

A General Framework for 2D Multiframe and 3D Surface-to-surface Motion Estimation*

Min Li[†], Chandra Kambhampettu[†] and Maureen Stone[‡]

[†] Video/Image Modeling and Synthesis Lab
Department of Computer and Information Sciences
University of Delaware, Newark, DE 19716 USA

[‡] Vocal Tract Visualization Lab
Department of Biomedical Sciences and Orthodontics
University of Maryland Dental School, Baltimore, MD 21201 USA

Abstract

A general framework for 2D multiframe and 3D surface-to-surface motion estimation is presented in this paper. By viewing a 2D contour sequence as a pseudo 3D surface, we solve the motion estimation problem for 2D multiframe and 3D surface-to-surface in a general framework, by estimating the motion of a "surface". The deformation of a "surface" is modeled using spline-based motion. This spline-based motion model does not constrain the motion type in the temporal domain for 2D multiframe motion estimation. For 3D motion estimation, we focus on the relationship between the underlying nonrigid motion and 3D surface properties. The spline motion model provides our method certain advantages over other nonrigid shape-based methods. For example, we do not need approximation of the orthogonal parameterization. The small deformation constraint introduced by the previous surface-to-surface motion estimation methods is also relaxed in our method. Experiments on both synthetic and real motion are presented in this paper.

1 Introduction

Nonrigid motion estimation and recovery of point correspondences are important problems in computer vision. The applications of nonrigid motion analysis include medical imaging, face modeling, tracking, among several others. The nonrigid motion estimation methods can be roughly classified into geometric and physical based methods. The advantage of geometric (shape-based) methods is that motion is solely estimated from the visual data.

1.1 Previous Shape-based Methods

Some shape-based methods assume that the change of normals and curvatures are the minimum during motion. Motion recovery is achieved by minimizing the differences between these geometry properties of the before-motion and after-motion objects. In [19],

*This research was funded in part by NIDCD/NIH grant number R01 DC01758.

the shape invariance is combined with geodesic distance to determine point correspondences between surfaces. In [7] the curvature invariance is used to measure the motion of deformable objects. The shape invariance properties are also used in [8], which is based on the popular ICP [3][20] registration method but has been extended to the nonrigid situation. Some other examples of the shape-based method are [18][16].

These shape-based methods can be applied to both 2D contour and 3D surface situations. However, in 2D case, none of these methods address the multiframe motion estimation problem which requires additional temporal constraint. Multiframe motion estimation is a more general problem in 2D since the object motion is generally represented by a sequence of contours (instead of a pair). Simple extension of these contour-to-contour registration methods to multiframe only takes into account the spatial constraint, however loses the temporal information.

In 3D case, surface-to-surface motion analysis is common since it is difficult to obtain real-time 3D sequences in order to perform 3D multiframe motion analysis. Noticing that the intuition behind shape-based methods referred above (we refer these methods as *direct shape-based* methods) is that the shape properties do not change between surfaces but curvature and normal are only rigid invariants, some researchers have been developing differential geometric methods [10][11][12][13] to capture the shape changes caused by the underlying nonrigid motion. These methods are referred as *nonrigid shape-based* methods in this paper.

1.2 A General Motion Estimation Framework

A 3D surface can be represented as set of 3D points $\mathbf{X}(x, y, z)$, while one point on the t^{th} frame of a sequence of 2D contours is given as $\mathbf{X}(x, y, t)$. If we take the temporal dimension t of a 2D sequence as a special "spatial" dimension, the 2D contour sequence can be viewed as a pseudo 3D surface in the $x - y - t$ space. We solve the motion estimation problem for 2D multiframe and 3D surface-to-surface in a general framework, by *estimating the motion of a "surface"*. The "surface" could be a real 3D surface in the $x - y - z$ space for the surface-to-surface registration or it could be a pseudo 3D surface in the $x - y - t$ space for 2D multiframe registration. Motion of a given 3D $x - y - z$ surface is between this surface and another $x - y - z$ surface while motion of a pseudo $x - y - t$ surface is within the same surface and is between points on consecutive frames.

We model the "surface" motion using a GRBF (Gaussian Radial Basis Function) to solve the 2D multiframe and 3D surface-to-surface motion estimation problem. 2D multiframe motion estimation for heart has been presented in [15] but this method is restricted to periodic motion and a periodic mapping function has to be estimated. By viewing a 2D contour sequence as a pseudo $x - y - t$ surface and modeling the "surface" motion with a GRBF, we do not restrict the motion type of the 2D contour. Additional temporal motion model of 2D multiframe is not required in our method and is incorporated into a single GRBF with the spatial motion.

For 3D surface-to-surface nonrigid motion estimation, we focus on the relationship between the underlying nonrigid motion and 3D shape properties. This is similar to *nonrigid shape-based methods* [10][11][12][13]. In these previous nonrigid shape-based methods, motion models are all defined in different local coordinate systems for different points on the before-motion surface since the basic relationship between underlying motion and shape properties is described in a local coordinate system. These definitions introduced two problems. First, the motion defined in the local coordinate system has no

explicit physical meaning. Second, the motion consistency over the entire motion field as a whole cannot be guaranteed with local motions defined in different local coordinate systems. We still utilize the nonrigid shape relationship defined in the *local* coordinate system but model the surface motion with a *single* GRBF. In addition, we relax small deformation constraint to allow higher order deformations. Details of the advantages of our 3D motion estimation method is presented in Section 4.

2 Nonrigid motion modeling: GRBF

In this paper, we model the nonrigid motion of a "surface" using the GRBF (Gaussian Radial Basis Function). The radial basis function provides both global and local control of deformation. It has been widely used in computer vision and medical imaging, such as image registration [4][9], image warping [2] and surface reconstruction [5]. The basis function of RBF can be thin plate spline (TPS), Multiquadric, linear or Gaussian [9][2][4]. Compared with other RBFs, Gaussian radial basis function (GRBF) is localized in scope and gives a significant response only in a neighborhood around each control point. It is more biologically plausible since its response is bounded.

Given a "surface", the motion estimation problem can be treated as recovering the displacement for each point on this surface. For 2D multiframe, the displacement of each point on the pseudo $x - y - t$ surface starts from this point and ends at a point on the same pseudo surface. The starting and ending points are located in consecutive frames, e.g. the t^{th} and $(t + 1)^{th}$ frames, of the 2D sequences. For 3D surface-to-surface, the displacement of each point on the given before-motion $x - y - z$ surface starts from this point but ends at a point on another $x - y - z$ surface (the after-motion surface).

The displacement field can be described using a GRBF which interpolates the displacements¹ at intermediate points using the displacement values at given control points. If there are n control points, a GRBF in d dimensions, denoted by $\mathbf{S}(\mathbf{X})$ is composed of d functions such that: $\mathbf{S}(\mathbf{X}) = [f_1(\mathbf{X}), \dots, f_k(\mathbf{X}), \dots, f_d(\mathbf{X})]$ where $f_k(\mathbf{X})$ is the displacement function of point \mathbf{X} in the k^{th} dimension. $f_k(\mathbf{X})$ has a global component and a local component: $f_k(\mathbf{X}) = Q_k(\mathbf{X}) + \sum_{i=1}^n \alpha_{ik} g(r_i^2)$.

The global component, $Q_k(\mathbf{X})$ is usually a global affine. i.e. $Q_k(\mathbf{X}) = a_{0k}x + a_{1k}y + a_{2k}z + a_{3k}$ when $d = 3$. The local component, $\sum_{i=1}^n \alpha_{ik} g(r_i^2)$, is the sum of a weighted elastic basis function $g(r_i^2) = e^{-r_i^2/\delta}$ where r_i denotes the distance from the point of interest, \mathbf{X} , to the i^{th} control point. δ is the Gaussian locality parameter. α_{ik} are the weighting parameters.

For convenience, we define a GRBF parameter vector $P = [P_1, \dots, P_k, \dots, P_d]^T$ where P_k is the parameter vector for the k^{th} dimension. In 3D case, the components of P_k are the global affine parameters $a_{0k}, a_{1k}, a_{2k}, a_{3k}$ and radial parameters $\alpha_{ik}, i = 1, \dots, n$.

3 2D Multiframe Motion Estimation

3.1 Motion Model for 2D Multiframe

Assume that the boundary of a moving object has been extracted from 2D images. A sequence of 2D contours is available after the extraction. For understanding and characterizing the object motion, one must first obtain the point correspondence between contours of consecutive frames. To ensure that the frame-to-frame correspondence gives a consistent point-tracked trajectory, point correspondence recovery and motion estimation based

¹In this paper, we use the terms *displacement* and *motion* interchangeably.

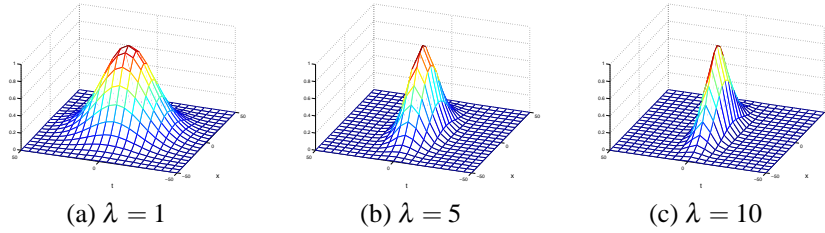


Figure 1: Spatio-temporal Gaussian functions with different λ . Note that when λ changes, the distribution of this Gaussian function varies in the t dimension while it is the same in the x dimension.

on multiframe is a better choice than contour pair. Multiframe estimation combines the spatial and temporal constraints into a single framework. The estimation between contour pair (contour-to-contour) loses the temporal information since it is only performed on consecutive frames.

A potential application of our 2D multiframe motion estimation method is tracking the motion of a contour representing human tongue surface from a sequence of 2D ultrasound images. In an ultrasound image, only part of human tongue surface is presented thus the surface is extracted as an open contour. Suppose a sequence of open contours is available and the open contour at time instance t has the Monge form $y = y_t(x)$. For a point $\mathbf{X}(x, t)$ which has spatial coordinate $(x, y_t(x))$ and located on the t^{th} frame, the displacement $\Delta x(\mathbf{X})$ of the x dimension will uniquely decide the motion of this point. This motion can only move point $\mathbf{X}(x, t)$ to some point $\mathbf{X}'(x', t')$ on the $(t + 1)^{\text{th}}$ frame where $x' = x + \Delta x, t' = t + 1$. $\Delta x(\mathbf{X})$ is modeled as a spatio-temporal GRBF in which the control points are sampled evenly from a $x - y - t$ pseudo surface formed by the 2D contour sequence. That is, control points are picked up from multi-frames of the sequence. With this spatio-temporal modeling, $\Delta x(\mathbf{X})$ is not only influenced by control points on the t^{th} frame, but also interpolated using displacements of control points through multi-frames. Temporal information is incorporated naturally since motion of each frame will affect not only itself but also other frames in the sequence. In another words, $\Delta x(\mathbf{X})$ is modeled as a function of the GRBF parameter vector P : $\Delta x(\mathbf{X}, P)$.

Note that for a point $\mathbf{X}(x, t)$, the spatial and the temporal parameters are of different nature, they must be treated differently. The distance from point $\mathbf{X}(x, t)$ to the i^{th} control point is defined as:

$$r_i = (\Delta s^2 + (\lambda \Delta t)^2)^{1/2} \quad (1)$$

where Δs is spatial distance in the x dimension and Δt is the frame difference between point $\mathbf{X}(x, t)$ and the i^{th} control point. λ is the nonuniform parameter which controls the temporal motion. Larger λ indicates less uniform temporal motion. With the distance r_i defined in Eq. 1, the Gaussian function $g(r_i^2)$ becomes a spatio-temporal kernel function which has nonuniform effect in the x and t dimensions. Examples of the spatio-temporal Gaussian functions are shown in Figure 1.

3.2 2D Multiframe Correspondence Recovery

The displacement $\Delta x(\mathbf{X}, P)$ from point $\mathbf{X}(x, t)$ to the corresponding point $\mathbf{X}'(x + \Delta x(\mathbf{X}, P), t + 1)$ is decided by the following criterion functions: Euclidean point distance E_d , differ-

ences in the normal E_n and curvature E_k , integrated over multiple frames:

$$E_d(P) = |y(x) - y'(x + \Delta x(\mathbf{X}, P))|^2 \quad E_n(P) = |\mathbf{n}(x) - \mathbf{n}'(x + \Delta x(\mathbf{X}, P))|^2 \quad E_k(P) = |k(x) - k'(x + \Delta x(\mathbf{X}, P))|^2. \quad (2)$$

Where \mathbf{n}' and k' denote the normal and curvature of point $\mathbf{X}'(x + \Delta x(\mathbf{X}, P), t + 1)$, respectively.

The GRBF parameter vector P that minimizes the above criterion functions is estimated using the Gauss-Newton optimization technique as [6]. The criterion functions are approximated using a Taylor approximation. For example:

$$E_k = |k(x) - k'(x + \Delta x(\mathbf{X}, P_0)) - \nabla k' \mathbf{J}_P (P - P_0)|^2 \quad (3)$$

where P_0 is the GRBF parameter vector estimated from the previous iteration or the initial value of P in the first iteration. $\nabla k'$ denotes the derivative of the curvature k' on the $(t + 1)^{th}$ frame. \mathbf{J}_P is the Jacobian matrix which denotes the matrix of partial derivatives of Δx with respect to the unknown components of P . For example, if the global affine of GRBF is the 1D affine $a_0x + a_1$ and the radial kernels are $g(r_1^2), \dots, g(r_n^2)$:

$$\mathbf{J}_P = [x, 1, g(r_1^2), \dots, g(r_n^2)]. \quad (4)$$

P can be resolved from set of linear equations derived from Eq. 3 and the similar approximations of E_d and E_n . The linear equation of E_k at one point $\mathbf{X}(x, t)$ is:

$$\nabla k' \mathbf{J}_P (P - P_0) = k(x) - k'(x + \Delta x(\mathbf{X}, P_0)). \quad (5)$$

Similar linear equations for E_d and E_n can be obtained. The set of equations is then solved as a *weighted least-square (WLS)* problem with weighting parameters w_d, w_n, w_c for E_d, E_n and E_c , respectively.

For fast and steady convergence, a coarse-to-fine algorithm is implemented to recover the GRBF parameter vector P . The 2D contours are first transformed to the Fourier domain [17]. High frequencies are filtered to smooth the contour. Curvature and normal are calculated according to the current smooth contours. Recovered GRBF parameters from current level are input to the higher level by increasing the filtering threshold in the Fourier domain. The whole process stops at a fine level.

4 3D Surface Motion Estimation

Similar to 2D multiframe motion estimation, motion between surfaces is modeled using a GRBF for 3D surface-to-surface motion estimation. Different with the direct shape-based method, the shape relationship between two surfaces is described by underlying nonrigid motion in our method. Although similar relationship has been used in some previous nonrigid shape-based methods, our method differs from these methods due to our GRBF motion modeling. Small deformation constraint is also relaxed in our method.

4.1 Background

In this section we present the background for the unit normal changes under nonrigid motion. Let $\mathbf{r} = [u, v, w]^T$ (usually we can write $w = f(u, v)$, thus \mathbf{r} can be represented in the parametric form: $\mathbf{r} = \mathbf{r}(u, v)$) denote a point in the local coordinate system which is defined for each point of interest on the before-motion surface. Let $\mathbf{s}(\mathbf{r})$ denote the

displacement function in the local coordinate system; the corresponding point of \mathbf{r} on the after-motion surface can be defined as $\mathbf{r}' = \mathbf{r} + \mathbf{s}$. The motion estimation problem now becomes the problem of finding the best displacement $\mathbf{s}(\mathbf{r})$ for each point of interest.

The *unit normal* \mathbf{n}' of the after-motion surface is related to the *unit normal* \mathbf{n} of the before-motion surface by the underlying motion \mathbf{s} as [14]:

$$\mathbf{n}' = \mathbf{n} - \frac{\mathbf{n} \times \text{rots}}{1 + \theta} \quad (6)$$

where θ is modulus of dilation, $\theta = \frac{D'-D}{D}$. $D = \sqrt{EG - F^2}$ and E, F, G denote the coefficients of the first fundamental form of a surface. $\text{rots} = \frac{1}{E}\mathbf{r}_u \times \mathbf{s}_u + \frac{1}{G}\mathbf{r}_v \times \mathbf{s}_v$.

The above equation is the basic nonrigid shape relationship we used for 3D surface-to-surface motion estimation. Note that the previous nonrigid shape-based method [10][11][12][13] has an additional approximation $\theta \ll 1$, which leads to the nonrigid shape relationship for small deformation:

$$\mathbf{n}' = \mathbf{n} - \mathbf{n} \times \text{rots}. \quad (7)$$

Note that the crucial requirement of Eqs. 6 and 7 is the orthogonal parameterization. Previous nonrigid shape-based methods approximate the orthogonality at points inside a local patch around each point of interest. In our approach, the local patch around each point of interest is not required for the motion recovery and only the orthogonal parameterization at the point of interest is needed, which can be easily guaranteed by constructing a *principal local coordinate system* at each point of interest [1].

4.2 3D Surface-to-surface Correspondence Recovery

With regularly subsampled control points, the displacement $\mathbf{S}(\mathbf{X})$ of each point \mathbf{X} on the before-motion surface is interpolated by the displacements of the control points with a GRBF.

Note that the GRBF displacement $\mathbf{S}(\mathbf{X})$ is defined over the whole surface, thus it is represented on the world coordinate system. While the relationship between unit normals of the before-motion and after-motion surfaces only hold in the *principal local coordinate system* constructed at each point of interest, an additional step that transforms the displacement \mathbf{S} in the world coordinate system to the displacement \mathbf{s} in the *principal local coordinate* is necessary. For each point of interest, given the unit normal \mathbf{n} as the z axis and two principal directions \mathbf{r}_1 and \mathbf{r}_2 as the other two axes of the *principal coordinate system*, a rotation matrix that transforms (omit the translation) a coordinate point from the world system to the local system can be defined: $R = [\mathbf{r}_1 \mathbf{r}_2 \mathbf{n}]^T$. Thus the displacement \mathbf{S} defined in the world coordinate system can be transformed to the *principal local coordinate system* with this rotation matrix: $\mathbf{s}(\mathbf{r}) = R\mathbf{S}(\mathbf{X})$.

Recall that P is the GRBF parameter vector, we can re-write $\mathbf{S}(\mathbf{X})$ and $\mathbf{s}(\mathbf{r})$ as $\mathbf{S}(\mathbf{X}, P)$ and $\mathbf{s}(\mathbf{r}, P)$. Eq. 6 now can be expressed with P as the unknown parameter:

$$\mathbf{n}'(\mathbf{r} + \mathbf{s}(\mathbf{r}, P)) = \mathbf{n}(\mathbf{r}) - \frac{\mathbf{A}(\mathbf{r}, P)}{1 + \theta}. \quad (8)$$

Where $\mathbf{A} = \mathbf{n} \times \text{rots}$.

We now have the following unit normal criterion to decide the point correspondence between surfaces:

$$E_N(P) = |\mathbf{n}(\mathbf{r}) - \frac{\mathbf{A}(\mathbf{r}, P)}{1 + \theta} - \mathbf{n}'(\mathbf{r} + \mathbf{s}(\mathbf{r}, P))|^2. \quad (9)$$

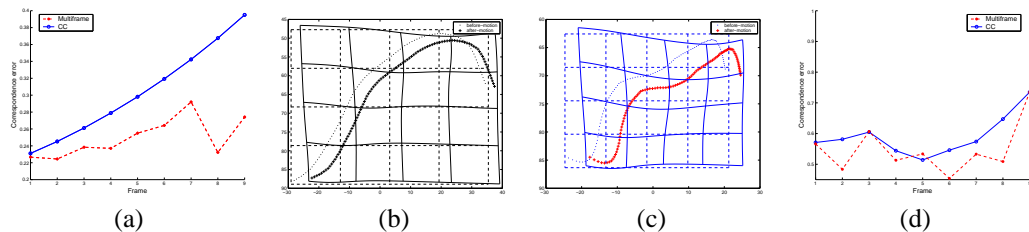


Figure 2: (a) Performance comparison between "Multiframe" and "CC" on the first synthetic sequence. (b) The second synthetic test: first two contours and deformation between them. (c) The second synthetic test: last two contours and deformation between them. (d) The second synthetic test: performance comparison between "Multiframe" and "CC".

Similar to 2D multiframe approach, the motion vector P is recovered using Gauss-Newton technique by combining the above unit normal criterion and the Euclidean distance constraint. Details of the optimization process can be found in [14].

5 Experiments

We performed several tests on 2D sequences and 3D surfaces to evaluate our nonrigid motion estimation method. In 2D tests, we compare our multiframe method ("Multiframe") to the contour-to-contour method ("CC"), which is the general approach used in [7][18] and [16]. In 3D tests, our nonrigid shape-based method is compared to Wang's shape-based method ("Wang's") [19] and the previous nonrigid shape-based method [12] which is restricted to small deformation ("SD").

In order to make quantitative comparisons, We define *correspondence error* $c_{err} = \|p' - p'_c\|$ for both 2D and 3D tests, where p'_c is the recovered corresponding point and p' is the true correspondence point for the point of interest p on the before-motion contour (for 2D) or on the before-motion surface (for 3D).

5.1 2D results

The first 2D test is conducted on a synthetic 2D sequence. A Monge form curve ($-50 \leq x \leq 50$), which is a part of an ellipse (semi-axes $a = 60, b = 40$), is chosen as the first contour of the 10-frame sequence. Consecutive frames are obtained by nonuniformly expanding the previous frame with scaling parameters 1.05 and 1.08 for the x and y dimensions, respectively. These 10 spatio-temporal contours form a pseudo $x - y - t$ surface. We apply our multiframe method to this "surface" and the point correspondence between consecutive contours is recovered. The nine *correspondence errors* by our multiframe method is shown in Figure 2(a) and compared with the results from contour-to-contour estimation. Significant improvement of "Multiframe" over "CC" is observed due to the fact that temporal information is incorporated in the multiframe method.

To test our method with contours in real life and with complex motion, we picked a tongue contour as the starting contour to create a contour sequence. Each subsequent contour in this sequence is obtained by deforming the previous contour with a TPS transformation. The coefficients of the TPS were sampled from a Gaussian distribution with zero mean. The first two contours of such a sequence are shown in Figure 2(b). Deformation between these two contours is also shown in the same figure by comparing the

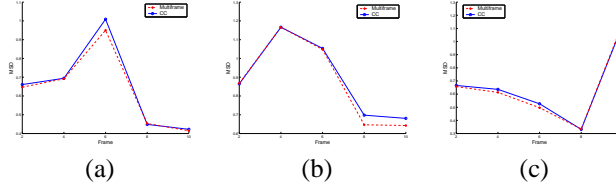


Figure 3: MSD value comparisons on real tongue sequence.

original regular square grid (dashed lines) with its transformed grid (solid lines). There are ten contours in this sequence. The last two contours of the same sequence and the deformation between these two contours are shown in Figure 2(c). Comparison between "Multiframe" and "CC" is presented in Figure 2(d). The average error of "Multiframe" is 0.55 while it is 0.59 in average for "CC". Overall performance of "Multiframe" is still better than "CC". We run the same test hundreds of times by creating the sequence with different TPS coefficients and similar results are observed.

The last 2D test is conducted on real tongue contour sequences extracted from ultrasound images with resolution 640x480. There is no ground-truth correspondence in tongue contour sequence thus c_{err} is not available for evaluation. Noticing the tongue motion is continuous, we interpolate contour C_{i-1} and C_{i+1} according to recovered correspondence to obtain an interpolated contour C'_i . C'_i is compared with the real i^{th} contour C_i . Smaller distance between C'_i and C_i indicates a better correspondence recovery. The distance measure between two contours U and V is defined as the Mean Sum of Distance (MSD): $MSD(U, V) = \frac{1}{2n} (\sum_{i=1}^n \min_j |v_i - u_j| + \sum_{i=1}^n \min_j |u_i - v_j|)$ where n is number of points on the contour. MSD is a measure of pixel errors between contours. Three tongue contour sequences are tested. There are eleven frames for each sequence. MSD value comparisons between "Multiframe" and "CC" at even frames are shown in Figure 3(a), (b) and (c) for three sequences, respectively. The first tongue sequence is obtained from the speech "ea". The second sequence is from the same speech by the same subject but in different session. The third sequence is the speech "golly" by a different subject.

5.2 3D Results

The synthetic experiments are first conducted to evaluate the performance of our algorithm quantitatively. Given the before-motion surface $\mathbf{X} = (x, y, \mathbf{X}(x, y))$ and the after-motion surface $\mathbf{X}' = (x', y', \mathbf{X}'(x', y'))$, we ran our algorithm to recover the correspondence and compare the result with the ground truth. The initial motion is given by a *trivial correspondence* which is defined as $x' = x$, $y' = y$ and $z' = \mathbf{X}'(x', y')$.

We show our results on the surface \mathbf{X} (a Monge patch, $-60 \leq x \leq 60$, $-60 \leq y \leq 60$) of an ellipsoid with semi-axes lengths $a = 120$, $b = 110$, $c = 90$. The after-motion surface \mathbf{X}' is obtained by scaling \mathbf{X} with different parameters δ_x, δ_y , and δ_z for the x , y and z dimensions, respectively. The synthetic experiments are conducted for several situations. We present a typical such experiment where $\delta_x = 1.15$, $\delta_y = 1.18$ are fixed while δ_z is varied from 1.1 to 1.2 with a step 0.01. The correspondence was recovered for these eleven situations.

The results of our algorithm are also compared against Wang's shape-based method [19]. We also recovered the correspondence with Eq. 7 which is used in the SD method

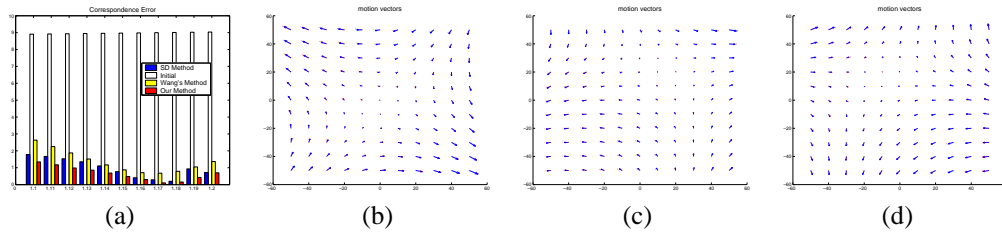


Figure 4: (a) Correspondence errors of the first synthetic 3D test. (b) (c) and (d) are three examples of the synthetic 3D test with complex motion. (b) Initialization: $c_{err} = 3.16$. SD: $c_{err} = 0.84$. Our method: $c_{err} = 0.60$. (c) Initialization: $c_{err} = 2.37$. SD: $c_{err} = 0.98$. Our method: $c_{err} = 0.87$. (d) Initialization: $c_{err} = 2.84$. SD: $c_{err} = 0.74$. Our method: $c_{err} = 0.67$.

[12]. The initial *correspondence errors*, recovered *correspondence errors* of our method, Wang’s method and SD method are shown in Figure 4(a). One can see from this figure that our method outperforms both Wang’s and SD method.

Similar to the 2D test, we tested our surface-to-surface motion estimation method with complex motion. The after-motion surface \mathbf{X}' is deformed from \mathbf{X} with a TPS transformation. Three examples are shown in Figure 4(b), (c) and (d). In these three figures, the recovered 3D displacement (dash) by our method is projected to the xy plane and compared with ground truth displacement (solid). Wang’s method fails for these complex motion and our method still outperforms the SD method. See Figure 4 for details.

Our method is also evaluated with real motion. The real motion of an object is recorded by stereo imaging at two time instances t_1 and t_2 . Object motion between t_1 and t_2 is reconstructed from the motion of some feature points. The object motion is then mapped from the object space to the space of the surface \mathbf{X} ; after-motion surface \mathbf{X}' is deformed from \mathbf{X} according to the mapped real motion. The details of the motion mapping can be found in [14].

The first evaluation of real motion is conducted with the paper bending motion. We also evaluated our algorithm with two types of real face motions. First is the motion from a neutral to a smile face and the second is from the neutral face to an open-mouth face. Feature points are marked in these two face tests.

In total, five deformations are tested for paper bending. The initial *trivial correspondence errors*, recovered *correspondence errors* by our method and SD method are shown in Figure 5(a). Results for these five small-to-large paper bending deformations are shown from left to right. In all situations our method is better than the SD method. As the deformation increases, correspondence error increases for both methods but the error of SD method increases faster than our method as expected. This shows that the small deformation requirement of Eq.7 is relaxed in Eq.6, which is used in our method. Correspondence errors of five small-to-large smile and five small-to-large open-mouth deformations are also shown from left to right in Figures 5(b) and (c), respectively. Similar results as the paper bending experiment are observed in smile and open-mouth experiments.

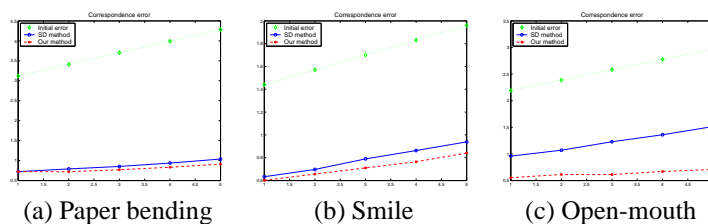


Figure 5: Correspondence errors.

6 Conclusion

A general framework for 2D multiframe and 3D surface-to-surface motion estimation is presented in this paper. Experiments show that our 2D multiframe method is better than the contour-to-contour motion estimation method and our surface-to-surface method outperforms previous nonrigid shape-based methods. Future works include developing a radial basis function with compact support for the 2D multiframe method in order to supply stronger control in the temporal domain. For the 3D surface-to-surface motion estimation method, combining the nonrigid Gaussian curvature relationship with a single spline-based motion is our ongoing research.

References

- [1] A.A. Amini and J.S. Duncan. Bending and stretching models for lv wall motion analysis from curves and surfaces. *IVC*, 10(6):418–430, 1992.
- [2] N. Arad and D. Reissfeld. Image warping using few anchor points and radial functions. *Comp. Graphics Forum*, 14(1):35–46, March 1995.
- [3] P.J. Besl and N.D. McKay. A method for registration of 3-d shapes. *PAMI*, 14(2):239–256, February 1992.
- [4] F.L. Bookstein. Principal warps: Thin-plate splines and the decomposition of deformations. *PAMI*, 11(6):567–585, June 1989.
- [5] J.C. Carr, W.R. Fright, and R.K. Beatson. Surface interpolation with radial basis functions for medical imaging. *MedImg*, 16(1):96–107, February 1997.
- [6] Y. Caspi and M. Irani. Spatio-temporal alignment of sequences. *PAMI*, 24(11):1409–1424, November 2002.
- [7] J.S. Duncan, R.L. Owen, L.H. Staib, and P. Anandan. Measurement of non-rigid motion using contour shape descriptors. In *CVPR91*, pages 318–324, 1991.
- [8] J. Feldmar and N.J. Ayache. Rigid, affine and locally affine registration of free-form surfaces. *IJCV*, 18(2):99–119, May 1996.
- [9] M. Fornefett, K. Rohr, and H.S. Stiehl. Elastic registration of medical images using radial basis functions with compact support. In *CVPR99*, pages I: 402–407, 1999.
- [10] C. Kambhamettu and D.B. Goldgof. Curvature-based approach to point correspondence recovery in conformal nonrigid motion. *CVGIP*, 60(1):26–43, July 1994.
- [11] C. Kambhamettu, D.B. Goldgof, and M. He. Determination of motion parameters and estimation of point correspondences in small nonrigid deformations. In *CVPR94*, pages 943–946, 1994.
- [12] C. Kambhamettu, D.B. Goldgof, M. He, and P. Laskov. 3d nonrigid motion analysis under small deformations. *IVC*, 21(3):229–245, March 2003.
- [13] P. Laskov and C. Kambhamettu. Curvature-based algorithms for nonrigid motion and correspondence estimation. *PAMI*, 25(10):1349–1354, October 2003.
- [14] M. Li, C. Kambhamettu, and M. Stone. Spline-based motion recovery for 3d surfaces using nonrigid shape properties. In *IEEE workshop on Articulated and NonRigid Motion 2004. In conjunction with CVPR'04. Washington, DC*, 2004.
- [15] J.C. McEachen, II, A. Nehorai, and J.S. Duncan. Multiframe temporal estimation of cardiac nonrigid motion. *IP*, 9(4):651–665, April 2000.
- [16] D. Meier and E. Fisher. Parameter space warping: Shape-based correspondence between morphologically different objects. *MedImg*, 21(1):31–47, January 2002.
- [17] L.H. Staib and J.S. Duncan. Boundary finding with parametrically deformable models. *PAMI*, 14(11):1061–1075, November 1992.
- [18] H.D. Tagare. Shape-based nonrigid correspondence with application to heart motion analysis. *MedImg*, 18(7):570–579, July 1999.
- [19] Y. Wang, B.S. Peterson, and L.H. Staib. Shape-based 3d surface correspondence using geodesics and local geometry. In *CVPR00*, pages II:644–651, 2000.
- [20] Z.Y. Zhang. Iterative point matching for registration of free-form curves and surfaces. *IJCV*, 13(2):119–152, October 1994.