10. TECHNIQUES FOR THE ACCURATE RECOVERY OF TIME-VARYING 3D SHAPES IN MEDICAL IMAGING

YOSHINOBU SATO, MASAMITSU MORIYAMA, TAKASHI UEUGUCHI, MASAYUKI HANAYAMA, HIROAKI NAITO AND SHINICH TAMURA

1 INTRODUCTION

Acquisition of 3D shape models is one of the most important topics in the computer vision field because of increasing demands for graphic displays in a virtual space as well as quantitative shape analysis in medical diagnosis and industrial inspection. For the purpose of acquiring 3D shape models directly from images, the use of occluding contours has received considerable attention. There have been two main approaches to 3D model acquisition using occluding contours: one approach integrates apparent contours from continuously varying viewpoints [1–3], and the other approach uses deformable models in which model constraints are incorporated such as symmetries and some other regularities [4–6]. The former has the advantage that accurate shape recovery is possible. Although several results have been reported for acquiring 3D models of rigid objects, it seems difficult to recover 3D models from fragmented contours. On the other hand, the latter has the advantage that 3D models can be recovered even using contours from one viewpoint or fragmented contours. Several recovery results including non-rigid moving objects have been demonstrated from one viewpoint [4,5]. More recently, rigid object recovery from multiple viewpoints was reported [6]. However, recovered models were imposed to have rotational or mirror symmetries, and then they were not regarded as accurate but only as plausible.

In this paper, we propose a method for acquiring accurate 3D shape models of non-rigid moving objects directly from images [7]. Especially, we aim at the recovery
of the left ventricular (LV) shapes, one of the most important types of non-rigid objects [8]. We use X-ray cineangiograms, that is, the 2D projections of LV, as an image data source. If we use ultra-fast CT or MRI synchronized with an electrocardiogram, we can directly obtain 3D cross-sectional information to recover 3D models [9,10]. However, X-ray cineangiography still has the advantages on temporal resolution, and spatial resolution along an axial direction of tomography as compared with ultra-fast CT and gated MRI. Furthermore, LV imaging by biplane or single-plane cineangiography is a procedure commonly performed in cardiac catheterization, which is a routine examination regarded as the most reliable and accurate method for cardiac diagnosis by physicians. From the clinical aspect, there is a need for a more accurate LV recovery method by improving conventionally used cineangiography without introducing any special examination such as ultra-fast CT and gated MRI. One of such efforts is to use the density profiles [11] as well as the apparent contours of LV. The problem of this approach is the difficulty of keeping the uniform density of contrast media.

Our recovery method is based on the integration of apparent contours from various viewpoints in order to acquire not only plausible but also accurate 3D models for quantitative shape analysis in cardiac diagnosis. We perform direct fitting to a closed surface model similar to a deformable model in order to deal with fragmented contours such as extracted from X-ray cineangiograms. While LV images are taken from one or two fixed viewing directions in conventional cineangiography, we vary viewing directions continuously when LV images are taken, which can be easily realized using conventional devices. There is previous work on such an image acquisition method motivated from purely clinical concerns [12]. However, the previous recovery method was very primitive, which reconstructed cross section shapes one by one and placed side by side along axial and temporal directions. Therefore, its formulation was quite insufficient as concerns the use of spatiotemporal smoothness of cardiac motion and shape and the evaluation of matches between the extracted contours and the projections of recovered shapes. Also, it was difficult to deal with fragmented contours. In this paper, in order to overcome those problems, we use a time-varying closed surface represented using B-spline functions having three variables (two as surface and one as time) to fit directly and simultaneously to all contour data extracted from time and viewpoint varying images.

The organization of the paper is as follows: In Section 2, we describe the assumption on non-rigid motion, and clarify the advantages of time and viewpoint varying images for non-rigid object recovery. In Section 3, we describe the representation of a time-varying closed surface and clarify the constraints relating the model surface to contour data extracted from images. An iterative recovery method is formulated in order to fit the time-varying model surface to contour data. In Section 4, we present experimental results using synthesized and real image sequences using balloon phantoms. In Section 5, we give discussion on the representation issue of time-varying surface, and the problems toward clinical application.
2 Obtaining Time and Viewpoint Varying Images

In general, non-rigid motion includes translation and rotation as well as the time-variations of object shape. In this paper, however, we assume that object motion originates from only shape variations, but does not include translational and rotational components. When we observe such non-rigid moving 3D objects in order to obtain their time-varying shape information, it is natural that we should vary viewpoints while observing them. We try to give a consideration to this intuition in this section.

A viewpoint can be represented as a point on a spherical surface, which can be parameterized using latitude $\phi$ and longitude $\theta$. In the case of LV recovery using a biplane X-ray system, we plan to vary LAO and RAO (left- and right-anterior-oblique view) angles from 0° to 90° and from 90° to 0°, respectively. When a patient body is aligned to the polar direction of the spherical coordinate system, the variations of LAO and RAO angles correspond to the variation of longitude $\theta$. Thus, we consider the variation of only longitude $\theta$ here. We consider a time–viewpoint space whose axes are time $t$ and viewpoint $\theta$. We assume that an object is observed by two cameras (a biplane X-ray system) whose viewing directions are orthogonal and given by $(\theta_1(t), \theta_2(t))$, where $|\theta_1(t) - \theta_2(t)| = \pi/2$. The sample points for image acquisition are taken along $\theta_1(t)$ and $\theta_2(t)$ in $t-\theta$ space. If fixed viewpoints are assumed, sample points in $t-\theta$ space are represented as shown in Figure 10.1(a). If time-varying viewpoints are assumed, more uniform sampling can be realized in $t-\theta$ space as shown in Figure 10.1(b). If we can assume both shape and motion are smooth, the recovery of more accurate 3D shapes can be expected using the combination of uniform sampling in $t-\theta$ space and an appropriate spatiotemporal interpolation method as compared with the combination of images obtained by dense sampling along either $t$ or $\theta$ and strong constraints on object shape or motion, except for the case of rigid objects (i.e. no motion) or rotationally symmetric shapes.

In general, cardiac motion can be approximated as periodic motion. When object motion can be assumed to be periodic, time-varying viewpoints are more advantageous. If observable time is long enough compared with one cycle, we can obtain dense uniform image sampling in $t-\theta$ space. In LV imaging by cineangiography, observable time is two or three seconds (that is, from three to five cardiac cycles) during one injection of contrast medium. The sampling pattern shown in Figure 10.1(c) is realized using viewpoint variations given by $(\theta_1(t), \theta_2(t)) = (\pi t/8T_0, \pi t/8T_0 + \pi/2)$ (where $T_0$ is one cycle of periodic motion) and image acquisition by sampling interval $T_0/9$. Different sampling patterns can be realized by changing viewpoint variations $(\theta_1(t), \theta_2(t))$ and the sampling interval (Figure 10.1(d)).

3 Recovery of Time-Varying 3D Shape Models

3.1 Representation of time-varying 3D shape models

We represent time-varying 3D shapes using uniform B-spline functions. In order to parameterize a closed surface, we use spherical coordinates. We specify 3D position using latitude $u$ and longitude $v$, and distance $r$ from the origin to the direction
Figure 10.1. Sampling patterns in time–viewpoint space \((t-\theta)\) space assuming the use of two cameras whose viewing directions are orthogonal. (a) Sampling pattern for time-varying but viewpoint fixed images. Viewpoints are given by \((\theta_1(t), \theta_2(t)) = (\pi/4, 3\pi/4)\). (b) Sampling pattern for time- and viewpoint-varying images of shapes with non-periodic motion. The velocity of viewpoint variation should be fast enough in order to obtain relatively uniform sampling. (c) Sampling pattern for periodic motion with cycle \(T_0\). The velocity of viewpoint variation can be slow enough in order to obtain dense uniform sampling. This pattern is realized by viewpoint variations \((\theta_1(t), \theta_2(t)) = (\pi t/8T_0, \pi t/8T_0 + \pi/2)\), and sampling interval \(T_0/9\). (d) Sampling pattern for periodic motion with cycle \(T_0\) realized by viewpoint variations \((\theta_1(t), \theta_2(t)) = (\pi t/4T_0, \pi t/4T_0 + \pi/2)\), and sampling interval \(2T_0/9\).

specified by latitude \(u\) and longitude \(v\). This means that recovered closed surfaces are limited to star-shaped surfaces with respect to the origin of the spherical coordinate system. Nevertheless, we believe that this class of surface is useful in many domains, especially in LV shape representation. Also, we assume that motion is periodic and its cycle is \(T_0\). Therefore, time-varying 3D shapes are represented by

\[
    r(u, v, t) = \sum_{i=-3}^{i_0-1} \sum_{j=0}^{j_0-1} \sum_{k=0}^{k_0-1} R_{ijk} U_i(u) V_j(v) T_k(t),
\]

(1)

where \(u \in [0, \pi]\), \(v \in [0, 2\pi]\), \(t \in [0, T_0]\), \(R_{ijk}\) is a coefficient, \(U_i(u)\) is the basis function of uniform cubic B-spline for non-periodic functions, and \(V_j(v)\) and \(T_k(t)\) are the basis functions for periodic functions.
3.2 Constraints for 3D recovery from contours

Given an image with known viewpoint \((\phi, \theta)\) and time \(t\), we want to derive the constraints which relate 2D coordinates of contour points in an image to 3D position and normal on time-varying surface \(r(u, v, t)\) (see Figure 10.2). For simplicity, we assume that the spherical coordinate system for representing a viewpoint is coincident with the one for representing a closed surface without loss of generality. Also, we assume orthography as an image projection model. A viewing direction can be given by

$$
v = (\cos \phi \cos \theta, \cos \phi \sin \theta, \sin \phi).
$$

We define two orthogonal directions of image axes as

$$
i = (-\sin \theta, \cos \theta, 0),
$$

and,

$$
\mathbf{j} = (-\sin \phi \cos \theta, -\sin \phi \sin \theta, \cos \phi).
$$

The optical ray corresponding to image coordinates \(\mathbf{I} = (x, y)\) is given by

$$
\mathbf{P} + \lambda \mathbf{v},
$$

where \(\mathbf{P} = xi + yj\), and \(\lambda\) is a scalar value. Here, we suppose that the optical ray passing through the origin of the spherical coordinate system is defined as \(\lambda \mathbf{v}\).

Now, we derive the constraints on a surface represented by \(r(u, v, t)\), given an image contour point \(\mathbf{I} = (x, y)\) at viewpoint \(\mathbf{v}\) and time \(t\). 3D position \(\mathbf{X}(u, v, t)\) on
its time-varying surface is obtained by the transformation from spherical coordinates to Cartesian coordinates. Because \( r(u, v, t) \) is the distance from the origin along 3D direction \((\cos v \cos u, \cos v \sin u, \sin v)\) at time \(t\), 3D position \(X(u, v, t)\) is given by

\[
X(u, v, t) = r(u, v, t)(\cos v \cos u, \cos v \sin u, \sin v) .
\]  

(6)

Surface normal \(n(u, v, t)\) at \(X(u, v, t)\) is given by

\[
n(u, v, t) = N \left[ \frac{\partial X(u, v, t)}{\partial u} \times \frac{\partial X(u, v, t)}{\partial v} \right],
\]

(7)

where \(N[x] = x/|x|\). If there is an image contour point \(I = (x, y)\) with viewpoint \(v\) and time \(t\) which is a projection of an occluding contour of a surface, the constraints given by

\[
X(u, v, t) = P + \lambda v,
\]

(8)

and

\[
n(u, v, t) \cdot v = 0
\]

(9)

must be satisfied at the corresponding surface coordinates \((u, v)\), where \(\lambda = X(u, v, t) \cdot v\), and \(P = xi + yj\).

3.3 Iterative method for time-varying 3D recovery

Based on the constraints given by Eqs (8) and (9), we formulate a method for estimating \(r(u, v, t)\). Equations (8) and (9) are the basic constraints for 3D recovery from occluding contours, which are also described in [3]. In our problems, however, it is difficult to directly obtain \(r(u, v, t)\) satisfying these constraints because it is unknown what coordinates \((u_\ell, v_\ell)\) correspond to the optical ray \(P_\ell + \lambda v_\ell\) determined by given image coordinates \((x_\ell, y_\ell)\).

In order to obtain an approximate solution, we decompose the problem into two stages: First, we use Eq. (9) to find the correspondence between surface coordinates \((u_\ell, v_\ell)\) and each optical ray determined by image coordinates \((x_\ell, y_\ell)\) of a given contour point. Second, we estimate \(r(u, v, t)\) by solving a linear equation system obtained from Eq. (8). We iterate these two stages to finally obtain a solution satisfying both Eqs (8) and (9).

The recovery algorithm is described as follows. We start the algorithm by setting initial shape \(r^{(0)}(u, v, t)\), and computing \(X^{(0)}(u, v, t)\) and \(n^{(0)}(u, v, t)\). (In the experiments, we used a sphere as initial shape \(r^{(0)}(u, v, t)\).) Let \(m\) be an iteration count. We set \(m = 0\) initially.

During the first stage, we find tentative correspondence between surface coordinates \((u_\ell, v_\ell, t_\ell)\) and an optical ray determined by image coordinates \((x_\ell, y_\ell)\) at time \(t_\ell\) (see Figure 10.3). We find this correspondence for every image contour point at every viewpoint and time. Given \(X^{(m)}(u, v, t)\), \(n^{(m)}(u, v, t)\), and the optical ray \(P_\ell + \lambda v_\ell\)
which corresponds to image contour point \((x_\ell, y_\ell)\), we find \(u_\ell, v_\ell,\) and \(\lambda_\ell\) satisfying the constraints

\[
\mathbf{n}^{(m)}(u_\ell, v_\ell, t_\ell) \cdot \mathbf{v}_\ell = 0,
\]

\[\text{(10)}\]

and

\[
\mathbf{P}_\ell + \lambda_\ell \mathbf{v}_\ell = \alpha_\ell (\cos v_\ell \cos u_\ell, \cos v_\ell \sin u_\ell, \sin v_\ell),
\]

\[\text{(11)}\]

where \(\alpha_\ell\) is a scalar coefficient. If we suppose that \(\lambda_\ell\) is given, the 3D position of \(\mathbf{P}_\ell + \lambda_\ell \mathbf{v}_\ell\) is determined. By representing this 3D position using spherical coordinates, \(u_\ell, v_\ell,\) and \(\alpha_\ell\) are uniquely determined using Eq. (11). We can check whether Eq. (10) is satisfied for determined \(u_\ell\) and \(v_\ell\). In order to find \(u_\ell, v_\ell,\) and \(\alpha_\ell\) satisfying the constraints, we continuously vary \(\lambda_\ell\) and search the value of \(\lambda_\ell\) which satisfies Eq. (10). If there are multiple values of \(\lambda_\ell\) satisfying Eq. (10), we select \(\lambda_\ell\) where \(|\alpha_\ell - r^{(m)}(u_\ell, v_\ell, t_\ell)|\) is the minimum.

During the second stage, based on \(u_\ell, v_\ell,\) and \(\alpha_\ell\) found at the first stage, we estimate \(r^{(m+1)}(u, v, t)\) by solving a set of linear equations derived from

\[
r^{(m+1)}(u_\ell, v_\ell, t_\ell) = \alpha_\ell.
\]

\[\text{(12)}\]

More precisely, combining with the smoothness constraint, we find \(R_{ijk}^{(m+1)}\) minimizing

\[
E^{(m+1)} = \frac{1}{\ell_0} \sum_{\ell=1}^{\ell_0} \left[ r^{(m+1)}(u_\ell, v_\ell, t_\ell) - \alpha_\ell \right]^2 + w_\ell \cdot \frac{1}{2T_0 \pi^2} \int \int \int \left( \pi \cdot \frac{\partial r^{(m+1)}(u, v, t)}{\partial u} \right)^2 \mathrm{d}u \mathrm{d}v \mathrm{d}t
\]

\[+ \left( 2\pi \cdot \frac{1}{\cos u} \cdot \frac{\partial r^{(m+1)}(u, v, t)}{\partial v} \right)^2 + \left( T_0 \cdot \frac{\partial r^{(m+1)}(u, v, t)}{\partial t} \right)^2 \mathrm{d}u \mathrm{d}v \mathrm{d}t \]

\[\text{(13)}\]
where

\[ r^{(m+1)}(u, v, t) = \sum_{i=-3}^{i_0-1} \sum_{j=0}^{j_0-1} \sum_{k=0}^{k_0-1} R^{(m+1)}_{ijk}(u) V_j(v) T_k(t), \]

(14)

and \( w \) is a weight parameter for the smoothness constraint. \( 1/\ell_0 \) and \( 1/(2T_0 \pi^2) \) are factors for obtaining average values from the summation and the integral, which normalizes the smoothness constraint and the data constraint based on Eq. (12). \( \pi, 2\pi \), and \( T_0 \) (by which the partial derivatives are multiplied) are factors for the normalization of each partial derivative. The partial derivative with respect to \( v \) is multiplied by \( 1/\cos u \) because of the reduction of length along \( v \) with approaching the poles. (In the experiments, the normalized partial derivatives were estimated using discrete approximations such as \( (R_{ijk} - R_{i+1,j,k})/(1/i_0) \)). If error \( E^{(m+1)} \) for newly estimated \( r^{(m+1)}(u, v, t) \) is almost the same as previous error \( E^{(m)} \), then we stop the algorithm, else we set \( m = m + 1 \) and go back to the first stage. (Empirically, four iterations were sufficient for an appropriate weight value of the smoothness constraint.)

4 EXPERIMENTAL RESULTS

We evaluated the method using synthesized and real image sequences. The method was implemented on a SPARC Station 20.

The synthesized image sequences were generated based on the stationary 3D shape recovered from viewpoint-varying X-ray images of a stationary balloon filled with contrast media. The recovered 3D shape was deformed using three different time-varying periodic scale functions along three orthogonal directions so that the time variation of its volume was similar to the one of LV. The periodic scale functions were described by sinusoidal functions. In a normal LV, its shape is roughly rotational symmetric, and its contraction is relatively uniform everywhere on heart wall. In a diseased LV, however, it is often that its shape is not rotational symmetric and/or the contraction is not uniform, which causes non-symmetric shapes. We generated time-varying shape so as not to be rotationally symmetric. The viewpoint-varying image sequences were generated under the assumption of orthographic projection.

The real image sequences were obtained by taking X-ray images of a balloon filled with contrast media using a biplane X-ray system (Siemens BICOR). Although the shape of a balloon is commonly rotationally symmetric, we deformed the balloon shape by covering carton frames. We controlled the volume of the balloon using a pump so that the time variation of the balloon volume was periodic and similar to the one of LV. Also, we took CT images of the balloon at several time phases and used the 3D models reconstructed from the CT images as the golden standard. In this case, perspective projection is more appropriate as an image projection model than orthography. However, we applied the recovery method assuming orthography.

Before the recovery of time-varying surfaces, the detection of image edges is necessary. We took the zero-crossings of the \( \nabla^2 G \) whose gradient magnitude values are large as image edges. Our method currently cannot discriminate “spurious” edges which should be regarded as outliers. So, we manually removed “spurious” edges
which did not originate from the occluding contours of the balloon. We randomly selected 25\% of all the extracted edges in each image and used them for the 3D recovery.

In the following experiments, we used a uniform function, that is a sphere, as initial shape $\mathbf{r}^{(0)}(u, v, t)$. The spherical coordinate system for surface representation was selected using the following method. The axis of elongation in the projected shape was manually specified in two images taken from orthogonal viewpoints at the systolic phase. A 3D line segment was determined as $z$-axis of the coordinate system so that its projections were coincident with the specified two axes in these images. The origin was set at the center of the 3D line segment. The directions of $x$-axis and $y$-axis were set to the two orthogonal directions from which the two images had been taken.

4.1 Synthesized image sequence

The image sequences were synthesized assuming the use of a biplane X-ray system by which projections from two orthogonal views can be obtained simultaneously. Three image sequences were generated using sampling patterns in the time–viewpoint space as shown in Figure 10.1. One of these sampling patterns was the time and viewpoint varying sequence generated using the sampling pattern shown in Figure 10.1(c). The other two sequences were generated using fixed viewpoints $(0, \pi/2)$ and $(\pi/4, 3\pi/4)$ with time interval $T_0/36$.

Figure 10.4 shows a part of the viewpoint-varying images. The size of each image was $220 \times 220$ (pixels). Figure 10.5 shows the shaded displays of time-varying 3D shapes recovered from the viewpoint-varying sequence. We used $i_0 = 12$, $j_0 = 12$, and $k_0 = 12$ as the number of knots of B-spline functions in Eq. (14), that is, the
Figure 10.5. Shaded displays of time-varying 3D shapes of one cycle recovered from time and viewpoint varying synthesized image sequence.

Figure 10.6. Time variations of the normalized volume of differences of the recovered shapes from viewpoint-varying sequence and viewpoint-fixed sequences.

grid was $12 \times 12 \times 12$. We used $w_s = 0.5 \times 10^{-3}$ as the weight parameter of the smoothness constraint in Eq. (13). Four iterations were needed for the recovery algorithm to almost converge. Figure 10.6 shows the time variations of error in the recovered shapes. We used the volume of the differences between the true shape and the recovered shape as a measure of error. We further divided the volume of differences by the true volume at each time phase to normalize the error. It should be noticed that the volume of differences is not the difference between the true volume and the estimated volume. That is, we did not use the difference of volumes, but used the difference of shapes. In Figure 10.6, the normalized volume of differences in the results recovered from the viewpoint-fixed images highly depended on the selection of two viewpoints. We synthesized the time-varying shapes by deforming a roughly rotational symmetric shape using time-varying scale functions along three orthogonal directions. Because two orthogonal viewpoints happened to be close to two of these three orthogonal directions when $(\theta_1(t), \theta_2(t)) = (0, \pi/2)$, the recovered shapes were relatively accurate (Figure 10.7). Nevertheless, the accuracy of the result recovered from the viewpoint-varying images was considerably higher through one cycle. Figure 10.8 shows the projection images of the recovered 3D shapes superimposed on the projection images of the true shapes. The projection images of the shape recovered from the viewpoint-varying images and the true shape were almost the same at not only
Computational Methods in Biophysics, Biomaterials, Biotechnology and Medical Systems
Algorithm Development, Mathematical Analysis and Diagnostics
Volume I: Algorithm Techniques
Volume II: Computational Methods
Volume III: Mathematical Analysis Methods
Volume IV: Diagnostic Methods
Leondes, C.T. (Ed.)
2002, 1461 p., Hardcover