Robust L_2E Estimation of Transformation for Non-Rigid Registration

Jiayi Ma, Weichao Qiu, Ji Zhao, Yong Ma, Alan L. Yuille, and Zhuowen Tu

***Abstract*—We introduce a new transformation estimation algorithm using the L_2E estimator, and apply it to non-rigid registration for building robust sparse and dense correspondences. In the sparse point case, our method iteratively recovers the point correspondence and estimates the transformation between two point sets. Feature descriptors such as shape context are used to establish rough correspondence. We then estimate the transformation using our robust algorithm. This 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. In the dense point case for non-rigid image registration, our approach consists of matching both sparsely and densely sampled SIFT features, and it has particular advantages in handling significant scale changes and rotations. The experimental results show that our approach greatly outperforms state-of-the-art methods, particularly when the data contains severe outliers.**

***Index Terms*—L_2E estimator, registration, dense correspondence, regularization, outlier, non-rigid.**

I. Introduction

Building "correct" registration, alignment or correspondence is a fundamental problem in computer vision, medical image analysis, and pattern recognition [1], [2], [3], [4]. Many tasks in these fields – such as stereo matching, shape matching, image registration and content-based image retrieval – can be formulated as matching two sets of feature points, more specifically, as a point matching problem [1]. The feature points in these tasks are typically the locations of sparse or dense 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 sparse correspondence or dense correspondence, depending on the specific application. The task of sparse correspondence aims to match sparsely distributed points, such as matching contour points in shape matching. It also frequently arises in image matching, for example, for images taken from different viewpoints while the scene consists of mostly rigid objects, sparse feature matching methods have proven highly effective. The problem of dense correspondence is typically associated with image alignment/registration, which aims to overlaying two or more images with shared content, either at the pixel level (e.g., stereo matching [5] and optical flow [6], [7]) or the object/scene level (e.g., pictorial structure model [8] and SIFT flow [4]). It is a crucial step in all image analysis tasks in which the final information is gained from the combination of various data sources, e.g., in image fusion, change detection, multichannel image restoration, as well as object/scene recognition.

The registration problem can also be categorized into rigid or non-rigid registration depending on the form of the data. Rigid registration, which only involves a small number of parameters, is relatively easy and has been widely studied [1], [2], [9], [10], [11]. By contrast, non-rigid registration is more difficult because the underlying non-rigid transformations are often unknown, complex, and hard to model [3]. But non-rigid 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 non-rigid case and introduce a new algorithm for both sparse point set registration and dense image correspondence. To illustrate our main idea, we take the point set registration problem as an example. 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 [2], [12], [3]. 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 as happens 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 key issue of our approach is to robustly estimate the transformations from the correspondences using a robust estimator named the L_2-Minimizing Estimate (L_2E) [13], [14]. 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. The correspondence set typically includes a large set of correspondence, and many of which are wrong. In the second step, we estimate the transformation using the robust estimator L_2E. This estimator enable us to deal with

J. Ma and Y. Ma are with the Electronic Information School, Wuhan University, Wuhan, 430072, China. E-mail: jyma2010@gmail.com; mayong@whu.edu.cn.

W. Qiu and A. Yuille are with the Department of Statistics, UCLA, Los Angeles, CA, 90095. E-mail: qiuwch@gmail.com; yuille@stat.ucla.edu.

J. Zhao is with the Samsung Advanced Institute of Technology, Beijing, 100027, China. E-mail: zhaoji84@gmail.com.

Z. Tu is with the Department of Cognitive Science, UCSD, La Jolla, CA, 92697. E-mail: zhuowen.tu@gmail.com.

the noise and outliers in the correspondences. The non-rigid transformation is modeled in a functional space, called the reproducing kernel Hilbert space (RKHS) [15], in which the transformation function has an explicit kernel representation.

We also apply our robust estimation of transformation to the dense image correspondence/registration problem, which aligns the images of the same scene taken under different imaging conditions. The majority of the methods are based on sparse feature correspondences and consists of four steps such as feature detection, feature matching, transformation estimation (rigid/non-rigid), and image resampling and transformation [16]. Recently, some new registration methods have also been developed, where the goal is to align different instances of the same object/scene category, such as the pictorial structure model for object recognition [8] and the SIFT flow for scene alignment [4]. Rather than operating on sparse feature correspondences, these methods match dense sampled features to obtain a higher level of image alignment (e.g., object/scene level). Here we generalize our robust algorithm, and introduce a novel image registration method which consists of matching both sparsely and densely sampled features. Our formulation contains two coupling variables: the non-rigid geometric transformation and the discrete dense flow field, where the former corresponds to the sparse feature matching and the latter corresponds to the dense feature matching. Ideally, the two variables are equivalent. We alternatively solve for one variable under the assumption that the other is known.

The main contributions of our work include: (i) the proposal of a new robust algorithm to estimate a spatial transformation/mapping from correspondences with noise and outliers; (ii) application of this algorithm to non-rigid sparse correspondence such as point set registration and sparse image feature correspondence; (iii) further application of this robust algorithm to non-rigid dense correspondence such as image registration, and it can handle significantly changes in scale and rotation; (iv) both the sparse and dense matching algorithms are demonstrably effective and outperform state-of-the-art methods. This article is an extension of our earlier work [17]. The primary new contributions are an expanded derivation and discussion of the robust L_2E estimator (Section III) and a new algorithm for non-rigid image registration based on the proposed robust algorithm (Section V).

The rest of the paper is organized as follows. Section 2 describes relevant previous work, followed by our robust L_2E algorithm for estimating transformations from point correspondences with unknown outliers in Section 3. In Section 4, we apply our robust algorithm to non-rigid sparse correspondence such as point set registration and sparse image feature correspondence. We further apply our robust algorithm to non-rigid dense correspondence, i.e., image registration, in Section 5. Section 6 demonstrates the experimental results of our sparse and dense matching algorithms on both 2D and 3D, synthetic and real data. Finally, we conclude this paper in Section 7.

II. Related Work

We apply our robust algorithm to sparse and dense correspondence, e.g., non-rigid point set registration and image registration. It is beyond the scope of this paper to give a thorough review on these topics. In this section we briefly overview some works that are most relevant to our approach.

A. Point Set Registration Methods

The iterated closest point (ICP) algorithm [2] 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 spatial transformation. Moreover, it requires a good initial transformation that places the two data sets in approximate registration. Belongie *et al.* [12] introduced a method for shape 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. These methods however ignore robustness when they recover the transformation from the correspondence.

For robust point feature matching, the random sample consensus (RANSAC) [18] is a widely used algorithm in computer vision. It uses a hypothesize-and-verify and tries to get as small an outlier-free subset as feasible to estimate a given parametric model by resampling. RANSAC has several variants such as MLESAC [19], LO-RANSAC [20] and PROSAC [21]. Although these methods are very successful in many situations they have had limited success if the underlying spatial transformation between point features are non-parametric, for example if the real correspondence is non-rigid. In related work, Chui and Rangarajan [3] established a general framework for estimating correspondence and transformations for non-rigid point matching. They modeled the transformation as a thin-plate spline and did robust point matching by an algorithm (TPS-RPM) which involved deterministic annealing [22] and soft-assignment [23]. Alternatively, the coherence point drift (CPD) algorithm [24] uses Gaussian radial basis functions instead of thin-plate splines. These algorithms are robust compared to ICP in the non-rigid case, but the joint estimation of correspondence and transformation increases the algorithm complexity.

Another interesting point matching approach is the kernel correlation (KC) based method [25]. The cost function of KC is proportional to the correlation of two kernel density estimates. Jian and Vemuri [26] presented a unified framework based on a divergence family that allows for interpreting several of the existing point set registration methods, e.g. [27], [3], [24], as special cases of the divergence family. In particular, the L_2 distance between Gaussians is strongly related to the L_2E estimator used in our method. Zheng and Doermann [28] 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, which was later generalized in [29] by introducing an optimal compatibility coefficient for the relaxation labeling process to solve a non-rigid point matching problem. Huang *et al.* [30] proposed to implicitly embed the shapes of interest in a higher-dimensional space of distance transforms, and align them similar to non-rigid image registration algorithms. The method performs well on relatively simple data sets. Other related work includes the EM-like algorithm for matching rigid and articulated shapes [31], as well as the graph matching approach for establishing feature correspondences [32].

B. Image Registration Methods

Image registration, as was mentioned above, is widely used in computer vision, medical imaging, remote sensing, etc. An early example of a widely-used image registration algorithm is the optical flow technique [6], [7]. It computes a dense correspondence field by directly minimizing pixel-to-pixel dissimilarities, and hence tends to operate on very similar images, e.g., two adjacent frames in a video sequence. Typical assumptions in optical flow algorithms include brightness constancy and piecewise smoothness of the pixel displacement field.

Due to changes of lighting, perspective and noise, the pixel values are often not reliable for registration [33]. Nevertheless, the development of various local invariant features has brought about significant progress in this area [34], [35]. The feature-based approaches are to some extent similar to the point set registration methods. They work by extracting a set of sparse features and then matching them to each other. The spatial form of the correspondence then can be described by parametric models such as affine and homography, or non-parametric models such as radial basis functions, thin-plate splines and elastic transforms, relating pixel coordinates in one image to pixel coordinates in another. This type of methods has the advantage of being more robust to typical appearance changes and scene movements, and is potentially faster, if implemented the right way. However, they are often designed for matching rigid scenes (e.g., in stereo matching and image stitching) or small non-rigid motions (e.g., in medical imaging), but less effective for handling objects and scenes with significant non-rigidity. This can be explained since the dense correspondence here is interpolated from sparse matching rather then through pixel-wise correspondence, which may be problematic when the real correspondence is non-rigid and the transform model is not known apriori. For comprehensive reviews of this class of methods please refer to [16], [36], [37].

Recently, Liu *et al.* [4] proposed a SIFT flow algorithm for non-rigid matching of highly different scenes. Instead of matching brightness of pixels in optical flow algorithms, the SIFT flow algorithm matches densely sampled SIFT features between two images. It has demonstrated impressive dense, pixel-wise correspondence results, however, it is not robust to significantly changes in scale and rotation. HaCohen *et al.* [38] proposed a non-rigid dense correspondence method (NRDC) which combines dense local matching with robustness to outliers based on Generalized PatchMatch [39]. Moreover, it can address large scale change and rotation. Ma *et al.* [40] proposed a regularized Gaussian field criterion for multimodal image registration, such as visible and infrared face images. In this paper, we introduce a novel dense correspondence method matching both sparsely and densely sampled SIFT features based on the robust L_2E estimator and SIFT flow. We show that our method significantly outperforms both SIFT flow and NRDC in a set of quantitative evaluation. Besides, Vemuri *et al.* [41] and Liu *et al.* [42] proposed a robust multimodal image registration method based on matching dominant local frequency image representations, and they are the first who applied the L_2E estimator in the image registration problem.

III. Estimating Transformation from Correspondences by L_2E

Given a set of point correspondences $S = \{(\mathbf{x}_i, \mathbf{y}_i)\}_{i=1}^n$, which are typically perturbed by noise and by outlier points which undergo different transformations, the goal is to estimate a transformation $\mathbf{f} : \mathbf{y}_i = \mathbf{f}(\mathbf{x}_i)$ 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 $(\mathbf{x}_i, \mathbf{y}_i)$ satisfies $\mathbf{y}_i - \mathbf{f}(\mathbf{x}_i) \sim N(\mathbf{0}, \sigma^2\mathbf{I})$, where $\mathbf{I}$ is an identity matrix of size $d \times d$, with d being the dimension of the point. The data $\{\mathbf{y}_i - \mathbf{f}(\mathbf{x}_i)\}_{i=1}^n$ can then be thought of as a sample set from a multivariate normal density $N(\mathbf{0}, \sigma^2\mathbf{I})$ 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 $\mathbf{f}$ for the inlier set. Next, we introduce a robust estimator named L_2-minimizing estimate (L_2E) which we use to estimate the transformation $\mathbf{f}$.

A. 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 the point matching problem, it is generally accepted that incorrect matches, or outliers, cannot be avoided in the matching process where only local feature descriptors are compared. In this case, a robust estimator of the transformation $\mathbf{f}$ is desirable 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 [43]. In this paper, we use the second method and adopt the L_2E estimator [13], [14], 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.

For the parametric setting with a density model $p(z|\theta)$, consider minimizing an estimate of L_2 distance with respect

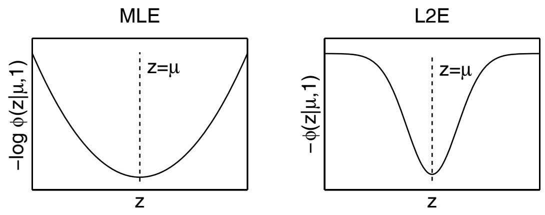


Fig. 1: The penalty curves of MLE (left) and L_2E (right) for an one-dimensional Gaussian model.

to θ:

$$
\begin{aligned}
\hat{\theta} &= \arg\min_{\theta} \int [p(z|\theta) - p(z|\theta_0)]^2 dz \\
&= \arg\min_{\theta} \Big[\int p(z|\theta)^2 dz - 2\int p(z|\theta) p(z|\theta_0) dz \\
&\qquad + \int p(z|\theta_0)^2 dz \Big] \\
&= \arg\min_{\theta} \Big[\int p(z|\theta)^2 dz - 2E[p(z|\theta)] \Big], \qquad (1)
\end{aligned}
$$

where the true parameter θ_0 is unknown, and we omit the (unknown) constant $\int p(z|\theta_0)^2 dz$. Based on the L_2 distance, the L_2E estimator for model $p(z|\theta)$ then recommends estimating the parameter θ by minimizing the criterion:

$$
\hat{\theta}_{L_2E} = \arg\min_{\theta} \left[\int p(z|\theta)^2 dz - \frac{2}{n} \sum_{i=1}^{n} p(z_i|\theta) \right]. \qquad (2)
$$

To get some intuition for why L_2E is robust, we consider a simple one-dimensional Gaussian model $p(z|\theta) = \phi(z|\mu, 1)$, where $\hat{\mu}_{MLE} = \arg\max_\mu \sum_{i=1}^{n} \log \phi(z_i|\mu, 1)$ and $\hat{\mu}_{L_2E} = \arg\min_\mu \left[\frac{1}{2\sqrt{\pi}} - \frac{2}{n} \sum_{i=1}^{n} \phi(z_i|\mu, 1) \right]$. The penalty curves of MLE (i.e. $-\log \phi(z|\mu, 1)$) and L_2E (i.e., $-\phi(z|\mu, 1)$) are shown in Fig. 1. From the figure, we see that the penalty curve of MLE is quadratic, and then it is reluctant to assign low probability to any points (which becomes infinite as $\phi(z_i|\mu, 1)$ tends to 0), including outliers, and hence tends to be biased by outliers. By contrast, the penalty of L_2E can be approximately seen as a truncated form of the MLE penalty, and then it can assign low probabilities to many points, hopefully to the outliers, without paying too high a penalty. Moreover, unlike a truncated penalty, the L_2E functional is derivable and it has the advantage of computational convenience.

To further demonstrate the robustness of L_2E, we present a line-fitting example which contrasts the behavior of MLE and L_2E, see Fig. 2. 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 $\hat{\alpha}_{L_2E} = \arg\min_\alpha \left[\frac{1}{2\sqrt{\pi}} - \frac{2}{n} \sum_{i=1}^{n} \phi(y_i - \alpha x_i|0, 1) \right]$. As shown in Fig. 2, 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 appropriate because, in this case, the contaminated data also comes from the same linear parametric model with slope $\alpha = 2$, e.g., $y = 2x$.

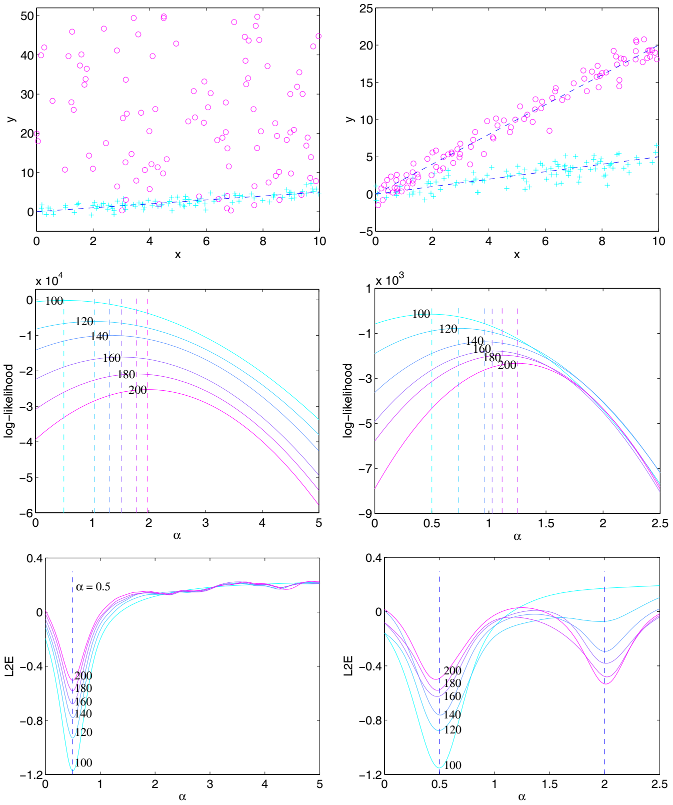


Fig. 2: Comparison between L_2E and MLE for linear regression as the number of outliers varies. Top row: data samples, where the inliers are shown by cyan pluses, and the outliers by magenta circles. The goal is to estimate the slope α of the line model $y = \alpha x$. We vary the number of data samples $n = 100, 120, \cdots, 200$, 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.

We now apply the L_2E formulation in (2) to the point matching problem, assuming that the noise of the inliers is given by a normal distribution, and obtain the following functional criterion:

$$
L_2E(\mathbf{f}, \sigma^2) = \frac{1}{2^d (\pi\sigma)^{d/2}} - \frac{2}{n} \sum_{i=1}^{n} \Phi\left(\mathbf{y}_i - \mathbf{f}(\mathbf{x}_i) | \mathbf{0}, \sigma^2 \mathbf{I}\right). \qquad (3)
$$

where Φ is a multi-dimensional Gaussian function. We model the non-rigid transformation $\mathbf{f}$ by requiring it to lie within a specific functional space, namely a reproducing kernel Hilbert space (RKHS) [15], [44], [45]. Note that other parameterized transformation models, for example, thin-plate splines (TPS) [46], [47], [48], can also be easily incorporated into our formulation.

We define an RKHS $\mathcal{H}$ by a positive definite matrix-valued kernel $\mathbf{\Gamma}(\mathbf{x}, \mathbf{x}') : \mathbb{R}^d \times \mathbb{R}^d \rightarrow \mathbb{R}^{d \times d}$. The optimal transformation $\mathbf{f}$ which minimizes the L_2E functional (3) then takes the form $\mathbf{f}(\mathbf{x}) = \sum_{i=1}^{n} \mathbf{\Gamma}(\mathbf{x}, \mathbf{x}_i)\mathbf{c}_i$ [45], [49], where the coefficient $\mathbf{c}_i$ 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 $\mathbf{c}_i$. 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 $\{\tilde{\mathbf{x}}_i\}_{i=1}^{m}$ to have nonzero coefficients in the expansion of the solution. This follows [50], [51] 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

$$\mathbf{f}(\mathbf{x}) = \sum_{i=1}^{m} \mathbf{\Gamma}(\mathbf{x}, \tilde{\mathbf{x}}_i)\mathbf{c}_i. \qquad (4)$$

The chosen point set $\{\tilde{\mathbf{x}}_i\}_{i=1}^{m}$ are somewhat analogous to "control points" [1]. By including a regularization term for imposing smooth constraint on the transformation, the L_2E functional (3) becomes:

$$L_2E(\mathbf{f}, \sigma^2) = \frac{1}{2^d(\pi\sigma)^{d/2}} - \frac{2}{n}\sum_{i=1}^{n} \frac{1}{(2\pi\sigma^2)^{d/2}} e^{-\frac{\left\|\mathbf{y}_i - \sum_{j=1}^{m} \mathbf{\Gamma}(\mathbf{x}_i, \tilde{\mathbf{x}}_j)\mathbf{c}_j\right\|^2}{2\sigma^2}} + \lambda\|\mathbf{f}\|_{\Gamma}^2, \qquad (5)$$

where $\lambda > 0$ controls the strength of regularization, and the stabilizer $\|\mathbf{f}\|_{\Gamma}^2$ is defined by an inner product, e.g., $\|\mathbf{f}\|_{\Gamma}^2 = \langle \mathbf{f}, \mathbf{f} \rangle_{\Gamma}$. By choosing a diagonal decomposable kernel [49]: $\mathbf{\Gamma}(\mathbf{x}_i, \mathbf{x}_j) = \kappa(\mathbf{x}_i, \mathbf{x}_j) \cdot \mathbf{I} = e^{-\beta\|\mathbf{x}_i - \mathbf{x}_j\|^2}\mathbf{I}$ with β determining the width of the range of interaction between samples (i.e. neighborhood size), the L_2E functional (5) may be conveniently expressed in the following matrix form:

$$L_2E(\mathbf{C}, \sigma^2) = \frac{1}{2^d(\pi\sigma)^{d/2}} - \frac{2}{n}\sum_{i=1}^{n} \frac{1}{(2\pi\sigma^2)^{d/2}} e^{-\frac{\left\|\mathbf{y}_i^{\mathrm{T}} - \mathbf{U}_{i,\cdot}\mathbf{C}\right\|^2}{2\sigma^2}} + \lambda \, \mathrm{tr}(\mathbf{C}^{\mathrm{T}}\mathbf{\Gamma}\mathbf{C}), \qquad (6)$$

where kernel matrix $\mathbf{\Gamma} \in \mathbb{R}^{m \times m}$ is called the Gram matrix with $\mathbf{\Gamma}_{ij} = \kappa(\tilde{\mathbf{x}}_i, \tilde{\mathbf{x}}_j) = e^{-\beta\|\tilde{\mathbf{x}}_i - \tilde{\mathbf{x}}_j\|^2}$, $\mathbf{U} \in \mathbb{R}^{n \times m}$ with $\mathbf{U}_{ij} = \kappa(\mathbf{x}_i, \tilde{\mathbf{x}}_j) = e^{-\beta\|\mathbf{x}_i - \tilde{\mathbf{x}}_j\|^2}$, $\mathbf{U}_{i,\cdot}$ denotes the i-th row of the matrix $\mathbf{U}$, $\mathbf{C} = (\mathbf{c}_1, \cdots, \mathbf{c}_m)^{\mathrm{T}}$ is the coefficient matrix of size $m \times d$, and tr$(\cdot)$ denotes the trace.

B. Estimation of the Transformation

Estimating the transformation requires taking the derivative of the L_2E cost function, see equation (6), with respect to the coefficient matrix $\mathbf{C}$, which is given by:

$$\frac{\partial L_2E}{\partial \mathbf{C}} = \frac{2\mathbf{U}^{\mathrm{T}}[\mathbf{P} \circ (\mathbf{Q} \otimes \mathbf{1}_{1\times d})]}{n\sigma^2(2\pi\sigma^2)^{d/2}} + 2\lambda\mathbf{\Gamma}\mathbf{C}, \qquad (7)$$

where $\mathbf{P} = \mathbf{U}\mathbf{C} - \mathbf{Y}$ and $\mathbf{Y} = (\mathbf{y}_1, \cdots, \mathbf{y}_n)^{\mathrm{T}}$ are matrices of size $n \times d$, $\mathbf{Q} = \exp\{\mathrm{diag}(\mathbf{P}\mathbf{P}^{\mathrm{T}})/2\sigma^2\}$ is an $n \times 1$ dimensional vector, 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, for example $(A \circ B)_{ij} = A_{ij} \cdot B_{ij}$, and $\otimes$ denotes the Kronecker product, for example, for $A \in \mathbb{R}^{m\times n}$ and $B \in \mathbb{R}^{p\times q}$, $A \otimes B = \begin{bmatrix} a_{11}B & \cdots & a_{1n}B \\ \vdots & \ddots & \vdots \\ a_{m1}B & \cdots & a_{mn}B \end{bmatrix}$.

Algorithm 1: Estimate of Transformation from Correspondences

```
Input: Correspondence set $S = \{(\mathbf{x}_i, \mathbf{y}_i)\}_{i=1}^{n}$,
       parameters $\gamma, \beta, \lambda$
Output: Optimal transformation $\mathbf{f}$
Construct Gram matrix $\mathbf{\Gamma}$ and matrix $\mathbf{U}$;
Initialize parameter $\sigma^2$ and $\mathbf{C}$;
Deterministic annealing:
    Using the gradient (7), optimize the objective function
    (6) by a numerical technique (e.g., the quasi-Newton
    algorithm with $\mathbf{C}$ as the old value);
    Update the parameter $\mathbf{C} \leftarrow \arg\min_{\mathbf{C}} L_2E(\mathbf{C}, \sigma^2)$;
    Anneal $\sigma^2 = \gamma\sigma^2$;
The transformation $f$ is determined by equation (4).
```

By using the derivative in equation (7), 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 (6) 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. Note that our proposed algorithm can be easily extended to the rigid or affine model, by using a rotation matrix $\mathbf{R}$ or a nonsingular matrix $\mathbf{A}$ together with a translation vector $\mathbf{t}$, i.e., $\mathbf{f}(\mathbf{x}) = \mathbf{R}\mathbf{x} + \mathbf{t}$ or $\mathbf{f}(\mathbf{x}) = \mathbf{A}\mathbf{x} + \mathbf{t}$. Here we omit the details of the derivation and focus on the non-rigid case only.

Computational complexity. By examining equations (6) and (7), 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 (e.g., the Matlab function *fminunc*), 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 $\mathbf{f}$ in equation (4) 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.

C. 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 $\mathbf{f}$ as the initial position plus a displacement function $\mathbf{v}$: $\mathbf{f}(\mathbf{x}) = \mathbf{x} + \mathbf{v}(\mathbf{x})$ [24], and solve for $\mathbf{v}$ instead of $\mathbf{f}$. This can be achieved simply by setting the output $\mathbf{y}_i$ to be $\mathbf{y}_i - \mathbf{x}_i$.

Parameter settings. Our method uses deterministic annealing to deal with the non-convexity, and so it will fail if the annealing fails. The deterministic annealing is more likely to work if we anneal slowly, and the annealing rate is controlled by parameter γ. Without considering the efficiency issue, we can always set it as large as possible (e.g., close to 1), and then use more iterations to produce satisfied results. Two other parameters are β and λ, which control the influence of the smoothness constraint on the transformation $\mathbf{f}$. Parameter β determines how wide the range of interaction between samples. Since we have used data normalization so that the points in the two sets both have zero mean and unit variance, the optimal value of parameter β in general will be similar on different examples. Parameter λ controls the trade-off between the closeness to the data and the smoothness of the solution. It is mainly influenced by the degrees of data degradation. We set $\gamma = 0.5$, $\beta = 0.8$ and $\lambda = 0.1$ throughout this paper. Finally, the parameter σ^2 and $\mathbf{C}$ in line 2 of Algorithm 1 were initialized to 0.05 and 0 respectively.

IV. Application to Non-Rigid Sparse Correspondence

In this section, we apply our robust algorithm to establishment of non-rigid sparse correspondence. Two tasks are considered such as non-rigid point set registration and non-rigid image feature correspondence.

A. Non-Rigid Point Set Registration

Point set registration aims to align two point sets $\{\mathbf{x}_i\}_{i=1}^n$ (the model point set) and $\{\mathbf{y}_j\}_{j=1}^l$ (the target point set). Typically, in the non-rigid case, it requires estimating a non-rigid transformation $\mathbf{f}$ 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. Since our method 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 [12] 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 [52] can be used as a feature descriptor, where the local similarity is measured by an improved correlation coefficient. Then the matching is performed by a method which encourages geometrically consistent groups.

Algorithm 2: Non-Rigid Point Set Registration

```
Input: Two point sets $\{\mathbf{x}_i\}_{i=1}^n$, $\{\mathbf{y}_j\}_{j=1}^l$
Output: Aligned model point set $\{\hat{\mathbf{x}}_i\}_{i=1}^n$
Compute feature descriptors for the target point set $\{\mathbf{y}_j\}_{j=1}^l$;
repeat
    Compute feature descriptors for the model point set $\{\mathbf{x}_i\}_{i=1}^n$;
    Estimate the initial correspondences based on the feature descriptors of two point sets;
    Solve the transformation $\mathbf{f}$ warping the model point set to the target point set using Algorithm 1;
    Update model point set $\{\mathbf{x}_i\}_{i=1}^n \leftarrow \{\mathbf{f}(\mathbf{x}_i)\}_{i=1}^n$;
until reach the maximum iteration number;
The aligned model point set $\{\hat{\mathbf{x}}_i\}_{i=1}^n$ is given by $\{\mathbf{f}(\mathbf{x}_i)\}_{i=1}^n$ in the last iteration.
```

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.

B. Non-Rigid Image Feature Correspondence

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

Once we have obtained the spatial mapping $\mathbf{f}$, we then have to establish accurate correspondences. We predefine a threshold τ and judge a correspondence $(\mathbf{x}_i, \mathbf{y}_j)$ to be an inlier provided it satisfies the following condition: $e^{-\|\mathbf{y}_j - \mathbf{f}(\mathbf{x}_i)\|^2/2\sigma^2} > \tau$. 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 due to the following two reasons: i) the initial correspondences between the given two feature point sets in general contain most of the ground truth correspondences; ii) we focus on determining the right correspondences which does not need precise recovery of the underlying transformation. Our approach then plays a role of rejecting outliers.

V. Application to Non-Rigid Dense Correspondence

We now focus on the image registration/dense matching problem. Given two input images, e.g. a model image and a target image, the goal is to establish dense correspondences between the two images. In this paper we assume that the underlying relationship between the input images is non-rigid. Next, we introduce a novel dense matching method which consists of matching both sparsely and densely sampled SIFT features.

A. Problem Formulation

To align two images, we extract a set of sparse feature correspondences $\{(\mathbf{x}_i, \mathbf{y}_i)\}_{i=1}^n$, e.g. SIFT feature correspondences [34], with $\mathbf{x}_i, \mathbf{y}_i \in \mathbb{R}^2$ being the spatial locations of the feature points in the model and target images respectively. Generally, the correspondences contain some unknown outliers, and the robust L_2E estimator can be used to recover a transformation $\mathbf{f}$ according to Algorithm 1. This sparse feature detection and matching has been a standard approach to image registration of the same scene [16]. However, for image pairs containing unknown non-rigid motions, this procedure in general cannot obtain accurate dense correspondences. To address this problem, here we also match densely sampled SIFT features based on the SIFT flow algorithm [4].

We define the transformation between the sparse correspondences as $\mathbf{f}(\mathbf{x}) = \mathbf{x} + \mathbf{v}(\mathbf{x})$, where $\mathbf{v}$ is the displacement function with the form in equation (4). To match dense SIFT features, we introduce a dense discrete flow $\mathbf{w}$: let $\mathbf{p}$ be the grid coordinate of the model image, $\mathbf{w}(\mathbf{p})$ then is the displacement vector at $\mathbf{p}$, i.e., the point $\mathbf{p}$ in the model image corresponds to the point $\mathbf{p} + \mathbf{w}(\mathbf{p})$ in the target image. In addition, $\mathbf{w}$ is only allowed to have integral entries. Let s_1 and s_2 be the per-pixel SIFT features for two images. The set $\boldsymbol{\epsilon}$ contains all the spatial neighborhoods (a four-neighbor system is used). Denote $\mathbf{w}(\mathbf{p}) = (\mathbf{w}^1(\mathbf{p}), \mathbf{w}^2(\mathbf{p}))^{\mathrm{T}}$, and l the total number of evaluation pixels. The cost function in our dense correspondence problem is then defined as:

$$\begin{aligned} E(\mathbf{v}, \mathbf{w}, \sigma^2) = \frac{1}{4\pi\sigma} - \frac{1}{n\pi\sigma^2} \sum_{i=1}^{n} e^{-\frac{\|\mathbf{y}_i - \mathbf{x}_i - \mathbf{v}(\mathbf{x}_i)\|^2}{2\sigma^2}} + \lambda \|\mathbf{v}\|_{\Gamma}^2 \\ + \frac{\delta}{l} \sum_{\mathbf{p}} \min(\|s_1(\mathbf{p}) - s_2(\mathbf{p} + \mathbf{w}(\mathbf{p}))\|_1, t) \\ + \frac{\delta}{l} \sum_{(\mathbf{p},\mathbf{q}) \in \boldsymbol{\epsilon}} \sum_{i=1}^{2} \min(\alpha |\mathbf{w}^i(\mathbf{p}) - \mathbf{w}^i(\mathbf{q})|, u) \\ + \frac{\delta}{l} \sum_{\mathbf{p}} \eta \|\mathbf{v}(\mathbf{p}) - \mathbf{w}(\mathbf{p})\|^2. \end{aligned} \tag{8}$$

We briefly go through all the components of the cost function. The first three terms are from the L_2E estimator as in equation (5), which performs a robust estimation of the displacement function $\mathbf{v}$ from sparse correspondences. The fourth, fifth and sixth terms correspond to the *data term*, *smoothness term* and *small displacement term* of the SIFT flow algorithm, which performs dense matching between two images. The only difference is that the last term in the SIFT flow algorithm is $\sum_{\mathbf{p}} \|\mathbf{w}(\mathbf{p})\|_1$, rather than $\sum_{\mathbf{p}} \|\mathbf{v}(\mathbf{p}) - \mathbf{w}(\mathbf{p})\|^2$ in our algorithm. Here this term acts as a bridge between the sparse and the dense feature matchings; it constraints the displacement function $\mathbf{v}$ and flow $\mathbf{w}$ to be consistent.

B. Optimization

There are two unknown variables in the cost function–the displacement function $\mathbf{v}$ and the flow $\mathbf{w}$. While solving for either variable without information regarding the other is quite difficult, an interesting fact is that solving for one variable once the other is known is much simpler than solving the original, coupled problem.

Now we first consider the terms of cost function (8) that related to $\mathbf{v}$, which can be expressed in the following matrix form:

$$\begin{aligned} E(\mathbf{C}, \sigma^2) = \frac{1}{4\pi\sigma} - \frac{1}{n\pi\sigma^2} \sum_{i=1}^{n} e^{-\frac{\|\mathbf{y}_i^{\mathrm{T}} - \mathbf{x}_i^{\mathrm{T}} - \mathbf{U}_{i,\cdot}\mathbf{C}\|^2}{2\sigma^2}} \\ + \lambda \mathrm{tr}(\mathbf{C}^{\mathrm{T}} \boldsymbol{\Gamma} \mathbf{C}) + \frac{\delta\eta}{l} \|\mathbf{V}\mathbf{C} - \mathbf{W}\|_F^2, \end{aligned} \tag{9}$$

where the matrices $\boldsymbol{\Gamma}$, $\mathbf{U}$ and $\mathbf{C}$ are defined as the same as in equation (6), matrix $\mathbf{V} \in \mathbb{R}^{l \times m}$ with $\mathbf{V}_{ij} = \kappa(\mathbf{p}_i, \tilde{\mathbf{x}}_j) = e^{-\beta\|\mathbf{p}_i - \tilde{\mathbf{x}}_j\|^2}$, $\mathbf{W} = (\mathbf{w}_1, \cdots, \mathbf{w}_l)^{\mathrm{T}}$ is the flow field of size $l \times 2$, and $\|\cdot\|_F$ denotes the Frobenius norm.

As described in Algorithm 1, the cost function (9) can be minimized by using efficient gradient-based numerical optimization techniques combining with deterministic annealing, where the derivative of the objective function with respect to the coefficient matrix $\mathbf{C}$ is given by:

$$\frac{\partial E(\mathbf{C}, \sigma^2)}{\partial \mathbf{C}} = \frac{\mathbf{U}^{\mathrm{T}}[\mathbf{P} \circ (\mathbf{Q} \otimes \mathbf{1}_{1\times 2})]}{n\pi\sigma^4} + 2\lambda\boldsymbol{\Gamma}\mathbf{C} + \frac{2\delta\eta}{l}\mathbf{V}^{\mathrm{T}}(\mathbf{V}\mathbf{C} - \mathbf{W}), \tag{10}$$

where $\mathbf{P} = \mathbf{X} + \mathbf{U}\mathbf{C} - \mathbf{Y}$ and $\mathbf{Q} = \exp\{\mathrm{diag}(\mathbf{P}\mathbf{P}^{\mathrm{T}})/2\sigma^2\}$ are defined similar to those in equation (7).

Next we consider the terms of cost function (8) related to $\mathbf{w}$, which involves the last three terms

$$\begin{aligned} E(\mathbf{w}) = \sum_{\mathbf{p}} \min(\|s_1(\mathbf{p}) - s_2(\mathbf{p} + \mathbf{w}(\mathbf{p}))\|_1, t) \\ + \sum_{(\mathbf{p},\mathbf{q}) \in \boldsymbol{\epsilon}} \sum_{i=1}^{2} \min(\alpha |\mathbf{w}^i(\mathbf{p}) - \mathbf{w}^i(\mathbf{q})|, u) \\ + \sum_{\mathbf{p}} \eta \|\mathbf{v}(\mathbf{p}) - \mathbf{w}(\mathbf{p})\|^2. \end{aligned} \tag{11}$$

To solve the flow $\mathbf{w}$, we utilize the SIFT flow algorithm [4] by modifying the small displacement term from $\sum_{\mathbf{p}} \|\mathbf{w}(\mathbf{p})\|_1$ to $\sum_{\mathbf{p}} \|\mathbf{v}(\mathbf{p}) - \mathbf{w}(\mathbf{p})\|^2$.

The two steps of estimating displacement function $\mathbf{v}$ and flow $\mathbf{w}$ are iterated to obtain a reliable result. Note that the value of objective function (8) will decrease in each step during the iteration, which guarantees the convergence of the algorithm. The number of iteration is fixed to 3.

Algorithm 3: Non-Rigid Dense Image Correspondence

```
Input: A pair of images, parameters $\gamma, \beta, \lambda, \alpha, \eta, u, \delta$
Output: A dense flow $\mathbf{w}$
Extract a set of sparse SIFT correspondence from the image pair: $S = \{(\mathbf{x}_i, \mathbf{y}_i)\}_{i=1}^n$;
Establish dense SIFT features $\{s_1\}$ and $\{s_2\}$ for two images respectively according to the orientations and scales of the sparse correspondences;
Construct Gram matrix $\mathbf{\Gamma}$, matrix $\mathbf{U}$ and $\mathbf{V}$;
repeat
    if the first iteration
        Initialize parameter $\sigma^2$;
        Compute $\mathbf{C}$ by using Algorithm 1;
    else
        Initialize parameter $\sigma^2$, and initialize $\mathbf{C}$ as its old value;
        Optimize the objective function (9) by a quasi-Newton method together with deterministic annealing to compute $\mathbf{C}$;
    end
    Update the displacement function $\mathbf{v} \leftarrow \mathbf{VC}$;
    Optimize the objective function (11) by using the SIFT flow algorithm to compute the flow $\mathbf{w}$;
until reach the maximum iteration number;
The dense image correspondence can be obtain from $\mathbf{w}$ after the iteration.
```

C. Implementation Details

Observe that the last term of objective function (9) is the average difference between $\mathbf{v}$ and $\mathbf{w}$ on the image lattice. Therefore, to estimate the coefficient matrix $\mathbf{C}$, it is possible to downsample the image lattice to achieve a significant speedup without much performance degradation. In our evaluation, we use the uniform sampling strategy with a sampling interval 10 pixels. Besides, we initialize $\mathbf{C}$ independent of $\mathbf{W}$ by using Algorithm 1. We also use data normalization so that the two sets of sparse feature points $\{\mathbf{x}_i\}_{i=1}^n$ and $\{\mathbf{y}_i\}_{i=1}^n$ both have zero mean and unit variance.

The SIFT flow algorithm suffers from the scale and rotation problem. Here we address this issue in our dense matching scheme. Note that the sparse SIFT matching provides the orientations and scales for each correspondence, and for each inlier correspondence, the orientation difference and scale ratio between the two feature points are in general constants. Based on this observation, we choose a small part of the sparse SIFT correspondences (e.g., 20%) with the highest matching scores which are in general inliers, and then set the orientation and scale of the dense SIFT features to the mean orientation difference and mean scale ratio of the chosen correspondences.

The parameters in our dense correspondence algorithm are set according to Algorithm 1 and the SIFT flow algorithm. There is an additional parameter δ need to be set in equation (9) after the first iteration, which controls the trade-off between the sparse matching and dense matching. We set it according to the equation $\delta\eta = 10^3$. Our dense image correspondence method is summarized in Algorithm 3.

VI. Experimental Results

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

A. Results on Non-Rigid Point Set Registration

We tested our method on the same synthesized data as in [3] and [28]. 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 [28].

Fig. 3a shows the registration results of our method on solving different degrees of deformations and occlusions. We also give a quantitative evaluation in Fig. 3b, where the *recall* is computed as the metric used in [26]. Here the recall 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, e.g., the Euclidean distance between a point in the warped model and the corresponding point in the target. 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. The average run time of our method on this dataset with about 100 points is about 0.5 seconds on an Intel Core 2.5 GHz PC with Matlab code.

To provide a quantitative comparison, we report the results of five state-of-the-art algorithms such as shape context [12], TPS-RPM [3], RPM-LNS [28], GMMREG [26] and CPD [24] which are implemented using publicly available codes. For the GMMREG algorithm, we choose the L2 distance and TPS kernel for evaluation. 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 Fig. 3c. In the deformation test results (e.g., 1st and 3rd rows), six algorithms achieve similar registration performance in both *fish* and *Chinese character* at low deformation levels, and our method generally gives better

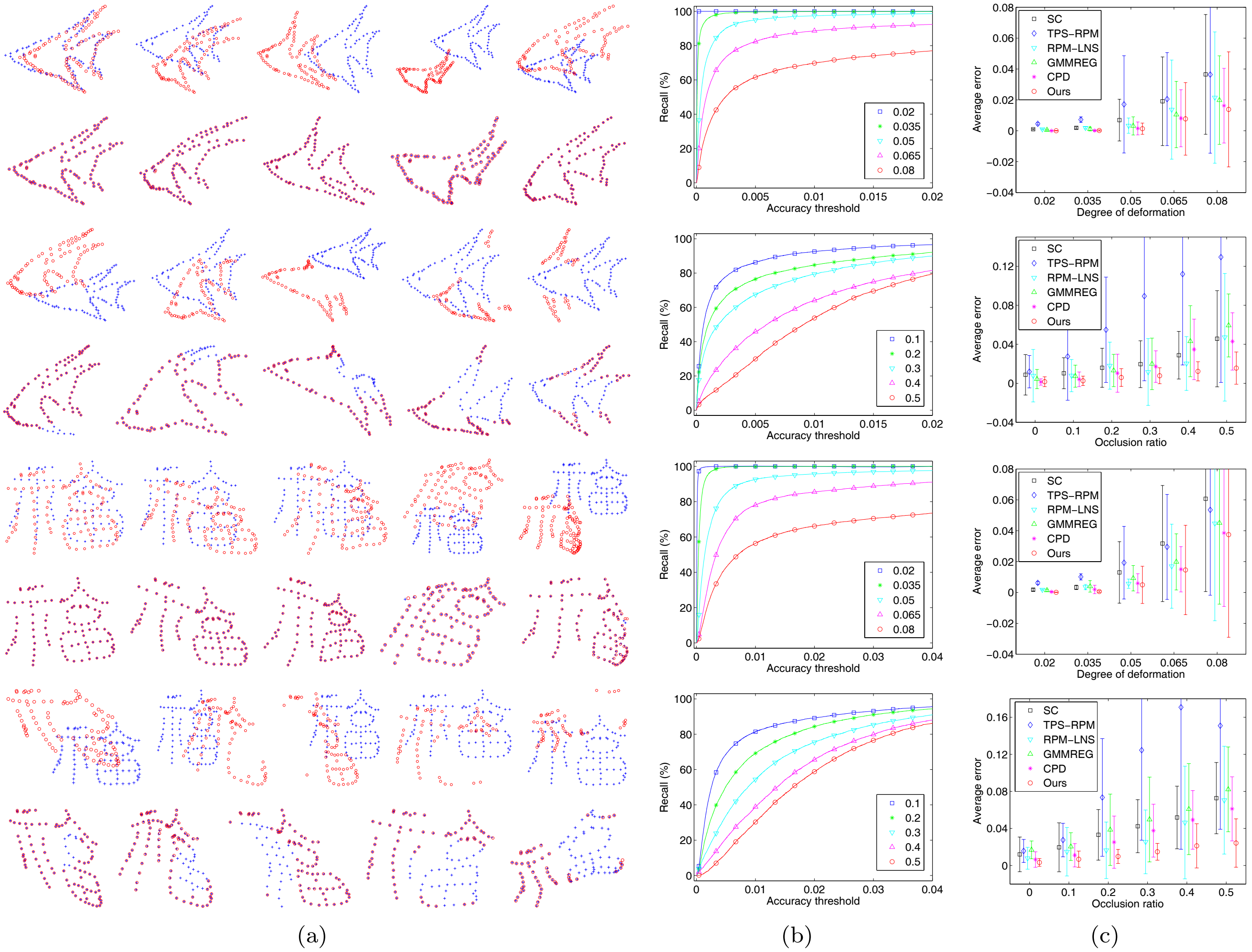


Fig. 3: Point set registration on 2D non-rigid shapes. (a) Results of our method on the *fish* (top) and *Chinese character* (bottom) shapes [3], [28], with deformation and occlusion presented in every two rows. The goal is to align the model point set (blue pluses) onto the target point set (red circles). 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. (b) Quantitative evaluations of our method on the corresponding datasets with different degrees of degradation. Each curve is generated based on 100 trials. (c) Comparisons of the registration performance of our method with shape context (SC) [12], TPS-RPM [3], RPM-LNS [28], GMMREG [26] and CPD [24] on the corresponding datasets. The error bars indicate the registration error means and standard deviations over 100 trials.

performance as the degree of 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 five algorithms. It is not surprised that our algorithm can achieve the best performance, since shape context and RPM-LNS do not consider the robust issue during estimating the transformation, while TPS-RPM, GMMREG and CPD do not consider using local shape features to help establish point correspondences. Note that the local shape features come from the point sets themselves.

More experiments on rotation, noise and outliers are also performed on the two shape models, as shown in Fig. 4. 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.

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.

B. Results on Non-Rigid Sparse 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 non-rigid.

Fig. 5 contains a newspaper with different amounts of spatial warps. We aim to establish correspondences between

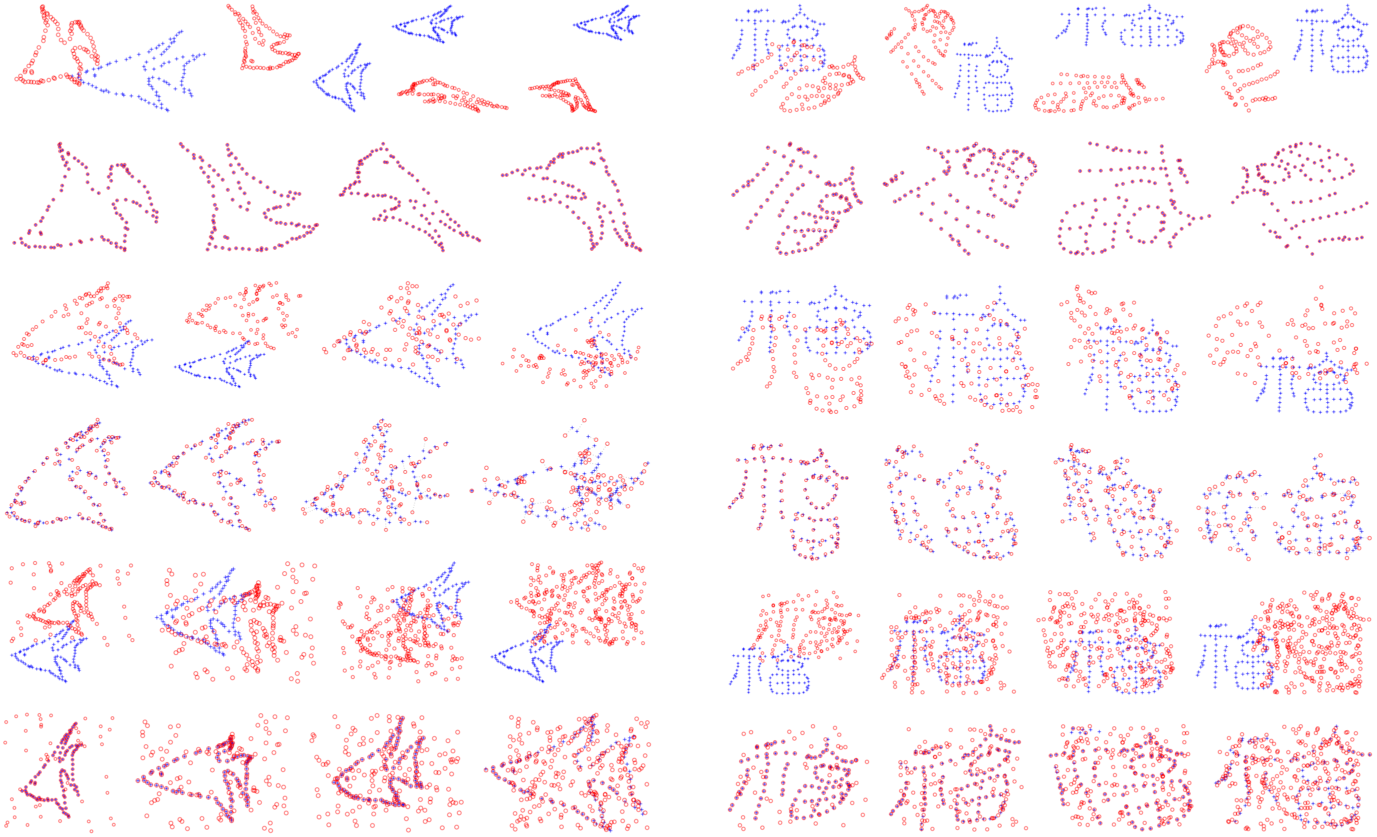


Fig. 4: From top to bottom, results on rotation, noise and outliers presented in every two rows. For each group of experiments, the upper figure is the data, and the lower figure is the registration result. From left to right, increasing degree of degradation.

sparse image features in each image pair. In our evaluation, we first extract SIFT [34] 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. 5. The inlier and true positive, false positive, true negative, false negative are defined within some tolerance radius on human annotated, per-pixel ground truth. 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 reject outliers, 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 0.4 seconds on an Intel Core 2.5 GHz PC with Matlab code.

TABLE I: Performance comparison on the image pairs in Fig. 5. The values in the first row are the inlier percentages (%), and the pairs are the precision-recall pairs (%).

Inlier pct.	79.61	56.57	51.84	45.71
RANSAC [18]	(100.00, 97.12)	(97.63, 91.56)	(92.26, 93.46)	(99.35, 92.22)
ICF [53]	(96.05, 98.38)	(83.95, 86.73)	(80.43, 95.48)	(75.42, 92.71)
VFC [49]	(100.00, 97.04)	(98.59, 99.53)	(98.09, 99.35)	(98.94, 97.91)
Ours (sparse)	(100.00, 99.73)	(99.06, 99.53)	(98.09, 99.35)	(100.00, 98.96)
Ours (dense)	(100.00, 99.73)	(100.00, 99.53)	(99.36, 100.00)	(100.00, 98.96)

In addition, we also compared our method to three state-of-the-art methods, such as random sample consensus (RANSAC) [18], identifying point correspondences by correspondence function (ICF) [53] and vector field consensus (VFC) [49], [54], [55]. We choose the affine model in RANSAC since recent work [56] justifies a simple RANSAC-driven deformable registration technique with an affine model that is at least as accurate as other methods based on the optimization of fully deformable models. 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 interpolation problem, and interpolates a smooth field to fit the potential inliers as well as estimates a consensus inlier set. The results are shown in Table I. We see that all the four algorithms work well when the deformation contained in the image pair is relatively slight. As the amount

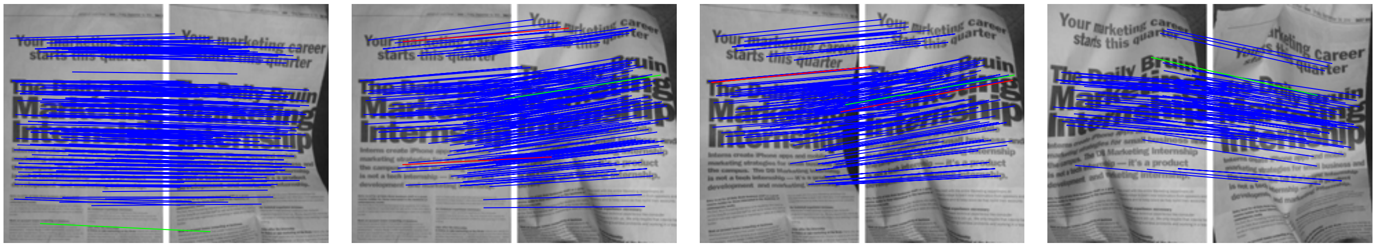


Fig. 5: Results of image feature correspondence on 2D image pairs of resolution 384×256 with deformable objects. From left to right, increasing degree of deformation. The inlier percentages in the initial correspondences are 79.61%, 56.57%, 51.84% and 45.71% respectively, and the corresponding precision-recall 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.

of deformation increases, the performance of RANSAC and ICF begin to degenerate, especially the ICF method. 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. Besides, we also test our dense correspondence method (i.e., Algorithm 3) on this task, as shown in the last row. More specifically, we first estimate a dense flow for each pixel in the image by using Algorithm 3, and then check the correctness of each sparse correspondence according to the flow. We see that the results are further improved by using our dense correspondence method.

Our next experiment involves feature point matching on 3D surfaces. We adopt MeshDOG and MeshHOG [57] as the feature point detector and descriptor to determine the initial correspondences. For the dataset, we use a surface correspondence benchmark[1] [58], which contains meshes representing people and animals in a variety of poses. Here we select four classes and each class four meshes with large level of deformations, as shown in Fig. 6. The ground truth correspondences are supplied by the dataset.

Fig. 6 presents the matching results of our method. We see that our method in general can obtain good performance, both in precision and recall. However, if there exists a body part with large degree of deformation and simultaneously few inlier correspondences, then all the correspondences on that part may be treated as outliers by our algorithm and hence removed, such as the legs of cat and dog, the hands of centaur and person. A quantitative comparison of our method with VFC [49] is reported in Table II (we do not compare to RANSAC and ICF since they are not applicable in this case). Clearly, our method has a better precision-recall trade-off. The average precision-recall pairs on the eight meshes are about $(94.45\%, 86.07\%)$ and $(95.98\%, 94.33\%)$ for VFC and our method.

C. Results on Non-Rigid Image Registration

In this section, we test our non-rigid dense correspondence method on a set of challenging image pairs with shared content, which includes image transformations such as viewpoint change, scale change, rotation, as well as non-rigid deformation. We also compare our method with existing state-of-the-art dense correspondence methods.

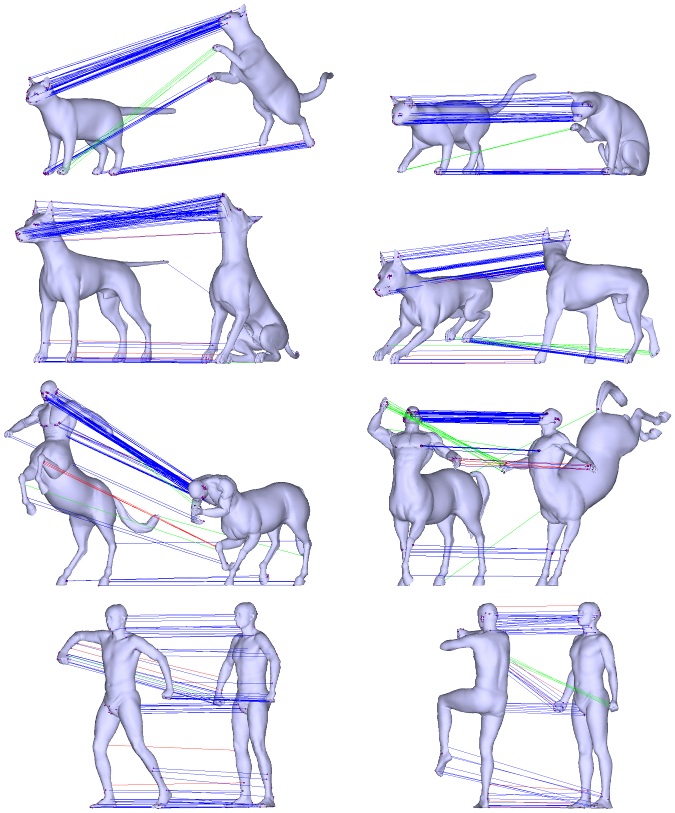


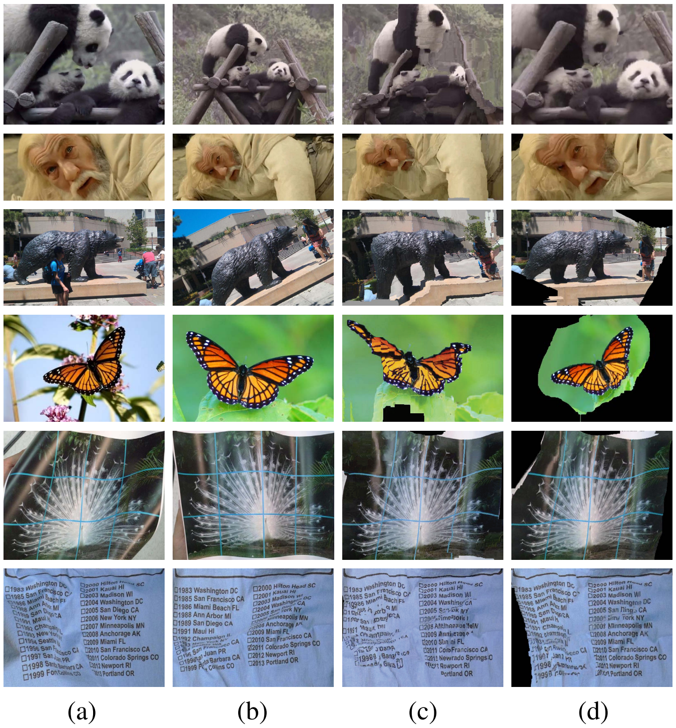
Fig. 6: Results on 3D surfaces of deformable objects. From left to right, top to bottom: *Cat1*, *Cat2*, *Dog1*, *Dog2*, *Centaur1*, *Centaur2*, *Person1*, *Person2*. The inlier percentages in the initial correspondences are 93.38%, 93.03%, 80.57%, 82.46%, 77.48%, 76.92%, 51.98% and 27.40% respectively, and the corresponding precision-recall pairs are $(99.43\%, 95.08\%)$, $(99.20\%, 94.03\%)$, $(96.37\%, 99.52\%)$, $(96.84\%, 93.40\%)$, $(95.51\%, 94.77\%)$, $(95.44\%, 86.79\%)$, $(95.02\%, 99.60\%)$ and $(86.99\%, 91.45\%)$ respectively. For visibility, only 100 randomly selected correspondences are shown. The lines indicate matching results (blue = true positive, green = false negative, red = false positive). Best viewed in color.

Fig. 7 shows a visual comparison between our method and SIFT flow on six pairs of real-world scenes. We aim to align the model images (second column) onto the target images (first column). In the first two rows and middle two rows, we test the robustness of our method to large scale change and rotation respectively. The warped model images of SIFT flow and our method are shown in the third and fourth columns. Clearly, SIFT flow will fail in these cases since in the algorithm

[1] Available at http://www.cs.princeton.edu/~vk/projects/CorrsBlended/doc_data.php.

TABLE II: Performance comparison on the surface pairs in Fig. 6. The values in the first row are the inlier percentages (%), and the pairs are the precision-recall pairs (%). From left to right: results of *Cat1*, *Cat2*, *Dog1*, *Dog2*, *Centaur1*, *Centaur2*, *Person1*, *Person2*.

Inlier pct.	93.38	93.03	80.57	82.46	77.48	76.92	51.98	27.40
VFC [49]	(99.76, 76.10)	(99.39, 93.46)	(97.17, 92.41)	(99.78, 74.53)	(100.00, 82.17)	(99.76, 78.87)	(92.88, 99.60)	(66.87, 91.45)
Ours	(99.43, 95.08)	(99.20, 94.03)	(96.37, 99.52)	(96.84, 93.40)	(95.51, 94.77)	(95.44, 86.79)	(95.02, 99.60)	(86.99, 91.45)



(a) (b) (c) (d)

Fig. 7: Qualitative comparison of matches on real-world scenes. The goal is to align the model images (b) onto the target images (a); (c) and (d) show the matching results (i.e., warped model images) of SIFT flow [4] and our dense correspondence method respectively. The upper two rows test the scale change, the middle two rows test the rotation, and the bottom two rows test the non-rigid deformation.

the dense SIFT features in the model and target images are computed with the same scale and orientation. However, due to the use of the scale and orientation information of sparse feature points for computing the dense feature, our method is not influenced by the rotation and scale change. It can produce almost perfect alignments under these image transformations. We also test the performance of our method in case of non-rigid deformation, as shown in the bottom two rows in Fig. 7. Again, our method displays its robustness and is able to generate good alignment even under extreme deformation (e.g., the last row). In contrast, the SIFT flow algorithm works quite well when the deformation is moderate (e.g., the next-to-last row), but fails in the case of extreme deformation (e.g., the last row).

Next, we give a quantitative comparison of our method with three existing state-of-the-art dense correspondence methods. The first method is implemented based on sparse SIFT feature correspondence, which has been a standard approach to image registration of the same scene. More specifically, we first exact a set of sparse SIFT feature correspondence from the image pair, and then utilize our robust L_2E method (i.e., Algorithm 1) to compute a mapping function/transformation from the sparse correspondence, finally we transform the model image to the target image by means of the mapping function. The second method is the SIFT flow method which matches dense sampled SIFT features. The third method is the NRDC method [38] which is based on matching image patches [39]. For the SIFT flow and NRDC algorithms, we implement them based on the publicly available codes.

Since ground truth data for real scenes of this kind is scarce, we turn to a standard dataset by Mikolajczyk *et al.* [59] that has been used for evaluating sparse feature based correspondence algorithms. The dataset contains image pairs either of planar scenes or taken by camera in a fixed position during acquisition. The images, therefore, always obey homography. The ground truth homographies are supplied by the dataset. The dataset contains six image transformations including rotation, scale change, viewpoint change, image blur, JPEG compression, and illumination. We focus our comparison on large geometric deformations, i.e., the first three transformations, and picked the related two subsets named “zoom + rotation” and “viewpoint”, as shown in Fig. 8. In addition, we also generate an image pair of non-rigid deformation for each subset. More specifically, in Fig. 8, we deform the model image in the second row in each subset by using the *Moving Least Squares* algorithm [60], [61], and then align the deformed model image to the target image. The ground truth correspondence can be obtained based on the homography and non-rigid transformation.

The results are shown in Fig. 8, in which we highlight only regions of correspondences that fall within a radius of 5 pixels from the ground-truth. The detail matching performance comparisons between the four methods are given in the last column. From the results, we see that the sparse correspondence based method works quite well in the case of homography, but degrades seriously after we add the non-rigid deformations. The SIFT flow method completely fails due to the scale change or rotation. The image patch matching based method method NRDC can obtain satisfied results in most cases, although the captured correspondences typically are not so many. Besides, it also degrades badly in extreme cases, e.g., in the last row of Fig. 8 which contains extreme scale change, rotation, as well as non-rigid deformation. In contrast, our dense correspondence method is much more robust to all these degenerations, and in general able to capture much larger and more reliable matches than the other methods.

VII. Conclusion

In this paper, we have presented a new approach for non-rigid registration. A key characteristic of our approach is the

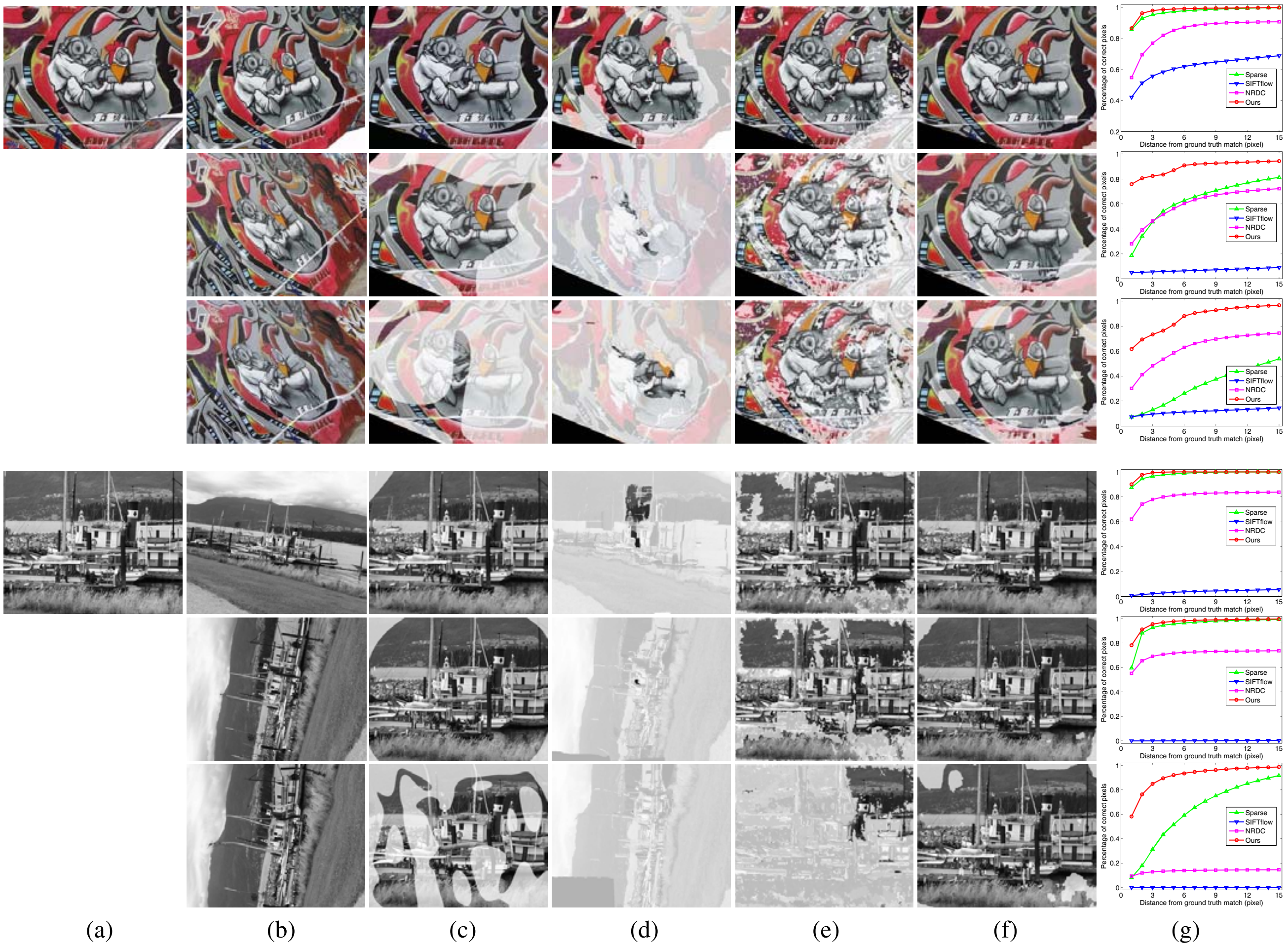


Fig. 8: Correspondence evaluation on two examples from the dataset of Mikolajczyk *et al*. [59]. The goal is to align the model images (b) onto the target images (a); (c), (d), (e) and (f) show the matching results (i.e., warped model images) of sparse correspondence based method, SIFT flow [4], NRDC [38], and our dense correspondence method, we highlight only regions of matches that fall within a radius of 5 pixels from the ground-truth, and the black regions denote that there are no ground truth correspondences; (g) shows the quantitative comparison of the four methods. The upper group of images contains view point change and non-rigid deformation; the bottom group contains scale change, rotation, as well as non-rigid deformation.

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 robust method to sparse correspondence such as non-rigid point set registration and sparse image feature correspondence, and dense correspondence such as non-rigid image registration. Experiments on public datasets for non-rigid point registration, real images for sparse image feature correspondence and non-rigid image registration demonstrate that our approach yields results consistently outperform those of state-of-the-art methods such as CPD and SIFT flow when there are significant deformations, occlusions, rotations, and/or scale changes 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 and NSF award IIS-1216528.

References

[1] L. G. Brown, “A Survey of Image Registration Techniques,” *ACM Computing Surveys*, vol. 24, no. 4, pp. 325–376, 1992.

[2] P. J. Besl and N. D. McKay, “A Method for Registration of 3-D Shapes,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 14, no. 2, pp. 239–256, 1992.

[3] H. Chui and A. Rangarajan, “A new point matching algorithm for non-rigid registration,” *Computer Vision and Image Understanding*, vol. 89, pp. 114–141, 2003.

[4] C. Liu, J. Yuen, and A. Torralba, “SIFT Flow: Dense Correspondence Across Scenes and Its Applications,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 33, no. 5, pp. 978–994, 2011.

[5] D. Scharstein and R. Szeliski, “A taxonomy and evaluation of dense two-frame stereo correspondence algorithms,” *International Journal of Computer Vision*, vol. 47, no. 1, pp. 7–42, 2002.

[6] B. D. Lucas and T. Kanade, “An Iterative Image Registration Technique with An Application to Stereo Vision,” in *Proceedings of the International Joint Conference on Artificial Intelligence*, 1981, pp. 674–679.

[7] B. K. P. Horn and B. G. Schunck, “Determining optical flow,” *Artificial intelligence*, vol. 17, no. 1, pp. 185–203, 1981.

[8] P. F. Felzenszwalb and D. P. Huttenlocher, “Pictorial Structures for Object Recognition,” *International Journal of Computer Vision*, vol. 61, no. 1, pp. 55–79, 2005.

[9] A. W. Fitzgibbon, “Robust Registration of 2D and 3D Point Sets,” *Image and Vision Computing*, vol. 21, pp. 1145–1153, 2003.

[10] S. Rusinkiewicz and M. Levoy, "Efficient Variants of the ICP Algorithm," in *3-D Digital Imaging and Modeling*, 2001, pp. 145–152.

[11] B. Luo and E. R. Hancock, "Structural Graph Matching Using the EM Algorithm and Singular Value Decomposition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 23, no. 10, pp. 1120–1136, 2001.

[12] S. Belongie, J. Malik, and J. Puzicha, "Shape Matching and Object Recognition Using Shape Contexts," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 24, pp. 509–522, 2002.

[13] D. W. Scott, "Parametric Statistical Modeling by Minimum Integrated Square Error," *Technometrics*, vol. 43, no. 3, pp. 274–285, 2001.

[14] 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.

[15] N. Aronszajn, "Theory of Reproducing Kernels," *Transactions of the American Mathematical Society*, vol. 68, no. 3, pp. 337–404, 1950.

[16] B. Zitová and J. Flusser, "Image Registration Methods: A Survey," *Image and Vision Computing*, vol. 21, pp. 977–1000, 2003.

[17] J. Ma, J. Zhao, J. Tian, Z. Tu, and A. Yuille, "Robust estimation of nonrigid transformation for point set registration," in *IEEE conference on Computer Vision and Pattern Recognition*, 2013, pp. 2147–2154.

[18] M. A. Fischler and R. C. Bolles, "Random Sample Consensus: A Paradigm for Model Fitting with Application to Image Analysis and Automated Cartography," *Communications of the ACM*, vol. 24, no. 6, pp. 381–395, 1981.

[19] P. H. S. Torr and A. Zisserman, "MLESAC: A New Robust Estimator with Application to Estimating Image Geometry," *Computer Vision and Image Understanding*, vol. 78, no. 1, pp. 138–156, 2000.

[20] O. Chum, J. Matas, and J. Kittler, "Locally Optimized RANSAC," in *Proceedings of Pattern Recognition Symposium of the German Association for Pattern Recognition (DAGM)*, 2003, pp. 236–243.

[21] O. Chum and J. Matas, "Matching with PROSAC - Progressive Sample Consensus," in *Proceedings of IEEE conference on Computer Vision and Pattern Recognition*, 2005, pp. 220–226.

[22] A. L. Yuille, "Generalized Deformable Models, Statistical Physics, and Matching Problems," *Neural Computation*, vol. 2, no. 1, pp. 1–24, 1990.

[23] A. Rangarajan, H. Chui, and F. L. Bookstein, "The Softassign Procrustes Matching Algorithm," in *Information Processing in Medical Imaging*. Springer, 1997, pp. 29–42.

[24] A. Myronenko and X. Song, "Point Set Registration: Coherent Point Drift," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 32, no. 12, pp. 2262–2275, 2010.

[25] Y. Tsin and T. Kanade, "A Correlation-Based Approach to Robust Point Set Registration," in *Proceedings of European Conference on Computer Vision*, 2004, pp. 558–569.

[26] B. Jian and B. C. Vemuri, "Robust Point Set Registration Using Gaussian Mixture Models," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 33, no. 8, pp. 1633–1645, 2011.

[27] ——, "A robust algorithm for point set registration using mixture of gaussians," in *IEEE International Conference on Computer Vision*, 2005, pp. 1246–1251.

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

[29] J.-H. Lee and C.-H. Won, "Topology Preserving Relaxation Labeling for Nonrigid Point Matching," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 33, no. 2, pp. 427–432, 2011.

[30] X. Huang, N. Paragios, and D. N. Metaxas, "Shape Registration in Implicit Spaces Using Information Theory and Free Form Deformations," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 28, no. 8, pp. 1303–1318, 2006.

[31] R. Horaud, F. Forbes, M. Yguel, G. Dewaele, and J. Zhang, "Rigid and Articulated Point Registration with Expectation Conditional Maximization," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 33, no. 3, pp. 587–602, 2011.

[32] L. Torresani, V. Kolmogorov, and C. Rother, "Feature Correspondence via Graph Matching: Models and Global Optimization," in *Proceedings of European Conference on Computer Vision*, 2008, pp. 596–609.

[33] W. E. L. Grimson, "Computational Experiments with A Feature Based Stereo Algorithm," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, no. 1, pp. 17–34, 1985.

[34] D. Lowe, "Distinctive Image Features from Scale-Invariant Keypoints," *International Journal of Computer Vision*, vol. 60, no. 2, pp. 91–110, 2004.

[35] J. Matas, O. Chum, M. Urban, and T. Pajdla, "Robust Wide Baseline Stereo from Maximally Stable Extremal Regions," in *British Machine Vision Conference*, vol. 1, 2002, pp. 384–393.

[36] R. Szeliski, "Image Alignment and Stitching: A Tutorial," *Foundations and Trends in Computer Graphics and Vision*, vol. 2, no. 1, pp. 1–104, 2006.

[37] A. Sotiras and N. Paragios, "Deformable Image Registration: A Survey," 2012.

[38] Y. HaCohen, E. Shechtman, D. B. Goldman, and D. Lischinski, "Non-Rigid Dense Correspondence with Applications for Image Enhancement," *ACM Transactions on Graphics*, vol. 30, no. 4, p. 70, 2011.

[39] C. Barnes, E. Shechtman, D. B. Goldman, and A. Finkelstein, "The Generalized Patchmatch Correspondence Algorithm," in *Proceedings of European Conference on Computer Vision*. Springer, 2010, pp. 29–43.

[40] J. Ma, J. Zhao, Y. Ma, and J. Tian, "Non-rigid visible and infrared face registration via regularized gaussian fields criterion," *Pattern Recognition*, vol. 48, no. 3, pp. 772–784, 2015.

[41] B. C. Vemuri, J. Liu, and J. L. Marroquín, "Robust multimodal image registration using local frequency representations," in *Information Processing in Medical Imaging*, 2001, pp. 176–182.

[42] J. Liu, B. C. Vemuri, and J. L. Marroquín, "Local Frequency Representations for Robust Multimodal Image Registration," *IEEE Trans. on Medical Imaging*, vol. 21, no. 5, pp. 462–469, 2002.

[43] P. J. Huber, *Robust Statistics*. New York: John Wiley & Sons, 1981.

[44] A. L. Yuille and N. M. Grzywacz, "A Mathematical Analysis of the Motion Coherence Theory," *International Journal of Computer Vision*, vol. 3, no. 2, pp. 155–175, 1989.

[45] C. A. Micchelli and M. Pontil, "On Learning Vector-Valued Functions," *Neural Computation*, vol. 17, no. 1, pp. 177–204, 2005.

[46] G. Wahba, *Spline Models for Observational Data*. SIAM, Philadelphia, PA, 1990.

[47] J. Ma, J. Zhao, Y. Zhou, and J. Tian, "Mismatch Removal via Coherent Spatial Mapping," in *Proceedings of International Conference on Image Processing*, 2012, pp. 1–4.

[48] J. Chen, J. Ma, C. Yang, and J. Tian, "Mismatch removal via coherent spatial relations," *Journal of Electronic Imaging*, vol. 23, no. 4, pp. 043 012–043 012, 2014.

[49] J. Zhao, J. Ma, J. Tian, J. Ma, and D. Zhang, "A Robust Method for Vector Field Learning with Application to Mismatch Removing," in *Proceedings of IEEE conference on Computer Vision and Pattern Recognition*, 2011, pp. 2977–2984.

[50] R. Rifkin, G. Yeo, and T. Poggio, "Regularized Least-Squares Classification," in *Advances in Learning Theory: Methods, Model and Applications*, 2003.

[51] J. Ma, J. Zhao, J. Tian, X. Bai, and Z. Tu, "Regularized vector field learning with sparse approximation for mismatch removal," *Pattern Recognition*, vol. 46, no. 12, pp. 3519–3532, 2013.

[52] A. E. Johnson and M. Hebert, "Using Spin-Images for Efficient Object Recognition in Cluttered 3-D Scenes," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 21, no. 5, pp. 433–449, 1999.

[53] X. Li and Z. Hu, "Rejecting Mismatches by Correspondence Function," *International Journal of Computer Vision*, vol. 89, no. 1, pp. 1–17, 2010.

[54] J. Ma, J. Zhao, J. Tian, A. L. Yuille, and Z. Tu, "Robust point matching via vector field consensus," *IEEE Transactions on Image Processing*, vol. 23, no. 4, pp. 1706–1721, 2014.

[55] J. Ma, Y. Ma, J. Zhao, and J. Tian, "Image feature matching via progressive vector field consensus," *IEEE Signal Processing Letters*, vol. 22, no. 6, pp. 767–771, 2015.

[56] Q.-H. Tran, T.-J. Chin, G. Carneiro, M. S. Brown, and D. Suter, "In defence of RANSAC for outlier rejection in deformable registration," in *ECCV*. Springer, 2012, pp. 274–287.

[57] A. Zaharescu, E. Boyer, K. Varanasi, and R. Horaud, "Surface Feature Detection and Description with Applications to Mesh Matching," in *Proceedings of IEEE conference on Computer Vision and Pattern Recognition*, 2009, pp. 373–380.

[58] V. G. Kim, Y. Lipman, and T. Funkhouser, "Blended intrinsic maps," *ACM Transactions on Graphics*, vol. 30, no. 4, p. 79, 2011.

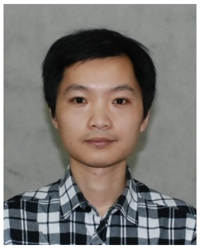
[59] K. Mikolajczyk, T. Tuytelaars, C. Schmid, A. Zisserman, J. Matas, F. Schaffalitzky, T. Kadir, and L. van Gool, "A Comparison of Affine Region Detectors," *International Journal of Computer Vision*, vol. 65, no. 1, pp. 43–72, 2005.

[60] S. Schaefer, T. McPhail, and J. Warren, "Image Deformation Using Moving Least Squares," *ACM Transactions on Graphics*, vol. 25, no. 3, pp. 533–540, 2006.

[61] D. Levin, "The Approximation Power of Moving Least-Squares," *Mathematics of Computation*, vol. 67, no. 224, pp. 1517–1531, 1998.

Jiayi Ma received the B.S. degree from the Department of Mathematics, and the Ph.D. Degree from the School of Automation, Huazhong University of Science and Technology, Wuhan, China, in 2008 and 2014, respectively. From 2012 to 2013, he was with the Department of Statistics, University of California at Los Angeles. He is now a Post-Doctoral with the Electronic Information School, Wuhan University. His current research interests include in the areas of computer vision, machine learning, and pattern recognition.

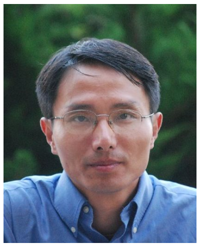
Weichao Qiu received the B.E. and M.S. degree from Huazhong University of Science and Technology, Wuhan, China in 2011 and 2014 respectively. He is currently a Ph.D. student of the Statistics department, UCLA. His research interest is computer vision.



Ji Zhao received the B.S. degree in automation from Nanjing University of Posts and Telecommunication in 2005. He received the Ph.D. degree in control science and engineering from HUST in 2012. Since 2012, he is a postdoctoral research associate at the Robotics Institute, Carnegie Mellon University. His research interests include image classification, image segmentation and kernel-based learning.

Yong Ma graduated from the Department of Automatic Control, Beijing Institute of Technology, Beijing, China, in 1997. He received the Ph.D. degree from the Huazhong University of Science and Technology (HUST), Wuhan, China, in 2003. His general field of research is in signal and systems. His current research projects include remote sensing of the Lidar and infrared, as well as Infrared image processing, pattern recognition, interface circuits to sensors and actuators. Between 2004 and 2006, he was a Lecturer at the University of the West of England, Bristol, U.K. Between 2006 and 2014, he was with the Wuhan National Laboratory for Optoelectronics, HUST, Wuhan, where he was a Professor of electronics. He is now a Professor with the Electronic Information School, Wuhan University.

Alan L. Yuille received his B.A. in mathematics from the University of Cambridge in 1976, and completed his Ph.D. in theoretical physics at Cambridge in 1980 studying under Stephen Hawking. Following this, he held a postdoc position with the Physics Department, University of Texas at Austin, and the Institute for Theoretical Physics, Santa Barbara. He then joined the Artificial Intelligence Laboratory at MIT (1982-1986), and followed this with a faculty position in the Division of Applied Sciences at Harvard (1986-1995), rising to the position of associate professor. From 1995-2002 Alan worked as a senior scientist at the Smith-Kettlewell Eye Research Institute in San Francisco. In 2002 he accepted a position as full professor in the Department of Statistics at the University of California, Los Angeles. He has over two hundred peer-reviewed publications in vision, neural networks, and physics, and has co-authored two books: Data Fusion for Sensory Information Processing Systems (with J. J. Clark) and Two- and Three- Dimensional Patterns of the Face (with P. W. Hallinan, G. G. Gordon, P. J. Giblin and D. B. Mumford); he also co-edited the book Active Vision (with A. Blake). He has won several academic prizes and is a Fellow of IEEE.



Zhuowen Tu is an assistant professor in the Department of Cognitive Science, and the Department of Computer Science and Engineering, University of California, San Diego. Before joining UCSD, he was an assistant professor at UCLA. Between 2011 and 2013, he took a leave to work at Microsoft Research Asia. He received his PhD from the Ohio State University and his ME from Tsinghua University. Zhuowen Tu received NSF CAREER award in 2009 and was awarded with David Marr prize in 2003.