function pair which maps points in one image to their corresponding points in another, and then reject outliers by the estimated correspondence functions. While the VFC converts the outlier rejection problem into a robust vector field learning problem, and learns a smooth field to fit the potential inliers as well as estimates a consensus inlier set. The results are shown in Table 1. We see that all the three algorithms work well when the deformation contained in the image pair is relatively slight. As the amount of deformation increases, the performance of ICF degenerates rapidly. But VFC and our method seem to be relatively unaffected even when the number of outliers exceeds the number of inliers. Still, our method gains slightly better results compared to VFC.

Our next experiment involves feature point matching on 3D surfaces. We adopt MeshDOG and MeshHOG [25] as the feature point detector and descriptor to determine the

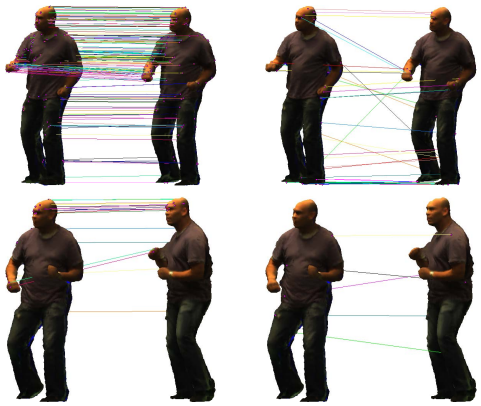

Figure 5. Results on 3D surfaces of deformable objects (INRIA *Dance-1* sequence). Top: results on frames 525 and 527; bottom: frames 530 and 550. For each group, the left pair denotes the identified suspect inliers, and the right pair denotes the removed suspect outliers.

initial correspondences. For the dataset, we use the INRIA *Dance-1* sequence [25], in which each surface is from the same moving person. Note that it is hard to give a quantitative performance comparison since the correctness of a correspondence is hard to decide. So we just schematically show our results in Fig. 5. In the upper group, the two frames are nearby, and the level of deformation is relatively slight. There are 191 initial correspondences, of which 156 are preserved after using our method to establish accurate correspondences. In the lower group, the two frames are far apart, and the level of deformation is relatively large, leading to less good initial correspondences. There are 23 initial correspondences, and 18 of which are preserved by our method.

5. Conclusion

In this paper, we have presented a new approach for nonrigid point set registration. A key characteristic of our approach is the estimation of transformation from correspondences based on a robust estimator named L_2E. The computational complexity of estimation of transformation is linear in the scale of correspondences. We applied our method to sparse image feature correspondence, where the underlying relationship between images is nonrigid. Experiments on a public dataset for nonrigid point registration, 2D and 3D real images for sparse image feature correspondence demonstrate that our approach yields results superior to those of state-of-the-art methods when there is significant noise and/or outliers in the data.

Acknowledgment

The authors gratefully acknowledge the financial support from the National Science Foundation (No. 0917141) and China Scholarship Council (No. 201206160008). Z. Tu is supported by NSF CAREER award IIS-0844566, NSF award IIS-1216528, and NIH R01 MH094343.

References

[1] N. Aronszajn. Theory of Reproducing Kernels. *Trans. of the American Mathematical Society*, 68(3):337–404, 1950.
[2] A. Basu, I. R. Harris, N. L. Hjort, and M. C. Jones. Robust and Efficient Estimation by Minimising A Density Power Divergence. *Biometrika*, 85(3):549–559, 1998.
[3] S. Belongie, J. Malik, and J. Puzicha. Shape Matching and Object Recognition Using Shape Contexts. *TPAMI*, 24(24):509–522, 2002.
[4] P. J. Besl and N. D. McKay. A Method for Registration of 3-D Shapes. *TPAMI*, 14(2):239–256, 1992.
[5] L. G. Brown. A Survey of Image Registration Techniques. *ACM Computing Surveys*, 24(4):325–376, 1992.
[6] H. Chui and A. Rangarajan. A new point matching algorithm for non-rigid registration. *CVIU*, 89:114–141, 2003.
[7] A. W. Fitzgibbon. Robust Registration of 2D and 3D Point Sets. *Image and Vision Computing*, 21:1145–1153, 2003.
[8] P. J. Huber. *Robust Statistics*. John Wiley & Sons, New York, 1981.
[9] B. Jian and B. C. Vemuri. Robust Point Set Registration Using Gaussian Mixture Models. *TPAMI*, 33(8):1633–1645, 2011.
[10] A. E. Johnson and M. Hebert. Using Spin-Images for Efficient Object Recognition in Cluttered 3-D Scenes. *TPAMI*, 21(5):433–449, 1999.
[11] J.-H. Lee and C.-H. Won. Topology Preserving Relaxation Labeling for Nonrigid Point Matching. *TPAMI*, 33(2):427–432, 2011.
[12] X. Li and Z. Hu. Rejecting Mismatches by Correspondence Function. *IJCV*, 89(1):1–17, 2010.
[13] D. Lowe. Distinctive Image Features from Scale-Invariant Keypoints. *IJCV*, 60(2):91–110, 2004.
[14] B. Luo and E. R. Hancock. Structural Graph Matching Using the EM Algorithm and Singular Value Decomposition. *TPAMI*, 23(10):1120–1136, 2001.
[15] J. Ma, J. Zhao, Y. Zhou, and J. Tian. Mismatch Removal via Coherent Spatial Mapping. In *ICIP*, 2012.
[16] C. A. Micchelli and M. Pontil. On Learning Vector-Valued Functions. *Neural Computation*, 17(1):177–204, 2005.
[17] A. Myronenko and X. Song. Point Set Registration: Coherent Point Drift. *TPAMI*, 32(12):2262–2275, 2010.
[18] R. Rifkin, G. Yeo, and T. Poggio. Regularized Least-Squares Classification. In *Advances in Learning Theory: Methods, Model and Applications*, 2003.
[19] S. Rusinkiewicz and M. Levoy. Efficient Variants of the ICP Algorithm. In *3DIM*, pages 145–152, 2001.
[20] D. W. Scott. Parametric Statistical Modeling by Minimum Integrated Square Error. *Technometrics*, 43(3):274–285, 2001.
[21] L. Torresani, V. Kolmogorov, and C. Rother. Feature Correspondence via Graph Matching: Models and Global Optimization. In *ECCV*, pages 596–609, 2008.
[22] Y. Tsin and T. Kanade. A Correlation-Based Approach to Robust Point Set Registration. In *ECCV*, pages 558–569, 2004.
[23] G. Wahba. *Spline Models for Observational Data*. SIAM, Philadelphia, PA, 1990.
[24] A. L. Yuille and N. M. Grzywacz. A Mathematical Analysis of the Motion Coherence Theory. *IJCV*, 3(2):155–175, 1989.
[25] A. Zaharescu, E. Boyer, K. Varanasi, and R. Horaud. Surface Feature Detection and Description with Applications to Mesh Matching. In *CVPR*, pages 373–380, 2009.
[26] J. Zhao, J. Ma, J. Tian, J. Ma, and D. Zhang. A Robust Method for Vector Field Learning with Application to Mismatch Removing. In *CVPR*, pages 2977–2984, 2011.
[27] Y. Zheng and D. Doermann. Robust Point Matching for Nonrigid Shapes by Preserving Local Neighborhood Structures. *TPAMI*, 28(4):643–649, 2006.