Shape Registration in Implicit Spaces using Information Theory and Free Form Deformations

Xiaolei Huang, *Member, IEEE*, Nikos Paragios, *Senior Member, IEEE*, and Dimitris N. Metaxas, *Senior Member, IEEE*

Abstract

We present a novel variational and statistical approach for shape registration. Shapes of interest are implicitly embedded in a higher dimensional space of distance transforms. In this implicit embedding space, registration is formulated in a hierarchical manner: the Mutual Information criterion supports various transformation models and is optimized to perform global registration; then a B-spline based Incremental Free Form Deformations (IFFD) model is used to minimize a Sum-of-Squared-Differences (SSD) measure and further recover a dense local non-rigid registration field. The key advantage of such framework is twofold: (1) it naturally deals with shapes of arbitrary dimension (2D, 3D or higher) and arbitrary topology (multiple parts, closed/open), and (2) it preserves shape topology during local deformation, and produces local registration fields that are smooth, continuous and establish one-to-one correspondences. Its invariance to initial conditions is evaluated through empirical validation, and various hard 2D/3D geometric shape registration examples are used to show its robustness to noise, severe occlusion and missing parts. We demonstrate the power of the proposed framework using two applications: one for statistical modeling of anatomical structures, another for 3D face scan registration and expression tracking. We also compare the performance of our algorithm with that of several other well-known shape registration algorithms.

Index Terms

shape registration, mutual information, free form deformations, correspondences, implicit shape representation, distance transforms, partial differential equations

X. Huang and D. N. Metaxas are with the Center for Computational Biomedicine Imaging and Modeling (CBIM), Division of Computer and Information Sciences, Rutgers University - New Brunswick, NJ, USA. Emails: {xiaolei, dnm}@cs.rutgers.edu.

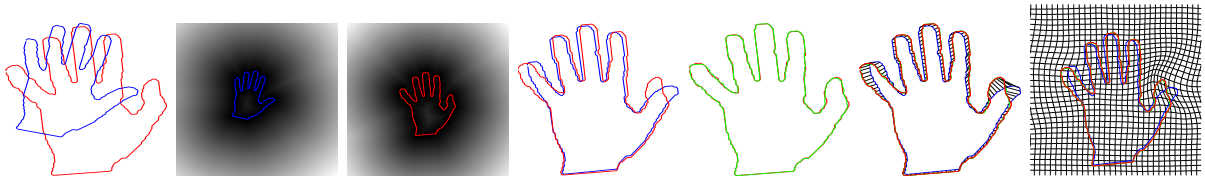
N. Paragios is with the Applied Mathematics and Systems laboratory (MAS), Ecole Centrale de Paris, Grande Voie des Vignes, 92 295 Chatenay-Malabry, FRANCE. Email: nikos.paragios@ecp.fr.

I. Introduction

Shape registration is critical to various imaging and vision applications [1]. Global registration, also known as shape alignment, aims to recover a global transformation that brings the pose of a source shape as close as possible to that of a target shape. The alignment has extensive uses in recognition, indexing and retrieval, and tracking. To further account for important local deformations, non-rigid local registration is needed to establish dense correspondences between the basic elements of shapes, such as points, curvature, etc. Medical imaging is a domain that requires local registration such as in building statistical models for internal organs [2], and intra-subject or atlas registration of 2D/3D anatomical structures.

There has been a lot of previous research on the shape registration problem [3], [4], [5], as well as on similar problems such as shape matching [6], [2], [7], [8], and point set matching [9]. The algorithms proposed differ in the following three main aspects.

1) **Shape Representation** is the selection of an appropriate representation for the shapes of interest. Clouds of points [6], [9], parametric curves/surfaces [7], [10], fourier descriptors [11], medial axes [12], and more recently, implicit distance functions [3], [13] are often considered.
2) **Transformation** refers to the selected global, local, or hierarchical (global-to-local) transformation model, which is used to transform the source shape to match with the target shape. Global transformation models apply to an entire shape; and examples are rigid, similarity, affine and perspective. Local transformation models can represent pixel-wise deformations that deform a shape locally and non-rigidly; and examples include optical flow [3], [14], Thin Plate Splines (TPS) [6], [9], Radial Basis Functions [15], and space deformation techniques such as Free Form Deformations (FFD) [16], [17]. Hierarchical models are also popular since they cover the entire transformation domain using both global and local transformations.
3) **Registration Criterion** is the approach used to recover the optimal transformation parameters given a shape representation and a transformation model. One can classify existing approaches into two sub-categories. The first is to establish explicit geometric feature correspondences and then estimate



(a) (b) (c) (d) (e) (f) (g)

Fig. 1. (a) Initial condition(source shape in blue, target shape in red). (b) The implicit source shape representation using a distance map; points on the shape overlap the zero level set (as drawn in color). (c) The implicit target shape representation. (d) Global alignment using Mutual Information; only the aligned shapes (zero level sets of the implicit representations) are shown. (e) Result after local non-rigid registration using IFFD; the transformed source shape (in green) is shown overlaid on the target shape (in red). (f) Established correspondences using IFFD. (g) The embedding space deformation to achieve local registration.

the transformation parameters using the correspondences [6], [18], [19]. The second is to recover the optimal transformation parameters through optimization of energy functionals [4], [5], [3], [9].

In this paper, we propose a hierarchical shape registration method using the implicit distance function shape representation in a variational framework. Our overall approach is depicted in [Fig. (1)]. The shapes of interest are represented in an implicit form [20], embedded in the space of distance functions of a Euclidean metric [Fig. (1).b-c]. Global alignment using an arbitrary motion model is achieved by maximizing mutual information [21], [22], [23] between two shape-embedding distance functions [Fig. (1).d]. For local non-rigid registration, parameters of a cubic B-spline based Incremental Free Form Deformation model (IFFD) are recovered by minimizing the sum-of-squared-differences between the two globally aligned distance functions [Fig. (1).e]. The resulting registration field [Fig. (1).g] preserves shape topology, is smooth, continuous and gives dense one-to-one correspondences between the source and the target shapes [Fig. (1).f].

The implicit shape representation we use in this paper is gaining increasing attention recently, both in shape registration [3] and in statistical shape modeling [13]. It is attractive in that it is a generic representation that handles naturally shapes of arbitrary dimension and arbitrary topology. This is because it represents shapes using their distance transform “images”, and does not require explicit parameterization of the shapes. The representation is also stable and robust to shape perturbations and noise [24].

Using a hierarchical approach, our algorithm accounts for the transformation between a source shape and a target shape globally as much as possible before applying local non-rigid deformations. This is standard practice adopted by many shape registration and model fitting algorithms [6], [18], [3], [10], because it minimizes shape distortion and produces more accurate local correspondences. In the global registration step, our method supports arbitrary global transformation models. In the local registration step, our method is novel in proposing the use of a space warping technique – Free Form Deformations (FFD)– for shape registration.

In our approach the optimal global and local transformation parameters are recovered by optimizing suitable energy functionals. The global alignment criterion - Mutual Information (MI) - is an information theoretical measure of the statistical dependencies between its inputs, and it has been proven robust to multi-modal image matching [22], [21], [23], [17]. In our method, we adapt MI to global shape registration for the first time. This adaptation is made possible because we use the implicit distance map "image" representation for shapes. Compared to other criteria such as Sum-of-Squared-Differences (SSD), MI does not require equality of its two inputs at its optimum. Instead it enforces statistical dependency. This property of MI makes it invariant to the changes in a shape's implicit representation when the shape undergoes a scale change, hence enables MI to deal with arbitrary global transformations robustly. In local non-rigid registration, the SSD criterion is used, because the many degrees of freedom in local deformations require input equality to achieve a global optimum.

As we will show in this paper through extensive experiments, the integration of implicit shape representation, mutual information global alignment and free-form local registration leads to a powerful shape registration framework that is generic, robust, efficient and guarantees coherent one-to-one correspondences between the basic shape elements on the source and target shapes.

The remainder of this paper is organized as follows. In section 2, we review previous work on shape registration. In section 3, we present in detail our registration framework. This includes the formulation of the implicit shape representation (Sec. 3.1), global registration (Sec. 3.2), and local registration (Sec.

3.3). In section 4, we show some example applications of our registration framework. In section 5, we evaluate the performance of our algorithm and compare it to previous methods. We conclude and discuss about directions of future research in section 6.

II. Background

In this section, we review previous work on shape registration that use some alternatives in terms of shape representation, transformation model and registration criterion. Literature on image registration such as surveyed in [25] is closely related, but we do not consider it in this paper.

A. Shape Representation

Other than the implicit shape representation, point clouds [9], [6] is another intuitive and generic shape representation, since it can easily represent shapes in 2D and 3D with arbitrary topology. A known limitation of the point cloud representation is that, it strongly depends on the sampling rule which affects the number of shape elements, their distribution, etc. For instance, given two shapes to be registered, each represented by a point cloud, the two point sets may not be sampled at corresponding locations due to low-resolution or improper sampling, and this can lead to inherent inconsistencies in the two point sets, thus cast problems when point correspondences are pursued between the two shapes.

Unlike point clouds, parametric curves/surfaces representation of shapes [7], [10] supports valid correspondences, although the explicit parameterization of shapes in arbitrary dimensions and topology is non-trivial in most situations. Fourier descriptors [11] and medial axis [12] are two other shape representations that are excellent when measuring the dissimilarity between shapes, but they are not suited for establishing dense correspondences between shape boundary elements.

B. Transformation

The choices for the global transformation model are fairly universal, and each transformation can be described by a small set of parameters. For instance, the rigid transformation has parameters related to

translation and rotation; the similarity transformation considers translation, rotation, and isotropic scaling; the affine transformation maps parallel lines to parallel lines and can account for translation, rotation, isotropic or anisotropic scaling, as well as shearing. Many shape alignment algorithms have been proposed to solve for the global transformation parameters between shapes, either based on feature correspondences [18], [19], [6], or by minimizing a dissimilarity cost function [4], [3]. In particular, a work closely related to ours [3] recovers parameters of a similarity transformation by minimizing a Sum-of-Squared-Differences (SSD) criterion between the implicit representations of two shapes. Since the implicit representation is invariant to translation and rotation but not to scaling [1], a scale factor was explicitly introduced into the SSD criterion in their method to account for isotropic scaling. However, it was unclear how to extend the framework to deal with anisotropic scaling and affine transformation. In this paper, we address this problem by optimizing the mutual information criterion, which accounts for scale variations implicitly, thus allows a generic formulation that supports arbitrary global transformation model.

For shapes that undergo local deformations, non-rigid local registration is important to further recover a dense deformation field that establishes correspondences between the basic shape elements. The model we use to represent local deformations in this paper is the Free Form Deformations (FFD) model [16], [17], [26], which consists of embedding an object inside a space, and deforming the shape of the object via deforming the space. It couples naturally with the implicit shape representation, which embeds shapes in a higher dimensional space. Underlying each FFD model, there is a choice of the interpolating spline basis function. The cubic B-spline basis function is used in this paper, since it has a local support, guarantees C^1 continuity at control points and C^2 continuity everywhere else. These properties are very desirable in representing continuous, smooth local deformations that preserve shape topology. Other than FFD, the optical flow like local deformation field (i.e. dense pixel-wise displacement field) is another commonly used model for non-rigid registration [3], [14]. Even with advanced regularization and smoothness constraints however, this model is not suitable for registering shapes, since it does not guarantee the preservation of

[1]It is shown that the representation changes linearly as the shape changes scale.

topology and coherence of a shape after deformation (e.g. it can deform a closed shape into an open one [3]), and it does not necessarily produce one-to-one correspondences (see Fig. 16). The Thin Plate Splines (TPS) [6], [9] and the Radial Basis Functions (RBF) [15] are two other popular non-rigid transformation techniques. Both TPS and RBF require explicitly finding two sets of corresponding landmark points, then the TPS or RBF interpolants are used to determine the deformations in other parts of the shapes. The main concern related to these two approaches is the automatic finding of landmark correspondences.

C. Registration Criterion

Recovering transformation parameters by optimizing shape dissimilarity energy functionals is a popular approach [4], [5], [3], [9]. In the Iterative Closest Points (ICP) method [4], [5], correspondences are first assumed between pairs of points in the source and target shapes based on proximity, then a dissimilarity energy function that defines the accumulative distance between the corresponding points is minimized to solve for the transformation parameters. Based on the new transformation, the assumed correspondences are updated, and a new iteration of minimization is initiated. This simple iterative algorithm works quite effectively when given a good initial estimate of the transformation. However, the ICP optimization may get stuck in local minima due to wrongly assumed correspondences, especially during local non-rigid registration or when the initial estimate is poor (see Fig. 17). The non-rigid point matching algorithm proposed in [9] jointly estimates point correspondences and non-rigid transformation parameters by minimizing a dissimilarity energy function defined on a softassign correspondence matrix and the thin plate spline parameters. The optimization of the energy function is based on deterministic annealing and the overall framework is robust to a small fraction of outliers. Because the method is only applicable to point set matching however, the established correspondences can be inherently inaccurate due to the sparse sampling problem. The more recent shape registration work in [3] minimizes energy functions defined on two shape-embedding distance functions. Our method differs from it in two main aspects: first, we use the mutual information criterion for global registration instead of SSD, and second, we use the more elegant free form deformations model for local registration instead of the optical flow like deformations.

Pursuing explicit feature correspondences and then estimating the transformation [6], [18], [19] is another approach to recover registration parameters. The performance of this approach relies on the robustness of feature extraction, and the accuracy of feature correspondences. It is often observed that algorithms adopting this approach may produce outlier correspondences whose effect can only be partially alleviated by robust estimation techniques. Furthermore, the registration problem can become under-constrained, especially in the case of non-rigid registration when many reliable correspondences are needed in order to solve for the many local deformation parameters.

III. METHODOLOGY

In this section, we present in detail our global-to-local shape registration framework.

A. The Implicit Shape Representation

Within the proposed framework, an implicit representation for the source and target shapes is considered. The Euclidean distance transform is used to embed a shape of interest as the zero level set of a distance function in the higher dimensional volumetric space.

In order to facilitate notation, we consider the 2D case. Let $\Phi : \Omega \rightarrow R^{+}$ be a Lipschitz function that refers to the distance transform of a shape $\mathcal{S}$. The shape defines a partition of the image domain Ω: the region that is enclosed by $\mathcal{S}$, $[\mathcal{R}_{\mathcal{S}}]$, and the background region $[\Omega - \mathcal{R}_{\mathcal{S}}]$. Given these definitions, the following implicit shape representation [Fig. (1).b-c] is considered:

$$\Phi_{\mathcal{S}}(x,y) = \begin{cases} 0, & (x,y) \in \mathcal{S} \\ +D((x,y),\mathcal{S}) > 0, & (x,y) \in \mathcal{R}_{\mathcal{S}} \\ -D((x,y),\mathcal{S}) < 0, & (x,y) \in [\Omega - \mathcal{R}_{\mathcal{S}}] \end{cases}$$

where $D((x,y),\mathcal{S})$ refers to the minimum Euclidean distance between the image pixel location (x,y) and the shape $\mathcal{S}$. If $\mathcal{S}$ is an open structure, the un-signed distance transform is used instead.

This representation provides a feature space in which objective functions that are optimized using a gradient descent method can be conveniently used. One can prove that the gradient of the embedding

distance function is a unit vector in the normal direction of the shape; and the representation satisfies a sufficient condition for the convergence of gradient descent methods, which requires continuous first derivatives. Furthermore, the use of the implicit representation provides additional support to the registration process around the shape boundaries and facilitates the imposition of smoothness constraints, since one would like to align the original structures as well as their clones that are positioned coherently in the image/volume plane.

One concern associated with the implicit representation is its efficiency in registration, since it has one dimension higher than the original shape. In our work, this problem is addressed by using only a narrow band around the shape in the embedding space as the sample domain for registration. This significantly speeds up the execution, while producing comparable results to that using the full image domain.

B. Global Registration by Maximizing Mutual Information

The selected implicit shape representation is inherently translation/rotation invariant [3]. When a shape undergoes scale variations, the intensity values of its associated distance map (i.e., its implicit representation) scale accordingly. Therefore the registration of distance maps of a shape in various scales is analogous to matching images in multiple modalities that refer to the same underlying scene elements. Mutual information, an information-theoretic criterion for measuring the global statistical dependency of its two input random variables, has been shown in the literature [27] to be able to address such matching objective. The integration of mutual information and the implicit representation gives rise to a global alignment framework that is invariant to translation, rotation, scaling, and accommodates transformations in arbitrary dimensions.

In order to facilitate notation let us denote the source shape representation $\Phi_{\mathcal{D}}$ as f [Fig. (1).b], and the target shape representation $\Phi_{\mathcal{S}}$ as g [Fig. (1).c]. Both f and g are intensity "images" where the intensity values refer to the distance values to the underlying shapes respectively. In the most general case, let us consider a sample domain Ω in the image domain of the source representation f [2], then global registration

[2]In practice, this sample domain contains the collection of pixels in a narrow band around the zero level set.

is equivalent to recovering the parameters $\mathbf{\Theta} = (\theta_1, \theta_2, ..., \theta_N)$ of a parametric transformation A, such that the mutual information between $f_\Omega = f(\Omega)$ and $g_\Omega^A = g\big(A(\mathbf{\Theta}; \Omega)\big)$ is maximized. The definition for such mutual information is:

$$MI(f_\Omega, g_\Omega^A) = \mathcal{H}\left[p^{f_\Omega}(l_1)\right] + \mathcal{H}\left[p^{g_\Omega^A}(l_2)\right] - \mathcal{H}\left[p^{f_\Omega, g_\Omega^A}(l_1, l_2)\right] \tag{1}$$

The terms in the above formula are: (i) l_1 and l_2 denote the intensity (distance value) random variables in the domains f_Ω and g_Ω^A respectively; (ii) $\mathcal{H}$ represents the differential entropy; (iii) p^{f_Ω} is the intensity probability density function (p.d.f.) in the source sample domain f_Ω; (iv) $p^{g_\Omega^A}$ is the intensity p.d.f. in the projected target domain g_Ω^A; and (v) p^{f_Ω, g_Ω^A} is their joint distribution.

This mutual information measures the general dependence between the target distance function and the transformed source distance function. It consists of three components: (i) the entropy of the source, $\mathcal{H}\left[p^{f_\Omega}(l_1)\right]$, (ii) the entropy of the projection of the source on the target given the transformation, $\mathcal{H}\left[p^{g_\Omega^A}(l_2)\right]$, and (iii) the joint entropy between the source and its projection on the target, $\mathcal{H}\left[p^{f_\Omega, g_\Omega^A}(l_1, l_2)\right]$. Maximizing this mutual information quantity encourages transformations where f_Ω statistically correlate with g_Ω^A.

We can further expand the formula in [Eq. 1] using the definition for differential entropy:

$$\mathcal{H}\left[p^{f_\Omega}(l_1)\right] = -\int_{\mathcal{R}^1} p^{f_\Omega}(l_1) log p^{f_\Omega}(l_1) dl_1 = -\iint_{\mathcal{R}^2} p^{f_\Omega, g_\Omega^A}(l_1, l_2) log p^{f_\Omega}(l_1) dl_1 dl_2 \tag{2}$$

$$\mathcal{H}\left[p^{g_\Omega^A}(l_2)\right] = -\int_{\mathcal{R}^1} p^{g_\Omega^A}(l_2) log p^{g_\Omega^A}(l_2) dl_2 = -\iint_{\mathcal{R}^2} p^{f_\Omega, g_\Omega^A}(l_1, l_2) log p^{g_\Omega^A}(l_2) dl_1 dl_2 \tag{3}$$

$$\mathcal{H}\left[p^{f_\Omega, g_\Omega^A}(l_1, l_2)\right] = -\iint_{\mathcal{R}^2} p^{f_\Omega, g_\Omega^A}(l_1, l_2) log p^{f_\Omega, g_\Omega^A}(l_1, l_2) dl_1 dl_2 \tag{4}$$

Combining [Eqs. 1, 2, 3 and 4], one can derive the criterion to perform global alignment using an arbitrary transformation model A with parameters $\mathbf{\Theta}$, by maximizing mutual information, which is equivalent to minimizing the following energy functional:

$$E_{Global}(A(\mathbf{\Theta})) = -MI(f_\Omega, g_\Omega^A) = -\iint_{\mathcal{R}^2} p^{f_\Omega, g_\Omega^A}(l_1, l_2) log \frac{p^{f_\Omega, g_\Omega^A}(l_1, l_2)}{p^{f_\Omega}(l_1) p^{g_\Omega^A}(l_2)} dl_1 dl_2 \tag{5}$$

The probability density functions in the energy functional are approximated using a nonparametric, differentiable Gaussian Kernel-based Density Estimation model. Using this model, the marginal probability

density functions are:

$$p^{f_\Omega}(l_1) = \frac{1}{V(\Omega)} \iint_\Omega G(l_1 - \underbrace{f(\mathbf{x})}_{\alpha}) d\mathbf{x} \tag{6}$$

$$p^{g_\Omega^A}(l_2) = \frac{1}{V(\Omega)} \iint_\Omega G(l_2 - \underbrace{g(A(\mathbf{\Theta}; \mathbf{x}))}_{\beta}) d\mathbf{x} \tag{7}$$

Where $\mathbf{x} = (x, y)$ refer to pixels in the sample domain Ω, $V(\Omega)$ represents the volume of the sample domain, and $G(a)$ represents the value of a one dimensional zero-mean Gaussian kernel at location a:

$$G(a) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{a^2}{2\sigma^2}}$$

where σ is a small constant controlling the kernel width (we set $\sigma = 4$ in all experiments).

Similarly, we can derive the expression for the joint probability density function using a two dimensional zero-mean Gaussian kernel:

$$p^{f_\Omega, g_\Omega^A}(l_1, l_2) = \frac{1}{V(\Omega)} \iint_\Omega G(l_1 - \underbrace{f(\mathbf{x})}_{\alpha}, l_2 - \underbrace{g(A(\mathbf{\Theta}; \mathbf{x}))}_{\beta}) d\mathbf{x} \tag{8}$$

where the 2D kernel $G(a, b)$ is given by:

$$G(a, b) = \frac{1}{2\pi\sigma_1\sigma_2} e^{-\frac{1}{2}(\frac{a^2}{\sigma_1^2} + \frac{b^2}{\sigma_2^2})}$$

and σ_1, σ_2 are the constants specifying the kernel widths in 2D.

The calculus of variations with a gradient descent method can now be used to minimize the cost function E_{Global} and recover the transformation parameters $\theta_i, i = 1, ..., N$. The parameter evolution equations are derived as follows:

$$\frac{\partial E_{Global}}{\partial \theta_i} = -\frac{1}{V(\Omega)} \iint_\Omega \Big[\iint_{\mathcal{R}^2} \Big(1 + log \frac{p^{f_\Omega, g_\Omega^A}(l_1, l_2)}{p^{f_\Omega}(l_1) p^{g_\Omega^A}(l_2)}\Big) \Big(- G_\beta(l_1 - \alpha, l_2 - \beta)\Big) dl_1 dl_2 \Big] \Big(\nabla g(A(\mathbf{\Theta}; \mathbf{x})) \cdot \frac{\partial}{\partial \theta_i} A(\mathbf{\Theta}; \mathbf{x})\Big) d\mathbf{x} \tag{9}$$

By substituting the general parameters $\mathbf{\Theta}$ with specific transformation parameters, the method supports registration using any global transformation model for shapes in 2D/3D.

Examples of such global alignment for 2D shapes using the similarity transformation model are given in [Fig. (2).1-4]; 2D examples using the affine transformation are shown in [Fig. (2).4]; 3D examples using the similarity transformation are shown in [Fig. (13).1] and [Fig. (14).b-c].

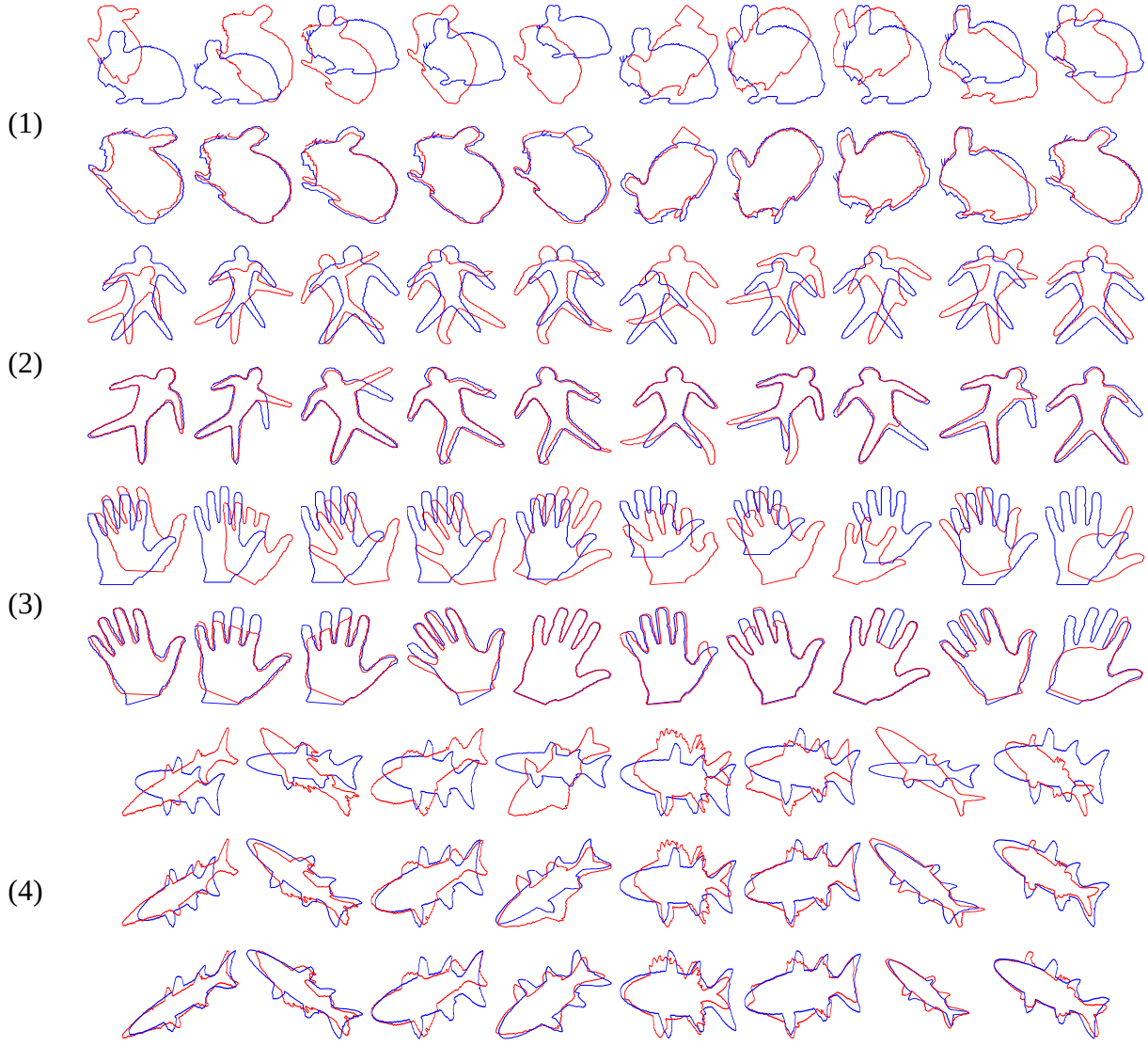
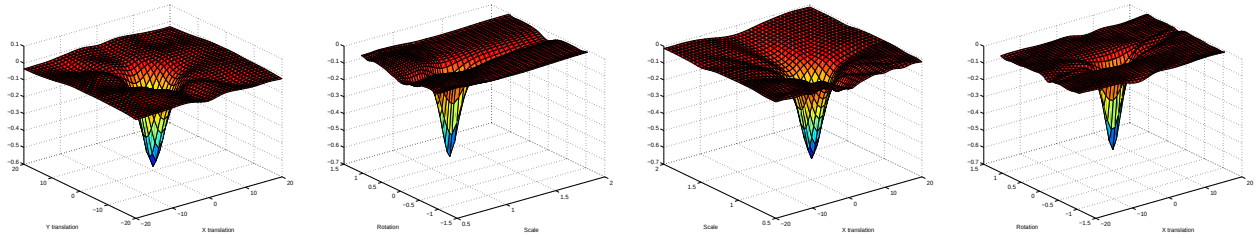


Fig. 2. Global registration examples. (1) Bunny, (2) Dude, (3) Hand, (4) Fish. (odd rows) Initial conditions (source in blue vs. target in red), (even rows) Alignment result using the similarity transformation model, (last row) Alignment result using the Affine transformation. Each column corresponds to a different trial. Only the zero level sets of the registered distance functions are shown in contour form.

1) Empirical Evaluation of the Global Criterion: Gradient descent optimization techniques often suffer from being sensitive to the initial conditions. The form of the objective function is a good indicator regarding the efficiency and stability of an optimization framework.

In order to perform a study on the performance of our global registration technique, we take the 2D similarity transformation model with four parameters: translations in x and y directions respectively, isotropic scale factor and the 2D rotation angle. Then we constrain the unknown parameter space in two dimensions, and empirically evaluate the form of the global registration objective function. For an example



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

Fig. 3. Empirical validation of global registration (a) Translations in x, y directions unknown, (b) Scale and rotation unknown, (c)Translation in x and scale unknown, (d) Translation in x and rotation unknown.

"dude" shape [Fig. (2).2], we have studied the following four cases: (1) translations in x, y directions are unknown [Fig. (3).a], (2) scale and rotation are unknown [Fig. (3).b], (3) translation in x and scale are unknown [Fig. (3).c], and (4) translation in x and rotation are unknown [Fig. (3).d]. In each case, we quantized the search space using a uniform sampling rule (100 elements) for all unknown parameters. Translations in (x, y) were in the range of $[-20, 20] \times [-20, 20]$, scale was in $[0.5, 2.0]$ and rotation in $\left[-\frac{\pi}{3}, \frac{\pi}{3}\right]$. Then, one can estimate the projections of the objective function in the space of two unknown parameters, by considering all possible combinations derived from the sampling strategy (the other two parameters are fixed). The resulting projections of the functional, as shown in [Fig. (3).a-d], have some nice properties: they are smooth and exhibit a single global minimum. Hence the objective function has a convex form for all combinations that involve two unknown registration variables and this is a good indicator for a well-behaved optimization criterion with smooth convergence properties.

C. Free Form Local Registration and Correspondences

Global registration can be an acceptable solution to a large number of computer vision applications. Medical imaging is an area where quite often global motion is not a valid answer when solving the dense registration and correspondences problem [28]. Local deformations are a complementary component to the global registration model. However, dense local motion (warping fields) estimation is an ill-posed problem since the number of variables to be recovered is often larger than the number of available constraints. Smoothness as well as other forms of constraints were employed to cope with this limitation.

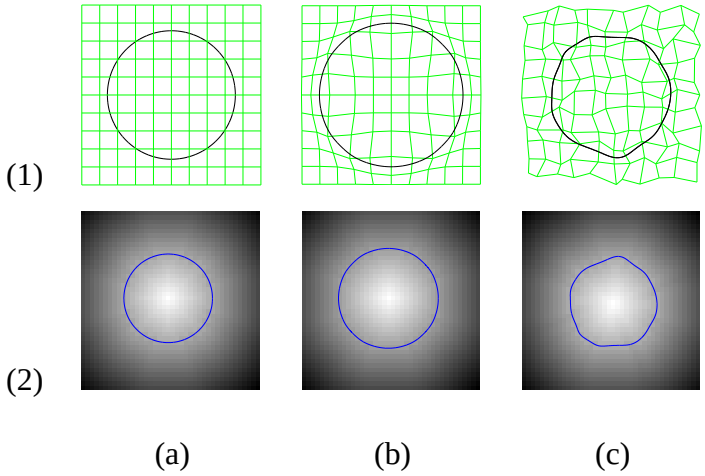


Fig. 4. Shape deformations based on FFD. (1) The deforming shapes. (2) The implicit representations for the shapes. (a) The initial shape. (b) Example FFD control lattice deformation to expand the shape. (c) Another FFD control lattice deformation to deform the shape in a free-form manner.

In the proposed framework, a global transformation A is recovered using the mutual information criterion. One can use such transformation to transform the source shape $\mathcal{D}$ to a new shape $\hat{\mathcal{D}} = A(\mathcal{D})$. Then, on top of this global registration result, local registration is equivalent to recovering a pixel-wise local deformation field that creates correspondences between the implicit representation $[\Phi_{\mathcal{S}}]$ of the target shape $\mathcal{S}$ and the implicit representation $[\Phi_{\hat{\mathcal{D}}}]$ of the transformed source shape $\hat{\mathcal{D}}$. Such a local deformation field $L(\mathbf{x})$ can be represented using space warping models such as Free Form Deformations [17], a popular approach in graphics, animation and rendering [29], [30], [31]. Contrary to optical flow techniques [3], FFD is able to implicitly enforce smoothness constraints, exhibit robustness to noise, and is suitable for modeling large and small non-rigid deformations via a multi-level approach. Furthermore, the recovered deformation field is smooth, continuous, preserves shape topology and guarantees a one-to-one mapping.

The essence of FFD is to deform the shape of an object by manipulating a regular control lattice P overlaid on its volumetric embedding space. The deformation of the control lattice consists of the displacements of all the control points in the lattice, and from these sparse displacements, a dense deformation field for every pixel in the embedding space can be acquired through interpolation using an interpolating basis function, such as Bezier spline or B-spline functions. One illustrative example is shown in [Fig. (4)]. A circular shape [Fig. (4).1.a] is implicitly embedded as the zero level set of a

distance function [Fig. (4).1.b]. A regular control lattice (drawn in green) is overlaid on this embedding space. When the embedding space deforms due to the deformation of the FFD control lattice as shown in [Fig. (4).b], the shape undergoes an expansion in its object-centered coordinate system. [Fig. (4).c] shows another example of free-form shape deformation given a particular FFD control lattice deformation.

In this paper, we consider an Incremental Free Form Deformations (IFFD) formulation using the cubic B-spline basis function for interpolation. To this end, dense registration is achieved by incrementally evolving a control lattice P according to a deformation improvement $[\delta P]$, and the inference problem is solved by minimizing a Sum-of-Squared-Differences criterion, with respect to the control lattice deformation improvements, which are the parameters of IFFD. The formulation of the local registration framework is presented in the following several subsections.

1) Local Deformation Representation: Let us consider a lattice of control points

$$P = \{P_{m,n}\} = \{(P^x_{m,n}, P^y_{m,n})\};\ m = 1, ..., M, \quad n = 1, ..., N$$

overlaid to a region $\Gamma_c = \{\mathbf{x}\} = \{(x, y)|l_x \leq x \leq h_x, l_y \leq y \leq h_y\}$ in the embedding space that encloses the source shape. Let us denote its initial regular configuration with no deformation as P^0 (e.g., [Fig. (4).1]), and the deforming configuration as $P = P^0 + \delta P$. Then the IFFD parameters are the deformation improvements of the control points in both x and y directions:

$$\mathbf{\Theta} = \delta P = \{(\delta P^x_{m,n}, \delta P^y_{m,n})\};\ (m, n) \in [1, M] \times [1, N] \tag{10}$$

Suppose the control lattice deforms from P^0 to P, the deformed position of any pixel $\mathbf{x} = (x, y)$ in the embedding space is defined by a tensor product of cubic B-splines:

$$L(\mathbf{x}) = \sum_{k=0}^{3} \sum_{l=0}^{3} B_k(u) B_l(v) P_{i+k,j+l} \tag{11}$$

where $i = \lfloor \frac{x - l_x}{h_x - l_x} \cdot (M - 1) \rfloor,\ j = \lfloor \frac{y - l_y}{h_y - l_y} \cdot (N - 1) \rfloor$. This is the familiar definition for cubic B-spline based interpolation, and the terms in the formula refer to:

1) $P_{i+k,j+l}$, $(k, l) \in [0, 3] \times [0, 3]$ are the coordinates of the sixteen control points in the neighborhood of pixel $\mathbf{x}$.

2) $B_k(u)$ represents the k^{th} basis function of cubic B-spline:

$$B_0(u) = (1-u)^3/6, \qquad B_1(u) = (3u^3 - 6u^2 + 4)/6$$

$$B_2(u) = (-3u^3 + 3u^2 + 3u + 1)/6, \qquad B_3(u) = u^3/6$$

with $u = \frac{x-l_x}{h_x-l_x} \cdot (M-1) - \lfloor \frac{x-l_x}{h_x-l_x} \cdot (M-1) \rfloor$.

$B_l(v)$ is similarly defined, with $v = \frac{y-l_y}{h_y-l_y} \cdot (N-1) - \lfloor \frac{y-l_y}{h_y-l_y} \cdot (N-1) \rfloor$.

According to our IFFD formulation $P = P^0 + \delta P$, we can re-write [Eq. 11] in terms of the IFFD parameters $\mathbf{\Theta} = \delta P$:

$$\begin{aligned} L(\mathbf{\Theta}; \mathbf{x}) &= \sum_{k=0}^{3} \sum_{l=0}^{3} B_k(u) B_l(v) (P^0_{i+k,j+l} + \delta P_{i+k,j+l}) \\ &= \sum_{k=0}^{3} \sum_{l=0}^{3} B_k(u) B_l(v) P^0_{i+k,j+l} + \sum_{k=0}^{3} \sum_{l=0}^{3} B_k(u) B_l(v) \delta P_{i+k,j+l} \end{aligned} \tag{12}$$

Based on the linear precision property of B-splines, a B-spline curve through collinear control points is itself linear, hence the initial regular configuration of control lattice P^0 generates the undeformed shape and its embedding space, i.e., for any pixel $\mathbf{x}$ in the sampling domain, we have:

$$\mathbf{x} = \sum_{k=0}^{3} \sum_{l=0}^{3} B_k(u) B_l(v) P^0_{i+k,j+l} \tag{13}$$

where i, j are derived the same way as in [Eq. 11].

Now combining [Eq. 12] and [Eq. 13], we have:

$$L(\mathbf{\Theta}; \mathbf{x}) = \mathbf{x} + \delta L(\mathbf{\Theta}; \mathbf{x}) = \mathbf{x} + \sum_{k=0}^{3} \sum_{l=0}^{3} B_k(u) B_l(v) \delta P_{i+k,j+l} \tag{14}$$

Compared to the traditional FFD, the IFFD formulation above simplifies the integration of smoothness constraints (see section III-C.2), and accounts for an efficient multi-level approach (see section III-C.3) to deal with both large and small local non-rigid deformations.

2) Local Registration Optimization Criterion and Gradient Descent: Under these local deformation definitions, local registration is equivalent to finding the control lattice deformation δP such that when applied to the embedding space of the source shape, the deformed source shape coincides with the target shape. Since the structures to be locally registered in our framework are the distance transform of the

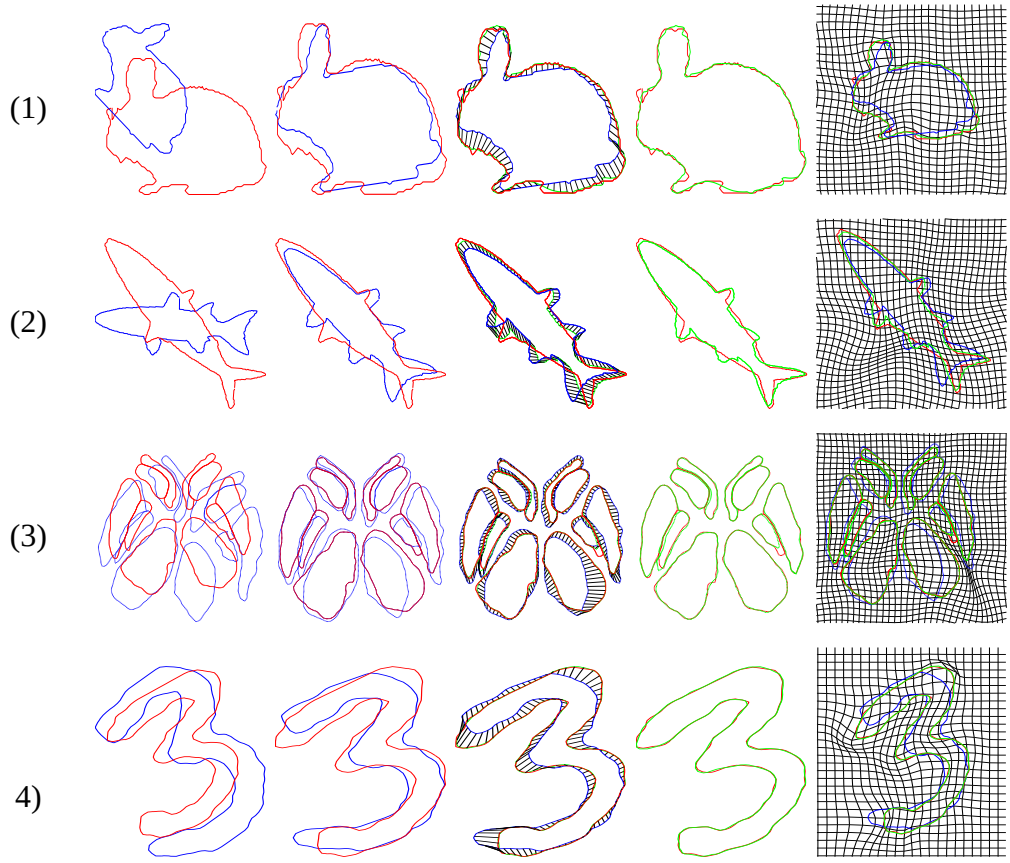


Fig. 5. Incremental B-spline FFD local registration. (1) Bunny, (2) Fish, (3) Brain Structure, (4) Digit 3. (a) Initial conditions (source shape in blue, target shape in red), (b) Result after global registration, (c) Established correspondences after local registration; only the zero level set (i.e., shape) correspondences are shown, (d) Locally deformed source shape (in green) overlaid on the target (in red), (e) Final IFFD control lattice configuration depicting the space warping to achieve local registration.

target shape - $[\Phi_{\mathcal{S}}]$, and the distance transform of its globally aligned source shape - $[\Phi_{\hat{\mathcal{D}}}]$, the Sum-of-Squared-Differences (SSD) criterion can be considered as the data-driven term to recover the deformation field $L(\mathbf{\Theta};\mathbf{x}))$:

$$E_{data}(\mathbf{\Theta}) = \iint_{\Omega} \left(\Phi_{\hat{\mathcal{D}}}(\mathbf{x}) - \Phi_{\mathcal{S}}(L(\mathbf{\Theta};\mathbf{x}))\right)^2 d\mathbf{x} \tag{15}$$

In order to further preserve the regularity of the recovered registration flow field, one can consider an additional smoothness term on the local deformation field δL. We consider a computationally efficient smoothness term:

$$E_{smoothness}(\mathbf{\Theta}) = \iint_{\Omega} \left(\left|\left| \frac{\partial \delta L(\mathbf{\Theta};\mathbf{x})}{\partial x} \right|\right|^2 + \left|\left| \frac{\partial \delta L(\mathbf{\Theta};\mathbf{x})}{\partial y} \right|\right|^2 \right) d\mathbf{x} \tag{16}$$

Such a smoothness term is based on an error norm with known limitations. One can replace this smoothness component with more elaborate norms. Within the proposed framework, an implicit smoothness constraint is also imposed by the B-Spline FFD, which guarantees C^1 continuity at control points and C^2 continuity everywhere else. Therefore there is no need for introducing complex and computationally expensive regularization components.

The data-driven term [Eq. 15] and the smoothness term [Eq. 16] can now be integrated into one energy functional to recover the IFFD parameters:

$$E_{Local}(\mathbf{\Theta}) = \iint_{\Omega} \left(\Phi_{\hat{\mathcal{D}}}(\mathbf{x}) - \Phi_{\mathcal{S}}(L(\mathbf{\Theta};\mathbf{x}))\right)^2 d\mathbf{x} + \alpha \iint_{\Omega} \left(\left|\left| \frac{\partial \delta L(\mathbf{\Theta};\mathbf{x})}{\partial x} \right|\right|^2 + \left|\left| \frac{\partial \delta L(\mathbf{\Theta};\mathbf{x})}{\partial y} \right|\right|^2 \right) d\mathbf{x} \quad (17)$$

where α is the constant balancing the contribution of the two terms. In our experiments, the typical values for α are in the range of $1 \sim 5$; smaller values lead to faster convergence, while larger values result in smoother deformation fields. The one-to-one mapping property is guaranteed regardless of the α value.

The calculus of variations and a gradient descent method can be used to optimize the local registration objective function [Eq. 17]. One can obtain the following evolution equation for each parameter θ_i in the IFFD control lattice deformation parameters $\mathbf{\Theta}$:

$$\begin{aligned} \frac{\partial}{\partial \theta_i} E_{Local}(\mathbf{\Theta}) = & -2 \iint_{\Omega} \left(\Phi_{\hat{\mathcal{D}}}(\mathbf{x}) - \Phi_{\mathcal{S}}(L(\mathbf{\Theta};\mathbf{x}))\right) \left(\nabla \Phi_{\mathcal{S}}(L(\mathbf{\Theta};\mathbf{x})) \cdot \frac{\partial}{\partial \theta_i} \delta L(\mathbf{\Theta};\mathbf{x})\right) d\mathbf{x} \\ & + 2\alpha \iint_{\Omega} \frac{\partial}{\partial x} \delta L(\mathbf{\Theta};\mathbf{x}) \cdot \frac{\partial}{\partial \theta_i} \left(\frac{\partial}{\partial x} \delta L(\mathbf{\Theta};\mathbf{x}) \right) + \frac{\partial}{\partial y} \delta L(\mathbf{\Theta};\mathbf{x}) \cdot \frac{\partial}{\partial \theta_i} \left(\frac{\partial}{\partial y} \delta L(\mathbf{\Theta};\mathbf{x}) \right) d\mathbf{x} \end{aligned} \quad (18)$$

The partial derivatives in the above formula can be easily derived from the model deformation equation in [Eq. 14]. Details are given in the Appendix.

Once the optimal IFFD parameters and the local registration field $\hat{L}$ are derived, dense one-to-one correspondences can be established between each point $\mathbf{x} = (x, y)$ on the source structure, with its deformed position $\hat{L}(\mathbf{x})$ on the target structure. These correspondences include not only the correspondences for those pixels located on the zero level set, which are points on the source and target shapes, but also correspondences between nearby level sets which are clones of the original shapes coherently positioned in the embedding image/volume space.

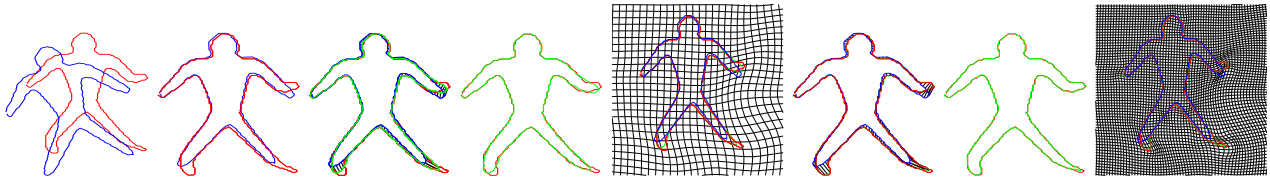
The performance of the proposed local registration paradigm is demonstrated for various 2D examples shown in [Figs. (5, 1)]. 3D examples are presented in Sec. IV-B (see [Figs. (13, 14)]).

3) Multi-resolution IFFD Control Lattices: To account for both large-scale and highly local non-rigid deformations, we can use an efficient multi-level implementation of the IFFD framework, as shown in [Fig. (6)]. To this end, multi-resolution control lattices are used according to a coarse-to-fine strategy. A coarser level control lattice is applied first to account for relatively global non-rigid deformations; then the space deformation resulting from the coarse level registration is used to initialize the configuration of a finer resolution control lattice, and at this finer level, the local registration process continues to deal with highly local deformations and achieve better matching between the deformed source shape and the target. Generally speaking, the hierarchy of control lattices can have arbitrary number of levels, but typically $2 \sim 3$ levels are sufficient. The layout of the control lattices in the hierarchy can be computed efficiently using a progressive B-spline subdivision algorithm [32]. At each level, we can solve for the incremental deformation of the control lattice using the scheme presented in section III-C.2. In the end, the overall deformation field is defined by the incremental deformations from all levels. In particular, the total deformation $\delta L(\mathbf{x})$ for a pixel $\mathbf{x}$ in a hierarchy of r levels is:

$$\delta L(\mathbf{x}) = \sum_{k=1}^{r} \delta L^k(\Theta^k; \mathbf{x}) \tag{19}$$

where $\delta L^k(\Theta^k; \mathbf{x})$ refers to the deformation improvement at this pixel due to the incremental deformation Θ^k of the kth level control lattice.

4) Incorporating Feature Point Constraints: The non-rigid registration framework we presented above operates directly on the distance values in the shape-embedding distance map images, without prior feature extraction. Because of the strong smoothness and coherence constraints, the resulting local registration field is continuous, smooth and guarantees a one-to-one mapping. Despite these properties, however, one can argue that this approach is optimal only when registering shapes that do not have distinct features. If the pair of shapes being registered does have distinct geometric feature correspondences, incorporating these feature information can greatly improve accuracy and efficiency [33].



(a) (b) (c) (d) (e) (f) (g) (h)

Fig. 6. Multi-level Incremental FFD for local registration. (a) Initial Condition, (b) After global registration, (c) Established correspondences using a coarse resolution IFFD control lattice for local registration, (d) Coarse resolution matching result, (e) Coarse resolution control lattice (space) deformation, (f) Refined correspondences by a finer resolution IFFD control lattice, (g) Finer resolution matching result, (h) Finer resolution control lattice (space) deformation.

When feature points and correspondences are available, these constraints can be conveniently integrated into our registration framework. Assuming the total number of features is n_c, and for each feature, there is a pair of corresponding points, $\mathbf{x}_{\mathcal{S}_i}$ $(i = 1, \ldots, n_c)$ on the target shape $\mathcal{S}$ and $\mathbf{x}_{\hat{\mathcal{D}}_i}$ on the globally transformed source shape $\hat{\mathcal{D}}$, then the following energy term incorporates the feature constraints:

$$E_{feature}(\mathbf{\Theta}) = \sum_i \left(L(\mathbf{\Theta}; \mathbf{x}_{\hat{\mathcal{D}}_i}) - \mathbf{x}_{\mathcal{S}_i}\right)^2; \quad i \in [1, n_c] \tag{20}$$

and its derivatives with respect to the IFFD deformation parameters are:

$$\begin{aligned} \frac{\partial E_{feature}(\mathbf{\Theta})}{\partial \theta_i} &= 2 \textstyle\sum_i \left[\left(L(\mathbf{\Theta}; \mathbf{x}_{\hat{\mathcal{D}}_i}) - \mathbf{x}_{\mathcal{S}_i}\right) \cdot \frac{\partial}{\partial \theta_i}\left(L(\mathbf{\Theta}; \mathbf{x}_{\hat{\mathcal{D}}_i})\right)\right] \\ &= 2 \textstyle\sum_i \left[\left(L(\mathbf{\Theta}; \mathbf{x}_{\hat{\mathcal{D}}_i}) - \mathbf{x}_{\mathcal{S}_i}\right) \cdot \frac{\partial}{\partial \theta_i}\left(\delta L(\mathbf{\Theta}; \mathbf{x}_{\hat{\mathcal{D}}_i})\right)\right] \end{aligned} \tag{21}$$

The feature term [Eq. 20], multiplied by a balancing constant β, can be added to the overall energy function [Eq. 17], and the derivative terms can be adjusted accordingly.

We show an example that uses feature point constraints in [Fig. (7)]. The goal is to register two humanoid shapes [Fig. (7).1.a]. Due to part-based deformations, the two shapes differ significantly, especially in the arms. Such large deformations make direct optimization of the energy functional in [Eq. 17] prone to get stuck in local minima [Fig. (7).2]. Hence incorporating a few feature correspondences plays a key role in correct local registration. We manually specified three feature points on the target shape and their correspondences on the source shape [Fig. (7).1.b-c]). Since we have high confidence in the accuracy of the features, we set a high value (usually between $5 \sim 10$) for the weight factor of the feature term, β,

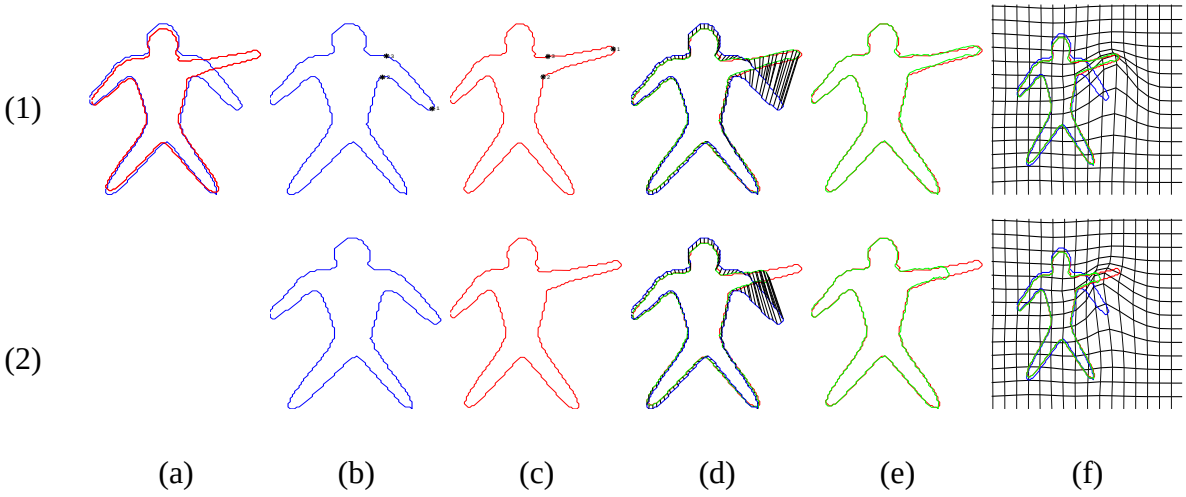
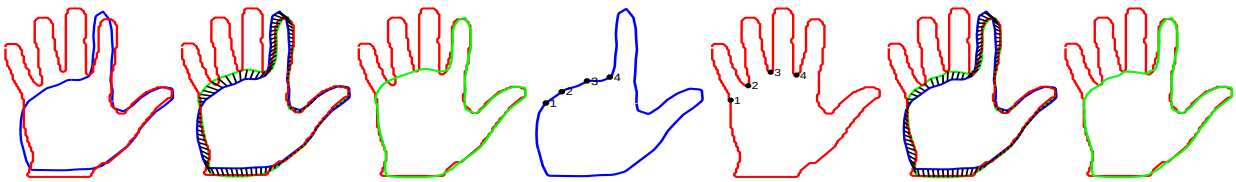


Fig. 7. Feature point constraints in the presence of large local deformations. (1) Local FFD registration with feature constraints. (2) For comparison purposes, local registration without feature constraints. (a) Two humanoid shapes rigidly aligned. (1.b & 1.c) source (in blue) and target (in red) shapes with features. The features are marked by asterisks and corresponding ones are marked by the same number. (2.b & 2.c) source and target shapes with no feature constraints specified. (d) Dense correspondences established by local FFD registration. (e) Locally deformed source shape (in green) overlaid on the target (in red). (f) Final IFFD control lattice configuration to achieve local registration.

during optimization. The registration results with vs. without feature constraints can be seen in [Fig. (7).1] and [Fig. (7).2] respectively.

Specifying feature point constraints also helps in the situation where one shape has spurious (or outlier) structures that can not be registered to the other shape. Although the mutual information based global registration can handle a great deal of outliers in the data (see [Fig. (2)]), the behavior of local registration is often ill-defined in the presence of significant spurious structures. One such example is shown in the rightmost panel in [Fig. (2).3], where one of the hand shapes has 4 fingers missing. In our local registration framework, without incorporating any prior knowledge or feature point constraints, the local deformation of the source shape stops when the smoothness term (Eq. 16) and the data term (Eq. 15) reach an equilibrium; this result is shown in [Fig. (8).b-c]. On the other hand, much more accurate and meaningful local registration can be achieved if we incorporate human knowledge about the shapes, for instance, the knowledge that the shapes are hand shapes and one shape has 4 fingers missing. This piece of knowledge can be integrated by hand-labeling corresponding feature points and adding the feature term (Eq. 20) in



(a) (b) (c) (d) (e) (f) (g)

Fig. 8. Local registration in the presence of spurious structures and outliers. (a) Two hand shapes rigidly aligned. There are 4 fingers missing on the source shape. (b-c) local registration without any feature constraints; (b) dense correspondences established by local FFD registration, (c) locally deformed source shape overlaid on the target. (d-e) source and target shapes with manually specified feature correspondences. (f-g) local registration result with the feature correspondences constraints.

the energy minimization process. In [Fig. (8).d-g], we show the local registration result with 4 pairs of corresponding feature points.

The incorporation of good feature constraints can generally improve registration accuracy and help to overcome local minima in the gradient-descent based optimization process. When manually specified feature correspondences are not easily available however, methods that automatically determine good feature correspondences [34], [6] can be considered. Particularly in our framework, an attractive approach is to use the distance map representation locally around each shape point to derive certain shape descriptive signatures, for instance the Gaussian Curvatures as in [34], then these signatures can be used to match points on the source and the target shapes and find feature correspondences.

D. Summary of the Global-to-Local Shape Registration Framework

In summary, our overall global-to-local shape registration protocol is as follows. Given a source and a target shape, the implicit representations for both shapes are derived. Then, parameters of a global transformation (such as similarity, affine) are recovered by maximizing mutual information between the source and target implicit representations. The source shape is transformed according to the global transformation, and further local registration between the transformed source shape and the target shape is done by solving for the IFFD parameters in the local registration framework. During local registration, one can also choose to use multi-resolution IFFD control lattices, or to incorporate explicit feature point

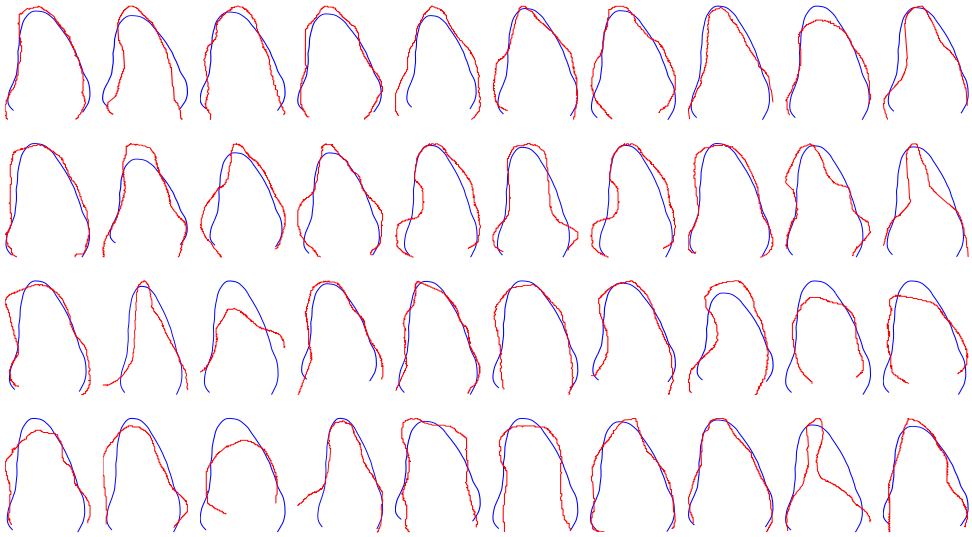


Fig. 9. Rigid Registration for User-Determined Ground Truth (Systole) shapes of the Left Ventricle from Ultrasonic Images (multiple views). (blue) target mean shape, (red) registered source shape.

constraints, or both, to achieve better registration accuracy.

IV. Applications

We present two applications of our registration framework to demonstrate its potential.

A. Statistical Modeling of Anatomical Structures

Organ modelling is a critical component of medical image analysis. To this end, one would like to learn a compact representation that can capture the variation in an anatomical structure of interest across individuals. Building such representation requires establishing dense local correspondences across a set of training examples. The registration framework proposed in this paper can be use to solve the dense correspondence problem [26].

As an example, we show the statistical modelling of systolic left ventricle (LV) shapes from ultrasonic images, using 40 pairs of hand-drawn LV contours. We first apply global rigid registration to align all contours to the same target, as shown in [Fig. (9)]. Local registration based on free form deformations is then used to non-rigidly register all these contours to the common target (see grid deformations in [Fig. (10)]). In order to establish dense one-to-one correspondences between all the aligned contours, we pick

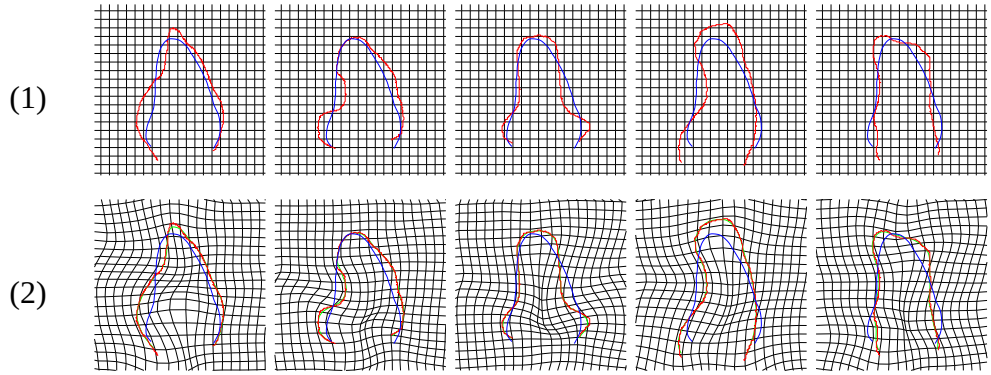


Fig. 10. Local non-rigid registration using IFFD. (1) initial un-deformed grid overlaid on global rigid registration result (blue - mean reference shape), (2) deformed grid to map the reference shape to various training shapes. Each column corresponds to a different trial.

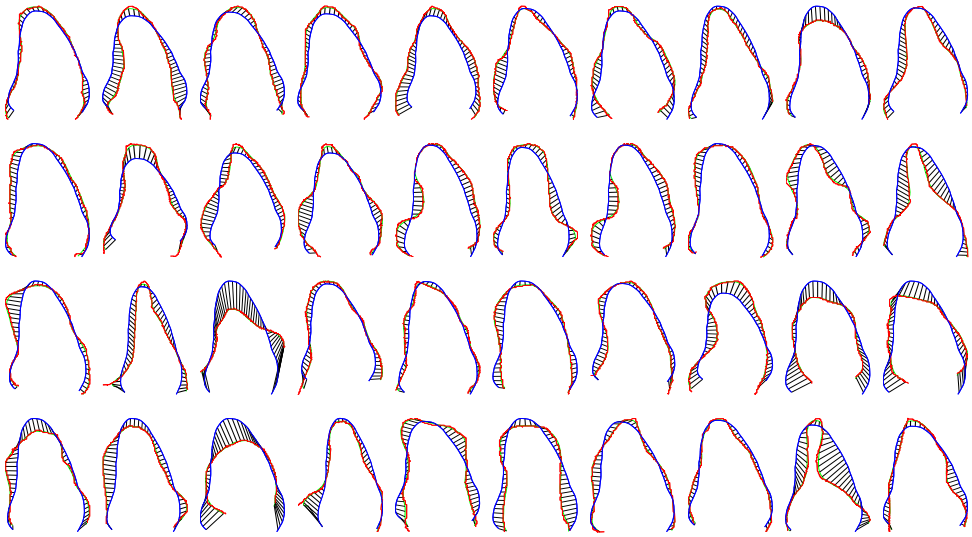


Fig. 11. Established correspondences using IFFD. (red) source shapes after global transformations, (blue) target mean shape, (dark lines) correspondences for a fixed set of points on the mean shape.

a set of sample points on the common target and compute their correspondences on each training contour based on the local registration result (see [Fig. (11)] for established local correspondences).

Using the established correspondences, the Principal Component Analysis (PCA) technique can be applied to build a Point Distribution Model (PDM) [2] to capture the statistics of the corresponding elements across the training examples. The resulting principal components for the statistical model of the set of systolic left ventricle shapes can be seen in [Fig. (12)]. The model captures the variations in the training set well, and generates new shapes that are consistent with the training examples. This justifies the validity of the established correspondences to some extent.

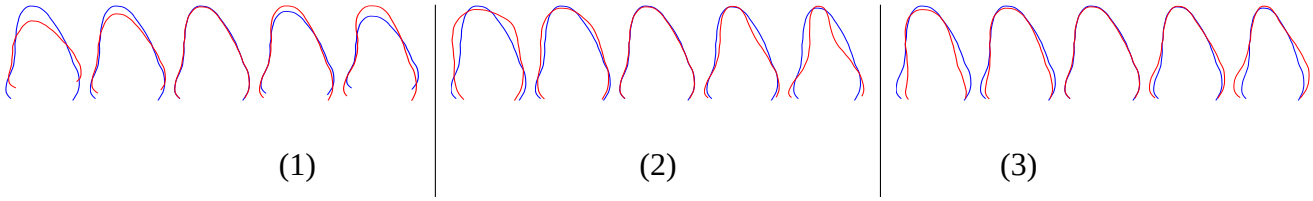


Fig. 12. PCA modelling for the systolic Left Ventricle shapes using the established local correspondences. (1) first mode, (2) second mode, (3) third mode; For each mode, from left to right shows the mode changing from $-2\sqrt{\lambda_i}$ to $2\sqrt{\lambda_i}$.

B. 3D Face Registration

The proposed framework and the derivations can be naturally extended to 3D. For global registration, parameters of a 3D transformation model can be solved by maximizing mutual information in the 3D sample domain; for local registration, free form deformations can be defined by the 3D tensor product of B-spline polynomials, and the SSD energy functional is defined in the 3D volumetric domain. Geometric feature constraints can be specified in 3D as well to increase registration accuracy. More details on the 3D formulation for non-rigid FFD registration can be found in [35].

We can use the 3D registration framework to align, register, and stitch 3D face scans captured from range scanners. This problem plays an important role in face modelling, recognition, etc. We show one set of such registration result in [Fig. (13)]. The global transformation model consists of translation, scaling, and quaternion-based rotation [Fig. (13).1]. The local incremental FFD model uses control lattices in the 3D space and a 3D tensor product of B-spline polynomials. Qualitatively the result after global-to-local registration can be seen from two views: the front view [Fig. (13).2(front)], and the side view [Fig. (13).2(side)]. Quantitatively, the sum-of-squared-differences matching error (Eq. 15) after global registration was 8.3. The IFFD based local registration used three resolutions of control lattices in a coarse-to-fine manner and ran 20 iterations for each resolution. After the coarsest-level (10*10 lattice) IFFD local registration, the matching error was reduced to 3.4; after the middle level (20*20 lattice), the matching error was reduced to 1.8; and after the finest level (40*40 lattice), the matching error was reduced to 1.2. The total time spent for global and multi-level local registration was 4.6 minutes.

The registration framework can also be used for facial expression tracking in 3D. Recent technological

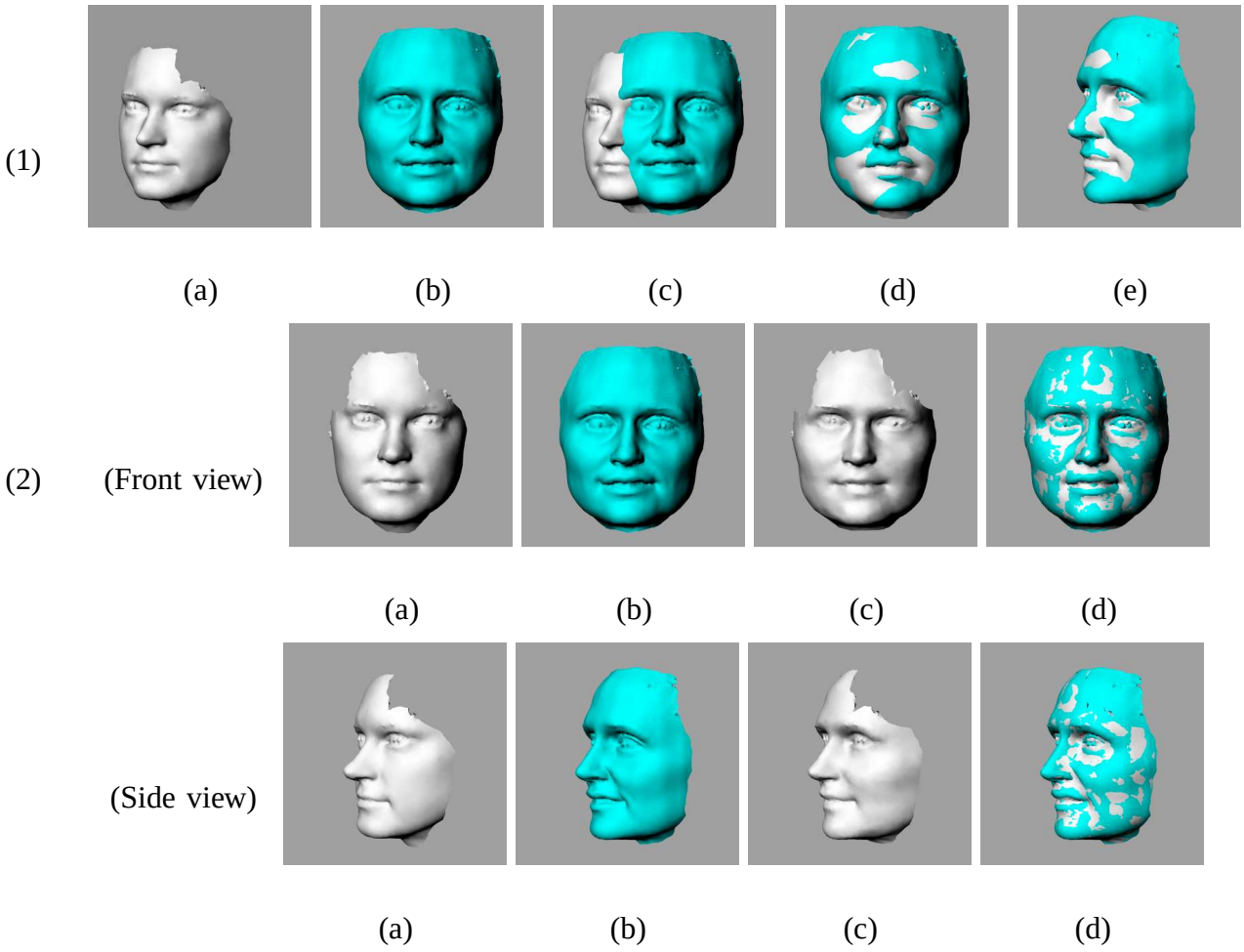


Fig. 13. Global-to-local registration for open 3D structures (both source and target shapes are from face range scan data). (1) Global registration using the 3D similarity transformation model: (a) source shape; (b) target shape; (c) initial pose of the source relative to the target; (d & e) globally transformed source shown overlaid on the target - front view (d) and side view (e). (2) Local registration using IFFD: (Front view & Side view): (a) source shape after rigid transformation; (b) target shape; (c) locally deformed source shape after IFFD registration; (d) locally deformed source shape shown overlaid on the target.

advances in digital projection, digital imaging and computers are making high resolution dynamic 3-D shape acquisition in real time increasingly available [36], [37]. Such high quality data are very attractive in the analysis of facial expressions as they can capture highly local and subtle details, such as wrinkles and furrows, in an expression. However, the point samples returned by these acquisition systems are not registered in object space, hence there is no guarantee of intra-frame correspondences, which would make tracking and temporal study of the expression difficult. Our registration framework provides a way to parameterize such a high amount of data and establish intra-frame correspondences. The basic idea is

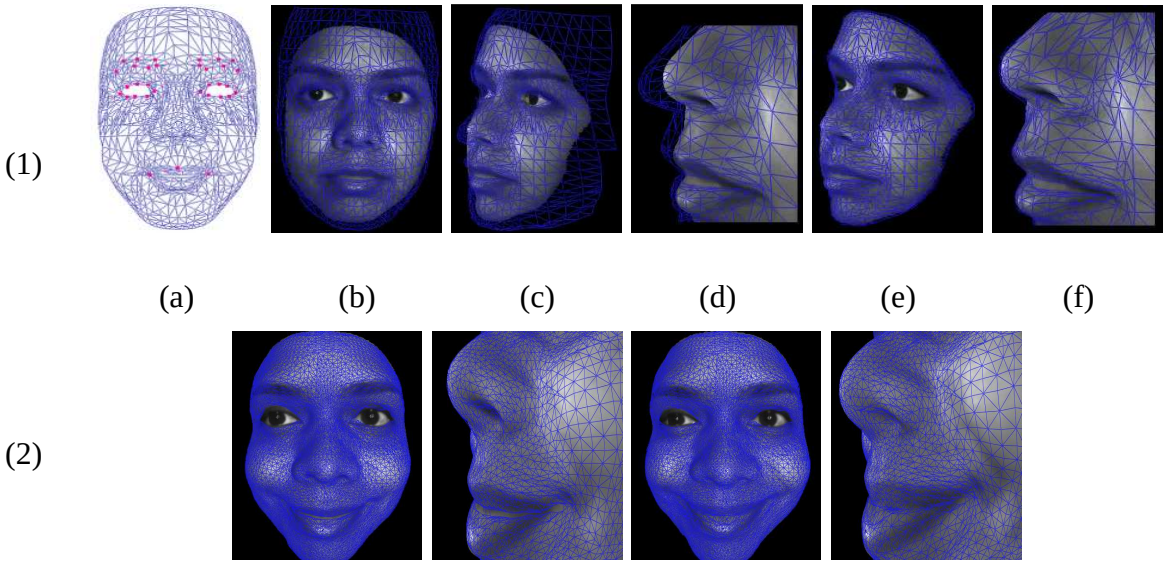


Fig. 14. Application to facial expression tracking using 3D dynamic face range scan data. (1) Registering a generic face mesh model to the first frame of a 3D sequence: (a) the mesh model, (b) front view, after global registration of the model (in blue) with the first frame, (c) side view after global registration, (d) close-up view after global registration, (e) side view, after local IFFD registration between the mesh model and the first frame, (f) close-up view of the local registration result. (2) Tracking result on a *smile* expression. The registered mesh model is superimposed on two intermediate frames. The 2nd column is the close-up view of the 1st column, and the 4th column is the close-up view of the 3rd column.

to register a generic face mesh model with the first frame of a dynamic sequence of face range scans, then keep deforming the mesh model to register it with the remaining frames in the sequence. Dense intra-frame correspondences are thus established between points on the dynamic range data that register to the same node (i.e. vertex) on the common mesh model. [Fig. (14)] shows some examples of the initial model fitting and subsequent tracking. The coarser mesh for initial fitting is subdivided to generate the dense mesh for tracking. Qualitatively, one can see from the results [Fig. (14).2] that our method is able to capture very fine details in the range data using a high resolution mesh model. Quantitatively, we perform a number of experiments on 3D facial expression sequences with attached markers to validate the accuracy of 3D tracking. The markers are for validation purposes only and are not used for tracking. In order to be detected successfully, the size of markers is around $4mm$ by $4mm$. An example tracking result is demonstrated in [Fig. (15).1], where the white dots are attached markers for validation purposes only. [Fig. (15).2] shows the algorithm's tracking errors of the markers on the mouth corner, upper mouth,

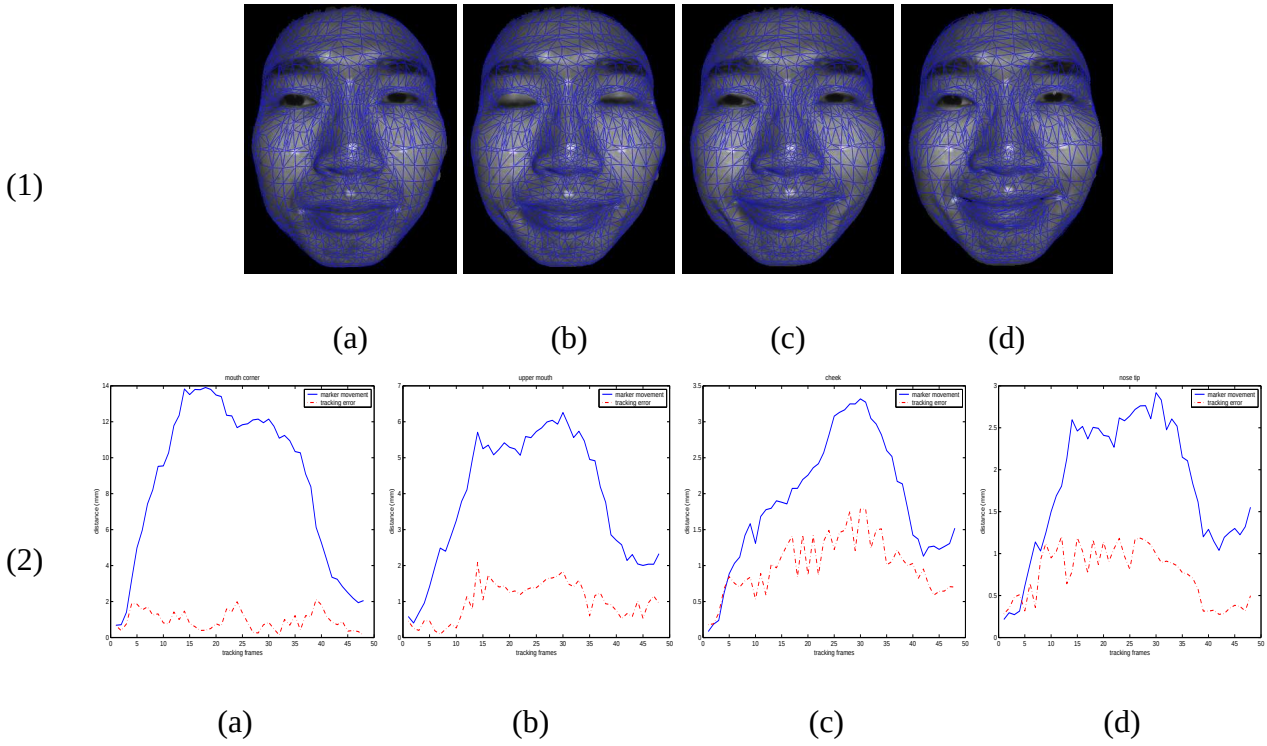


Fig. 15. Quantitative validation of 3D tracking error using the registration framework. (1) Selected tracking results of a 'smile' sequence, with 50 frames in total. The resulting meshes are illustrated in blue color and white dots are attached markers for verification purposes only. (1.a) frame 1, (1.b) frame 5, (1.c) frame 10, and (1.d) frame 37. (2) Tracking errors of the markers. (2.a) average tracking error of the markers on the mouse corners. (2.b) tracking error of the marker on the upper mouth. (2.c) average tracking error of the markers on the cheeks. (2.d) tracking error of the marker on the nose tip. In all axes, the X-axis denotes the frame number, and the Y-axis denotes the estimated error magnitude.

cheek, and nose tip respectively. From the error distributions, we can see that in most cases the tracking error in 3D is around $1mm$. This error is very low given that the resolution of the 3D range scan data is $0.5mm$ in X and Y directions and $0.2mm$ in Z direction.

V. Discussion and Evaluation

The proposed registration method is generic since it handles naturally shapes of arbitrary topology in any dimension. By using a narrow band around the source shape as the registration sample domain, the algorithm is also efficient. On a 1.5GHZ Pentium PC station, the average convergence time of our global rigid registration on two 2D shapes is about 50ms. We use a 2-level Cubic B-spline incremental FFD for local non-rigid registration, and it takes less than 30ms to converge for each level. This performance is

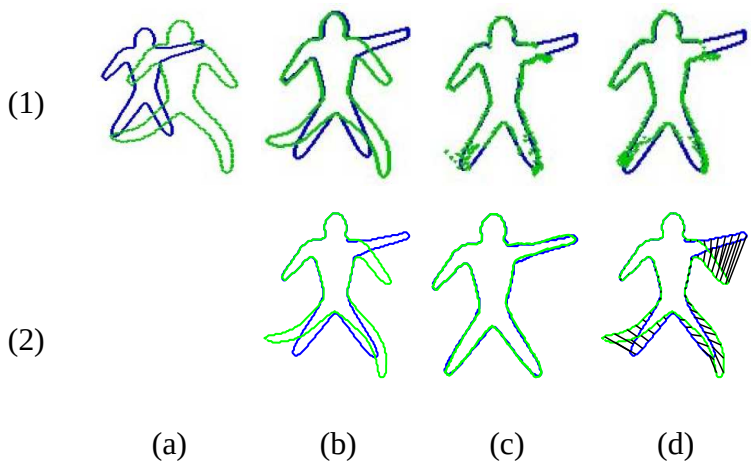


Fig. 16. Comparison with the non-rigid shape registration algorithm presented in [3]. (1) Results from the algorithm in [3]. (2) Results from our algorithm. (a) Initial poses of the source (in green) and target (in blue) shapes. (b) Alignment after global registration. (1.c) Local registration result; (1.d) local registration with regularization constraints. (2.c) Local registration result; (2.d) established correspondences.

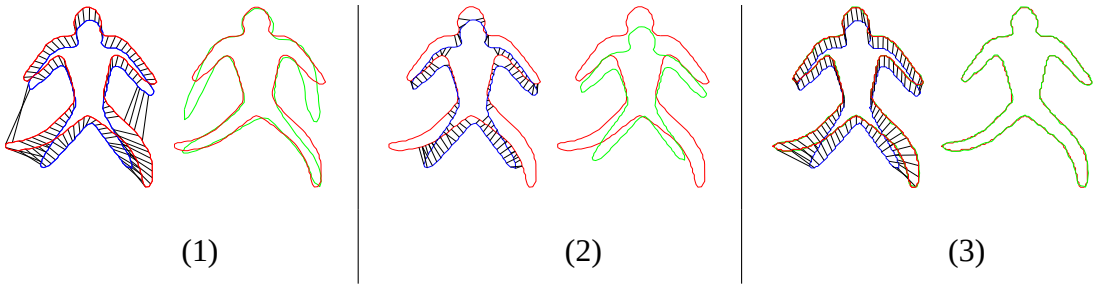


Fig. 17. Comparison with Shape Context (SC) and ICP algorithms. Source shape shown in blue and target shape in red. (1) Results from SC (first two columns), (2) Results from ICP (middle two columns), (3) Results from our method (last two columns). (odd columns) Correspondences established. (even columns) Registration result with final transformed source (green) overlaid on target (red).

comparable to the reported times of other recent approaches on shape matching/registration [6], [9], [3].

Besides computational efficiency, accuracy is another important performance measure of a registration algorithm. First, we compare the result of our algorithm to that presented in [3], which also uses the implicit shape representation for registration. Visual comparisons can be seen from [Fig. (16)]. The first row presents results from the method in [3], and the second row are results from our method. These results make clear the main advantage of our method, which is a more elegant local deformation model - IFFD. IFFD couples naturally with the implicit shape representation, and generates smooth and continuous non-rigid deformations that preserve shape topology [Fig. (16).2.c] and guarantee one-to-one coherent correspondences [Fig. (16).2.d]. On the other hand, the pixel-wise optical flow model used in [3] does not

guarantee preserving the topology and coherence of a shape after deformation [Fig. (16).1.c]. Even with advanced regularization and smoothness constraints, a closed shape can be deformed to become an open structure after registration [Fig. (16).1.d]. Second, we also compare the dense correspondences established by our algorithm to that by two other well-known shape matching/registration algorithms: Shape Context (SC) [6] and Iterative Closest Points (ICP) [4], [5]. (The SC code is from the original authors, and the ICP is based on our own implementation.). In both methods, the Thin Plate Spline (TPS) model is used to recover the non-rigid deformations based on the correspondences established by the methods. The comparison results are shown in [Fig. (17)]. In the figure, the black lines separate the results from different methods (group 1 - SC; group 2 - ICP; group 3 - our method). From the results, one can see that, due to insufficient smoothness/coherence constraints, some correspondences produced by SC and ICP algorithms are outliers and these outlier false matches degrade the registration result of these methods. On the other hand, our proposed method imposes smoothness constraints both along and around the shape boundary, thus producing good registration results with coherent one-to-one correspondences.

VI. Conclusion

In this paper, we have proposed a variational framework for global-to-local shape registration. Such a framework has been derived by integrating a powerful implicit shape representation with (i) a robust global registration technique (mutual information) and (ii) an elegant local deformation model (incremental free form deformations). The resulting paradigm can deal with arbitrary forms of global transformation [Fig. (2)], can efficiently solve the dense local correspondences problem [Fig. (1,5, 11, 14)]. Furthermore, it is shown to be relatively free from the initial conditions [Fig. (2, 3)]; it exhibits robustness to noise [Fig. (2).4], large local variations [Fig. (2).2], severe occlusions and missing parts [Fig. (2).3]; it handles structures of arbitrary topology, such as multi-components [Fig. (5).4] and open structures [Fig. (13, 11, 14)], and can be extended in higher dimensions [Fig. (13, 14)].

The extension of our framework to image and volume registration is the main future direction of our work. To this end, one can couple the registration of shape and surface structures with registration in the

intensity spaces. While the use of the SSD criterion is efficient and appropriate for local shape registration using the implicit representation, it cannot account for the complex image intensity dependencies in image/volume registration, hence it will have to be replaced by other multi-modal image matching criterions such as mutual information. For local registration, instead of IFFD that uses a regular control lattice, we can consider Free Form Deformation models that allow the placing of control points at arbitrary locations such as the Dirichlet Free Form Deformations [38]. It is also interesting to investigate a mathematical justification of the method in terms of existence and uniqueness of the solution within a limited parameter search space.

APPENDIX

It is straightforward to derive the partial derivatives (in section III-C.2) with respect to the incremental B-Spline Free Form Deformation parameters $\theta_i, i = 1, ..., M \times N$, which are:

$$\delta P_{m,n} = (\delta P^x_{m,n}, \delta P^y_{m,n});\ m = 1, ..., M, \quad n = 1, ..., N$$

Without loss of generality one can consider the $(m, n)th$ control point and its deformation in both directions. Then, for the data driven term the following relation holds:

$$\frac{\partial\ \delta L(\mathbf{\Theta}; \mathbf{x})}{\partial \delta P^x_{m,n}} = \begin{cases} \begin{bmatrix} B_{m-i}(u)\ B_{n-j}(v) \\ 0 \end{bmatrix}, 0 \leq m-i, n-j \leq 3 \\ \\ \mathbf{0}, \qquad otherwise \end{cases}$$

$$\frac{\partial\ \delta L(\mathbf{\Theta}; \mathbf{x})}{\partial \delta P^y_{m,n}} = \begin{cases} \begin{bmatrix} 0 \\ B_{m-i}(u)\ B_{n-j}(v) \end{bmatrix}, 0 \leq m-i, n-j \leq 3 \\ \\ \mathbf{0}, \qquad otherwise \end{cases}$$

While for the smoothness term one can obtain analytically the following partial derivatives:

$$\frac{\partial}{\partial x}\delta L(\mathbf{\Theta};\mathbf{x}) = \sum_{k=0}^{3}\sum_{l=0}^{3}\big(\frac{\partial}{\partial u}B_k(u)\cdot\frac{du}{dx}\big)B_l(v)\delta P_{i+k,j+l}$$

$$\frac{\partial\ \frac{\partial}{\partial x}(\delta L(\mathbf{\Theta};\mathbf{x}))}{\partial \delta P^x_{m,n}} = \begin{cases} \begin{bmatrix} \big(\frac{\partial}{\partial u}B_{m-i}(u)\cdot\frac{du}{dx}\big)B_{n-j}(v) \\ 0 \end{bmatrix}, & 0\le m-i \\ & n-j\le 3 \\ \mathbf{0}, & otherwise \end{cases}$$

$$\frac{\partial\ \frac{\partial}{\partial x}(\delta L(\mathbf{\Theta};\mathbf{x}))}{\partial \delta P^y_{m,n}} = \begin{cases} \begin{bmatrix} 0 \\ \big(\frac{\partial}{\partial u}B_{m-i}(u)\cdot\frac{du}{dx}\big)B_{n-j}(v) \end{bmatrix}, & 0\le m-i \\ & n-j\le 3 \\ \mathbf{0}, & otherwise \end{cases}$$

The derivation for $\frac{\partial}{\partial y}\delta L(\Theta;\mathbf{x})$, etc. can be similarly obtained.

ACKNOWLEDGEMENTS

The authors are thankful to Dr. Sebastian and Prof. Kimia, from Brown Univ. for providing their shape database [12] as well as to Dr. Mokhtarian and Prof. Kittler for providing the fish shapes. The authors are grateful to S. Zhang, Y. Wang, Prof. Huang and Prof. Samaras from the State University of New York - Stony Brook, for providing the 3D dynamic face range scan data and helping with the quantitative validation. The two 3D face range scans in [Fig. (13)] are provided publicly by Cyberware, Inc.

REFERENCES

[1] R. Veltkamp and M. Hagedoorn, "State-of-the-art in Shape Matching," Utrecht University, Tech. Rep. UU-CS-1999-27, 1999.

[2] T. Cootes, C. Taylor, D. Cooper, and J. Graham, "Active shape models - their training and application," *Computer Vision and Image Understanding*, vol. 61, pp. 38–59, 1995.

[3] N. Paragios, M. Rousson, and V. Ramesh, "Non-Rigid Registration Using Distance Functions," *Computer Vision and Image Understanding*, vol. 89, pp. 142–165, 2003.

[4] Z. Zhang, "Iterative Point Matching for Registration of Free-form Curves and Surfaces," *International Journal of Computer Vision*, vol. 13, no. 2, pp. 119–152, 1994.

[5] 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.

[6] S. Belongie, J. Malik, and J. Puzicha, "Matching Shapes," in *IEEE Int'l Conf. on Computer Vision*, 2001, pp. 456–461.

[7] R. H. Davies, C. J. Twining, T. F. Cootes, J. C. Waterton, and C. J. Taylor, "3D Statistical Shape Models using Direct Optimization of Description Length," in *European Conf. on Computer Vision*, 2002, pp. 3–20.

[8] T. Sebastian, P. Klein, and B. Kimia, "Alignment-based Recognition of Shape Outlines," *Lecture Notes in Computer Science*, vol. LNCS-2059, pp. 606–618, 2001.

[9] H. Chui and A. Rangarajan, "A New Algorithm for Non-Rigid Point Matching," in *IEEE Conf. on Computer Vision and Pattern Recognition*, 2000, pp. II: 44–51.

[10] D. Metaxas, *Physics-Based Deformable Models*. Kluwer Academic Publishers, 1996.

[11] L. H. Staib and J. S. Duncan, "Boundary Finding with Parametrically Deformable Models," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 14, no. 11, pp. 1061–1075, 1992.

[12] T. Sebastian, P. Klein, and B. Kimia, "Recognition of Shapes by Editting Shock Graphs," in *IEEE Int'l Conf. on Computer Vision*, 2001, pp. 755–762.

[13] M. Leventon, E. Grimson, and O. Faugeras, "Statistical Shape Influence in Geodesic Active Contours," in *IEEE Conf. on Computer Vision and Pattern Recognition*, 2000, pp. I:316–322.

[14] C. Chefd'Hotel, G. Hermosillo, and O. Faugeras, "A Variational Approach to Multi-Modal Image Matching," in *IEEE Workshop in Variational and Level Set Methods*, 2001, pp. 21–28.

[15] M. Fornefett, K. Rohr, and H. Stiehl, "Elastic Registration of Medical Images using Radial Basis Functions with Compact Support," in *IEEE Conf. on Computer Vision and Pattern Recognition*, vol. 1, 1999, pp. 1402–1409.

[16] T. Sederberg and S. Parry, "Free-Form Deformation of Solid Geometric Models," in *ACM SIGGRAPH*, 1986, pp. 151–160.

[17] D. Rueckert, L. Sonoda, C. Hayes, D. Hill, M. Leach, and D. Hawkes, "Nonrigid Registration Using Free-Form Deformations: Application to Breast MR Images," *IEEE Transactions on Medical Imaging*, vol. 8, pp. 712–721, 1999.

[18] Y. Gdalyahu and D. Weinshall, "Flexible Syntactic Matching of Curves and its Application to Automatic Hierarchical Classification of Silhouettes," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 21, no. 12, pp. 1312–1328, 1999.

[19] A. E. Johnson and M. Hebert, "Recognizing Objects by Matching Oriented Points," in *IEEE Conf. on Computer Vision and Pattern Recognition*, 1997, pp. 684–689.

[20] S. Osher and J. Sethian, "Fronts Propagating with Curvature-dependent Speed: Algorithms based on the Hamilton-Jacobi Formulation," *J. of Computational Physics*, vol. 79, pp. 12–49, 1988.

[21] A. Collignon, F. Maes, D. Vandermeulen, P. Suetens, and G. Marchal, "Automated Multimodality Image Registration using Information Theory," in *Information Processing in Medical Images*, 1995, pp. 263–274.

[22] P. Viola and W. Wells, “Aligment by Maximization of Mutual Information,” in *IEEE Int’l Conf. on Computer Vision*, 1995, pp. 16–23.

[23] C. Studholme, D. L. G. Hill, and D. J. Hawkes, “An Overlap Invariant Entropy Measure of 3D Medical Image Alignment,” *Pattern Recognition*, vol. 32, no. 1, pp. 71–86, 1999.

[24] W. Zhu and T. Chan, “Stability for Shape Comparison Model,” UCLA-CAM, Tech. Rep. 0308, 2003.

[25] B. Zitova and J. Flusser, “Image Registration Methods: A Survey,” *Image and Vision Computing*, vol. 21, no. 11, pp. 977–1000, 2003.

[26] X. Huang, N. Paragios, and D. Metaxas, “Establishing Local Correspondences towards Compact Representations of Anatomical Structures,” in *Int’l Conf. on Medical Imaging Copmuting and Computer-Assisted Intervention*, 2003, pp. 926–934.

[27] J. P. W. Pluim, J. B. A. Maintz, and M. A. Viergever, “Mutual Information Based Registration of Medical Images: A Survey,” *IEEE Trans. on Medical Imaging*, vol. 22, no. 8, pp. 986–1004, 2003.

[28] J. Feldmar and N. Ayache, “Rigid, Affine and Locally Affine Registration of Free-Form Surfaces,” *Int’l J. of Computer Vision*, vol. 18, pp. 99–119, 1996.

[29] P. Faloutsos, M. van de Panne, and D. Terzopoulos, “Dynamic Free-Form Deformations for Animation Synthesis,” *IEEE Trans. Visualization and Computer Graphics*, vol. 3, pp. 201–214, 1997.

[30] S. Lee, K.-Y. Chwa, J. Hahn, and S. Shin, “Image Morphing Using Deformation Techniques,” *Journal of Visualization and Computer Animation*, vol. 7, pp. 3–23, 1996.

[31] S. Lee, K.-Y. Chwa, and S. Shin, “Image Metamorphosis with Scattered Feature Constraints,” *IEEE Transactions on Visualization and Computer Graphics*, vol. 2, pp. 337–354, 1996.

[32] D. Forsey and R. Bartels, “Hierachical B-spline Refinement,” *ACM Transactions on Computer Graphics*, vol. 22, pp. 205–212, 1988.

[33] T. Hartkens, D. Hill, A. Castellano-Smith, D. Hawkes, C. Maurer, A. Martin, W. Hall, H. Liu, and C. Truwit, “Using Points and Surfaces to Improve Voxel-based Nonrigid Registration,” in *Int’l Conf. on Medical Imaging Copmuting and Computer-Assisted Intervention*, 2002, pp. II:565–572.

[34] P. Laskov and C. Kambhamettu, “Curvature-based Algorithms for Nonrigid Motion and Correspondence Estimation,” *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 25, no. 10, pp. 1349–1354, 2003.

[35] X. Huang, S. Zhang, Y. Wang, D. Metaxas, and D. Samaras, “A Hierarchical Framework for High Resolution Facial Expression Tracking,” in *Third IEEE Workshop on Articulated and Nonrigid Motion, in conjunction with CVPR’04*, July 2004.

[36] P. S. Huang, Q. Hu, F. Jin, and F. P. Chiang, “Color-encoded Digital Fringe Projection Technique for High-speed Three-dimensional Surface Contouring,” *Optical Engineering*, vol. 38, no. 6, pp. 1065–1071, 1999.

[37] L. Zhang, B. Curless, and S. M. Seitz, “Spacetime Stereo: Shape Recovery for Dynamic Scenes,” in *IEEE Conf. on Computer Vision and Pattern Recognition*, 2003, pp. 367–374.

[38] S. Ilic and P. Fua, “Using Dirichlet Free Form Deformation to Fit Deformable Models to Noisy 3-D data,” in *European Conf. on Computer Vision*, vol. LNCS 2351, 2002, pp. 704–717.