TRACKING TONGUE MOTION IN THREE DIMENSIONS USING TAGGED MR IMAGES

Xiaofeng Liu[1], *Maureen Stone*[3], *and Jerry L.Prince*[2]

[1]Computer Science, The Johns Hopkins University, Baltimore, MD, 21218
[2]Electrical and Computer Engineering, The Johns Hopkins University, Baltimore, MD, 21218
[3]Biomedical Sciences and Orthodontics , University of Maryland Dental School, Baltimore, MD, 21201

ABSTRACT

Harmonic phase (HARP) analysis has been used in tagged magnetic resonance imaging (MRI) to measure two-dimensional (2D) in-plane motion and strain, and was recently applied in characterizing the motion of tongue during speech. The 3D-HARP method extended the HARP method to track three-dimensional (3D) cardiac motion from short- and long-axis tagged MR images by diffusing 2D in-plane motion on a sparse 3D mesh. In this paper, we propose a new 3D-HARP method to calculate the 3D tongue motion from three orthogonal tag orientations imaged within sagittal, coronal, and axial image planes. Our method iteratively tracks 2D in-plane motions on points on a compact mesh using 2D-HARP followed by thin-plate spline (TPS) interpolation to extend the 2D motion to the whole mesh. Experiments on real tongue data show that our method is capable of accurate motion tracking in simple utterances.

1. INTRODUCTION

The harmonic phase (HARP) technique was developed for automatic motion tracking and strain analysis of tagged magnetic resonance (MR) imaging [1]. It was originally applied to cardiac motion tracking and was also been applied to the tracking of tongue motion during speech and swallowing [2]. The original HARP method was limited to tracking 2D tissue motion in the image plane. Recently, several research groups have extended HARP to 3D [3, 4, 5]. The basic difficulty faced in making this extension is the lack of dense motion information in three dimensions. This can be solved through extensive data acquisition [3] or by constructing models of the object to be tracked [4, 5].

Our approach to 3D motion tracking extends that of Pan et al. [5], which uses a mesh model of the object undergoing deformation. In their case, the mesh was cup-shaped in order to resemble the shape of the left ventricle of the heart. In contrast, our approach uses small planar rectangular meshes that can be placed anywhere in the tongue. In order to track motion, the approach of Pan et al. formed a deformation field on their mesh by extrapolating sparse and partial displacement information measured at intersections of the mesh with the (tagged) image planes. The extrapolation was formed using a fixed Gaussian kernel whose size was chosen empirically based on the mesh size and separation of the observed images. In this paper, we use a thin plate spline in order to interpolate (and extrapolate) the sparse displacement information. This methodology is more appropriate to the geometry of our meshes and less arbitrary in its selection of parameters. Results on 3D tongue images acquired during speech show that 3D motion can be tracked smoothly and consistently using our new method.

This research was supported by NIH grants R01 HL047405 and R01 DC001758

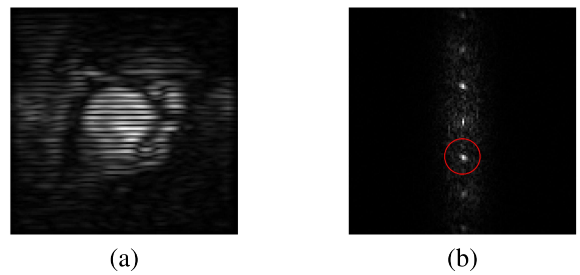

Fig. 1. (a) A tagged MR image of a sagittal cross section of the tongue and (b) the magnitude of its Fourier transform.

2. BACKGROUND

2.1. 2D HARP Method

Fig. 1(a) shows a sinusoidally tagged MR image of the tongue, and Fig. 1(b) shows the magnitude of its Fourier transform. In the HARP method, a bandpass filter placed on one of the harmonic peaks [as shown in Fig. 1(b)] is applied on the spectrum, and its inverse Fourier transform is calculated. The resulting harmonic image is then given by [1]

$$I(\mathbf{x}, t) = D(\mathbf{x}, t)e^{j\phi(\mathbf{x},t)} , \quad (1)$$

where $D(\mathbf{x}, t)$ is the harmonic magnitude image, and $\phi(\mathbf{x}, t)$ is the harmonic phase (HARP) image.

The HARP value is a material property of the tagged tissue, which means the harmonic phase of a material point does

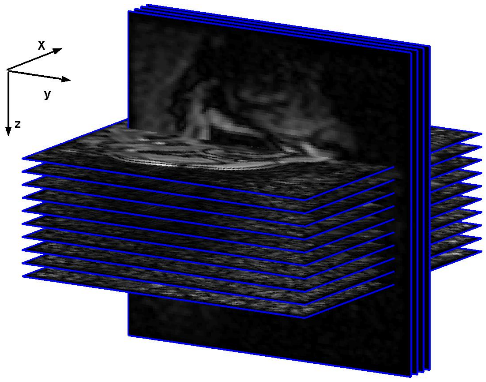

Fig. 2. 3D tongue MR images.

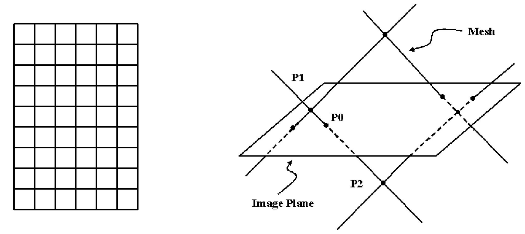

Fig. 3. (a) The shape of a 3D tongue mesh. (b) A 3D mesh intersects with an image plane.

not change when it moves. This is called the phase invariance property and can be expressed as:

$$\phi(\mathbf{x}(t_n)) = \phi(\mathbf{x}(t_{n+1}))\,, \tag{2}$$

where $\phi(\mathbf{x}(t_n))$ is the phase of material point $\mathbf{x}$ at time t_n.

In 2D HARP, two tag orientations are required to track the in-plane motion of a material point. Hence the phase $\phi = [\phi_1, \phi_2]$ is a 2D vector, where ϕ_i is the HARP value at i^{th} tag orientation. The 2D in-plane motion is tracked using optical flow method based on this property [1].

2.2. Thin-plate Spline Interpolation

The thin-plate spline [6] is a widely used interpolation method for scattered data samples. In our method, the 3D motion of material points is interpolated using 3D thin-plate spline. Let the velocity vector be $\mathbf{v} = (v_x, v_y, v_z)^t$. If the motions of a set of materials points $\mathbf{x_i} = (x_i, y_i, z_i)^t$ are known as $\mathbf{v_i} = (v_x^i, v_y^i, v_z^i)^t$, the 3D motion on any point $\mathbf{x}$ in the 3D space is expressed as

$$v_x(x,y,z) = a_0^x + a_1^x x + a_2^x y + a_3^x z + \sum_{i=1}^{N} w_i^x r_i^2 \ln r_i^2 \tag{3}$$

$$v_y(x,y,z) = a_0^y + a_1^y x + a_2^y y + a_3^y z + \sum_{i=1}^{N} w_i^y r_i^2 \ln r_i^2 \tag{4}$$

$$v_z(x,y,z) = a_0^z + a_1^z x + a_2^z y + a_3^z z + \sum_{i=1}^{N} w_i^z r_i^2 \ln r_i^2 \tag{5}$$

where $r_i^2 = (x-x_i)^2 + (y-y_i)^2 + (z-z_i)^2$. The $3(N+3)$ coefficients can be solved in the similar way as 2D TPS in [6].

3. METHOD

In order to track 3D motion, three tag orientations are required, which implies that at least two image orientations — we use sagittal and axial in the tongue — are required. In order to cover 3D space, image stacks are acquired in each orientation, as illustrated in Fig. 2. The three tag orientations are acquired by having both vertical and horizontal tags in the axial images and horizontal tags in the sagittal images.

Our TPS-based 3D-HARP works in an iterative manner. A rectangular planar mesh is initialized in the reference time and the initial phase values of grid points of this mesh are computed based on their position in space and knowledge of the tag pattern. 2D-HARP methods are applied on each of the image sequences, thereby giving partial knowledge of the components of 3D motion on all image planes. The points of intersection between the mesh and the image planes are then computed, which give a sparse set of motion observations on the mesh (from the previous 2D HARP computations). Each component of motion is then interpolated using a 3D thin-plate spline so that the motion on all nodes of the mesh are known. The mesh position is updated until no motion is implied anywhere on the mesh (within a small tolerance) and then time is incremented, and these steps are repeated. We now provide details on this overall approach.

We assume that the three tag directions (two on one image orientation, and one on the other) are orthogonal. The image coordinate system's axes are defined as follows (Fig. 2.2): The x and y axes are orthogonal to the two tag orientations in the first view, and the z axis is orthogonal to the tag orientation in the second image view. The image planes of the first view are parallel to the xy plane, and the image planes of the second view are parallel to the yz plane.

Step 1. Mesh Initialization

The initial mesh M^0 is a rectangle in the reference time frame, as depicted in Fig. 3(a). It is parallel to the xz plane so that it intersects both of the two view images. The mesh is placed completely inside the tongue in order to avoid tracking error. The mesh is represented by m by n grid points connected by straight lines. We denote the initial 3D phase of grid point $\mathbf{P}$ as $\phi^0(\mathbf{P}) = [\phi_x^0(\mathbf{p})), \phi_y^0(\mathbf{p}), \phi_z^0(\mathbf{p})]^T$. For every grid point, its initial phase is achieved either by knowledge from the imaging process or from the reference time frame. For every time $t > 0$, the mesh is initialized as the computed mesh from the previous time $t-1$. The following steps are then executed iteratively.

Step 2. Calculation of the intersection points on the xy image planes.

For each image slice parallel to xy plane, the intersection points of the mesh and the image plane are calculated. This is done by finding mesh edges whose vertices are on different sides of the image plane, as illustrated in Fig. 3(b). The intersection points are then calculated using linear interpolation. The initial phase values of these points are also calculated as linear interpolation of the phase values of the edge vertices.

Step 3. 2D motion tracking in x and y axes and interpolation.

For every intersection point $\mathbf{x}_i$ calculated in the previous step, its 2D in-plane motion in x and y axes, $v_x(\mathbf{x}_i)$ and $v_y(\mathbf{x}_i)$, are tracked using the basic 2D HARP method. Then for every grid point $\mathbf{x}_i^t$ of the mesh, the x and y motion, $v_x(\mathbf{x}_i^t)$ and $v_y(\mathbf{x}_i^t)$, is then interpolated using equation (3) and (4).

Assuming the motion in z-direction is zero, the mesh is updated by

$$\mathbf{x}^t = \mathbf{x}^t + \mathbf{v}(\mathbf{x}^t) \tag{6}$$

where $\mathbf{x}^t$ is any mesh grid point at time t, and $\mathbf{v}(\mathbf{x}^t) = [v_x(\mathbf{x}^t), v_y(\mathbf{x}^t), 0]^T$. By updating the mesh after each coordinate update (unlike [5]), there is less risk of a tracking error, and the algorithm should converge faster.

Step 4. Calculation of the intersection points on yz image planes.

The intersection points are calculated in the same way as in Step 2.

Step 5. 1D motion tracking in z axis and interpolation.

Since there are only horizontal tags in the xz image slices, it is assumed that the intersection points do not move in the y direction. Accordingly, the points are tracked in the z direction only, and the z component of motion on every mesh grid point is calculated using (5). The mesh is updated using (6) where $\mathbf{x}^t$ is any mesh grid point at time t, and $\mathbf{v}(\mathbf{x}^t) = [0, 0, v_z(\mathbf{x}^t)]^T$.

Step 6. Phase invariance checking

Because HARP phase is a material property, the correct positioning of the mesh is found when the phase at all intersection points agree with the data. Therefore, the phase invariance property is checked on all intersection points of the updated mesh and all image planes. The initial phase of the k-th intersection point at the current time is denoted as $\phi_{\mathbf{k}}^{\mathbf{t}} = [\phi_{x,k}^t, \phi_{y,k}^t, \phi_{z,k}^t]^T$. If this point is on the first view image planes, only the x and y phase can be calculated. The phase difference of this point is defined as:

$$\delta_k = \max(|\phi_{x,k}^t - \phi_{x,k}^0|, |\phi_{y,k}^t - \phi_{y,k}^0|) \tag{7}$$

If this point is on the second view image planes, only the z phase can be calculated. the phase difference is then

$$\delta_k = (|\phi_{z,k}^t - \phi_{z,k}^0|) \tag{8}$$

If phase differences of all intersection points are all less than a threshold ϵ, the phase invariance condition is satisfied.

Steps 2–6 are repeated until the phase invariance condition is satisfied. Then time is incremented and the process is repeated. When all time frames have been used, the process is complete.

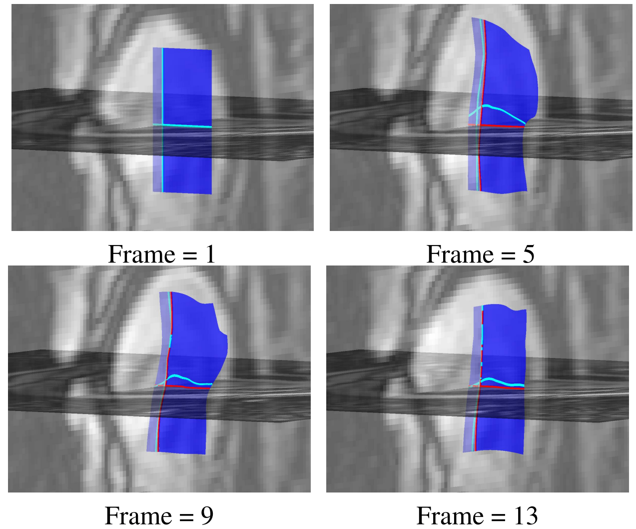

Fig. 4. The mesh motion at four different time frames. On each time frame, the up-down direction is the top-bottom direction of head, and left-right direction is the anterior-posterior direction of head. The red lines are the intersection of the mesh and the displayed image planes. The intersection line on the first time frame is tracked in the subsequent frames and displayed in cyan.

4. EXPERIMENT RESULTS

We applied the algorithm on the tagged MR images of the tongue. The MR image slices were acquired on a Philips Eclipse 1.5 T scanner when the subject uttered “deGoose.” The acquired image data includes horizontally and vertically tagged images in the axial view, and horizontally tagged images in the sagittal view. The imaging parameters were: the image size is 128×128, pixel spacing is 1.56 mm, slice thickness is 5 mm, tag spacing is 5 mm, flip angle is 10. Ten axial and four sagittal image slices were acquired at each of 18 time frames. The separation between sagittal images is 7 mm, and between axial images is 8 mm. The interval between the time frames was 49 ms. The 10 axial slices covered the whole tongue, and the 4 sagittal slices went from the middle to the left side of the tongue.

The mesh was initialized at the reference time frame. It is 19 mm by 38 mm rectangle that is parallel to the xz plane.

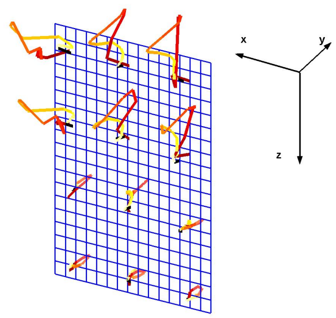

Fig. 5. The 3D pathlines of selected points on the mesh tracked temporally. The brighter color means the path of a later time frame.

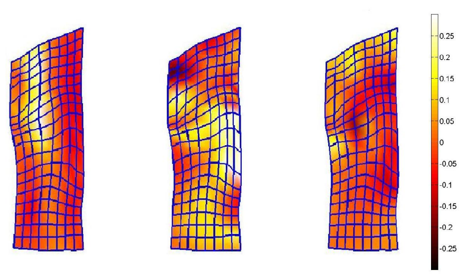

Fig. 6. The strain map along three axes directions at time frame 4. From left to right: the strain map along x, y, and z axes.

At the reference time frame, the mesh intersects with 5 axial slices and 3 sagittal slices. The relative location of the mesh and the image slices is illustrated in Fig. 4. After initialization, the mesh was tracked in all 18 time frames.

Fig. 4 shows the tongue mesh tracking result. The displayed image is on the mid-sagittal plane passing through the center of the tongue. Fig. 5 shows the pathlines of several selected points that are tracked in time. From these figures we can see the tongue moves more in y and z axes, and less in x axis. The upper part of the tongue moves more than that on the lower part. Fig. 6 shows the strain map on the mesh in the three axes. The mesh is projected onto xz plane. Fig. 7 shows the positions of the intersections of tracked meshes with a tagged sagittal image. This enables direct comparison of automatic mesh tracking with visual assessment of tag motion.

This algorithm was implemented in Matlab on a 2.8 GHZ Intel Pentium 4 computer with 512 MB RAM. In our implementation it took about 10 seconds for each time frame.

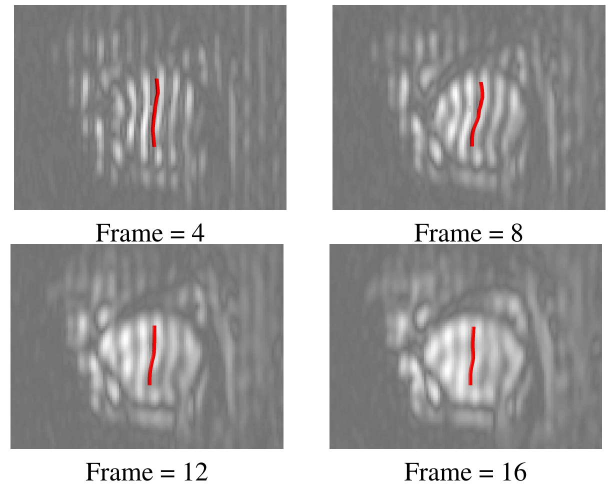

Fig. 7. The intersection lines of the tracked mesh one vertically tagged sagittal image in different time frame. The image orientation is the same as in Fig. 4. The intersection lines are shown in red.

5. CONCLUSION

In this paper we presented a fast method to track tissue motion in three dimensions from tagged MR images using thin-plate spline interpolation. Experiments on real MR images show this method can track tongue motion smoothly and accurately.

6. REFERENCES

[1] Nael F. Osman and Jerry L. Prince, "Visualizing myocardial function using harp mri," *Phys. Med. Biol.*, vol. 45, pp. 1665–1682, 2000.

[2] Vijay Parthasaranthy, Moriel Ness-Aiver, and Jerry L. Prince, "Tracking tongue motion from tagged magnetic resonance images using harmonic phase mri (harp-mri)," *Proc. Of 15th International Congress of Phonetic Sciences*, 2003.

[3] Salome Ryf, Marcus A. Spiegel, Michael Gerber, and Peter Boesiger, "Myocardial tagging with 3d-cspamm," *J. Magn. Res Imag.*, vol. 16, no. 3, pp. 320–325, 2002.

[4] I. Haber and C.F. Westin, "Model-based 3d tracking of cardiac motion in harp images," in *Int. Soc. Mag. Reson. Med.*, 2002.

[5] Li Pan, Jerry L. Prince, Joao A. C. Lima, and Nael F. Osman, "Fast tracking of cardiac motion using 3d-harp," *IEEE Trans. Biomed. Eng.*, vol. 52, no. 8, pp. 1424–1435, 2005.

[6] Fred L. Bookstein, "Principal warps: thin-plate splines and decomposition of deformations," *IEEE Trans. Patt. Anal. and Mach. Intell.*, vol. 11, no. 6, pp. 567–585, 1989.