LEARNING-BASED 3D MYOCARDIAL MOTION FLOW ESTIMATION USING HIGH FRAME RATE VOLUMETRIC ULTRASOUND DATA

Yang Wang, Bogdan Georgescu, Dorin Comaniciu
Siemens Corporate Research
Princeton, NJ 08540

Helene Houle
Siemens Ultrasound
Mountain View, CA 94043

ABSTRACT
The estimation and analysis of cardiac motion provides important information for the quantification of the elasticity and contractility of the myocardium. Taking advantage of the recent progress on real-time ultrasound imaging, unstitched volumetric data can be captured in a high frame rate. In this paper, we propose a learning-based method to automatically estimate the 3D displacements and velocities of the myocardial motion. To achieve robust tracking on ultrasound image sequences, multiple information is fused together in our framework to handle noisy and missing data, including speckle patterns, boundary detection and motion prediction. Preliminary results on clinical data confirmed these findings in a qualitative manner. The estimated displacement and velocity values have a strong agreement with the results from other systems and modalities. The proposed method is efficient and achieves high speed performance of less than 1 second per frame for volumetric ultrasound data.

Index Terms— Cardiac Deformation, Boundary Detection, Motion Tracking, Learning-Based Methods

1. BACKGROUND
Recent developments in the field of echocardiography have allowed the cardiologist to quantify cardiac deformation in a non-invasive manner. However, most existing methods for measuring myocardial motion are often limited to measurements in one or two dimensions, including Doppler myocardial imaging [1], speckle tracking [2], and image registration approaches [3]. Although visual wall motion scoring is the clinically established method towards the assessment of myocardial function, this methodology has been proven to be highly variable between observers [4]. Furthermore, a manual initialization and segmentation step is often required for the tracking process [5, 3, 6, 7, 8]. A few studies have been carried out to analyze the change of myocardium volume during a cardiac cycle [9]. The common conclusion is that the volume of the ventricular wall remains relatively consistent during the cardiac cycle and the change is less than 5% [10]. Since myocardial tissue is virtually incompressible, it deforms in all three dimensions simultaneously. Therefore, it is important to compute the cardiac deformation in the three-dimensional space. It was validated through several clinical studies for quantification of left ventricle (LV) function as reviewed in [11].

In this paper we propose a new approach to estimate the 3D velocity of the ventricular wall from high-frame rate volumetric ultrasound images in an automatic manner. Compared to the existing methods, such as image registration [7] and optical flow [8], the proposed framework has the following advantages:

- This system is automatic. Real-time boundary detection is performed in the first frame using the marginal space learning (MSL) approach [12] to initialize both the endocardial and epicardial boundaries.
- Information from multiple cues, including speckle patterns, image gradients, boundary detection and motion prediction, is fused into a single Bayesian objective function to improve tracking accuracy and robustness.
- Image quality measurements based on image intensities and speckleness scores are integrated in a weighted likelihood estimation to handle noise and signal dropouts in ultrasound data.

To demonstrate the performance we evaluated our method on clinical data. Preliminary results confirmed these findings in a qualitative manner, as the estimated displacement and velocity values are consistent with the results from other systems and modalities.

2. FRAMEWORK
In this section, we present the new framework to estimate 3D myocardial motion. As illustrated in Fig. 1, our system includes the following main steps:
1. Automatic tracking initialization: The initial endocardial and epicardial boundaries of LV are first detected using Marginal Space Learning (MSL) [12].

2. Dense tracking of the myocardium: The 3D deformation of the volumetric myocardium is captured by fusing information from multiple cues, including speckle tracking, boundary detection and motion prior.

3. 3D myocardial displacement and velocity estimation: Based on the dense 3D tracking of the myocardium volume, the displacements and velocities in both the Cartesian and local heart coordinate system.

2.1. Initialization by Detection
In the starting frame (typically the end-diastole cardiac phase), we initialize the tracker automatically by detecting the endocardial and epicardial boundaries of the left ventricle (LV). A 3D detector is learned to locate the pose, including the position \( X = (x, y, z) \), orientation \( \theta = (\alpha, \beta, \gamma) \) and scale \( S = (s_x, s_y, s_z) \), of the LV using the marginal space learning (MSL) approach [12]. The local deformations of the myocardial boundaries are further estimated based on the posterior distribution \( p_t(X|I) \) of each control point on the surface, which is learned using the steerable features and the probability boosting-tree (PBT) [13].

2.2. 3D Motion Tracking
In order to estimate myocardium deformations, dense tracking of the cardiac motion is required to establish the inter-frame correspondences for each point on both the endocardial and epicardial boundaries. This task is particularly challenging for the ultrasound data because of the noise and signal dropouts [8]. Inspired by [14] we propose to fuse information from multiple cues into a single Bayesian framework as follows

\[
\arg \max_{\vec{X}_t} p(\vec{X}_t|\vec{Y}_{0:t}) = \arg \max_{\vec{X}_t} p(\vec{Y}_{0:t}|\vec{X}_t)p(\vec{X}_t|\vec{Y}_{0:t-1})
\]

where \( \vec{Y}_{0:t} = \vec{Y}_0, \ldots, \vec{Y}_t \) are the measurements from the input image sequence \( I_{0:t} = I_0, \ldots, I_t \). For clarity, we use \( \vec{X}_t \) to denote a concatenation of the mesh point positions, \( \vec{X}_t = [X_1, \ldots, X_n] \), which need to be estimated at the current time instance \( t \), and \( n \) is the total number of points in the mesh model.

Assuming the Markovian structure of the motion, Eqn. 1 can be solved in a recursive manner:

\[
\arg \max_{\vec{X}_t} p(\vec{X}_t|\vec{Y}_{0:t}) = \arg \max_{\vec{X}_t} p(\vec{Y}_t|\vec{X}_t) \prod_{t=1}^{T} p(\vec{X}_t|\vec{X}_{t-1})p(\vec{X}_{t-1}|\vec{Y}_{0:t-1})
\]

Prediction Estimation: In Eqn. 2, \( p(\vec{X}_t|\vec{X}_{t-1}) \) is the transition probability function \( \tilde{p}(\vec{X}_t|\vec{X}_{t-1}) \) augmented by an incompressibility constraint \( p(f_V(\vec{X}_t)|f_V(\vec{X}_{t-1})) \) as follows

\[
p(\vec{X}_t|\vec{X}_{t-1}) = \tilde{p}(\vec{X}_t|\vec{X}_{t-1})p(f_V(\vec{X}_t)|f_V(\vec{X}_{t-1}))
\]

where \( f_V(\vec{X}) \) is the volume enclosed by the mesh \( \vec{X} \) and \( \tilde{p}(\vec{X}_t|\vec{X}_{t-1}) \) can be obtained from the training data set based on a non-linear manifold learning technique [15]. For simplicity, we model \( p(f_V(\vec{X}_t)|f_V(\vec{X}_{t-1})) = p(f_V(\vec{X}_t) - f_V(\vec{X}_{t-1})) \) by a zero mean Gaussian distribution \( \mathcal{N}(0, \sigma_V) \) with \( \sigma_V \) learned from the training data.

Likelihood Estimation: To maximize the accuracy and robustness of the tracking performance, the likelihood term \( p(\vec{Y}_t|\vec{X}_t) \) is computed from both boundary detection and speckle template matching as follows,

\[
p(\vec{Y}_t|\vec{X}_t) = (1 - \lambda_k)p(y_b|\vec{X}_t) + \lambda_k p(T_t|\vec{X}_t)
\]

where \( y_b \) is the steerable feature response [12], \( T_t \) is the speckle pattern template, and \( \lambda_k \) is the weighting coefficient of the matching term. In our approach \( \lambda_k \) is computed based on the speckleness measure as follows

\[
\lambda_k = \frac{1}{1 + e^{-f_c(I_t(\vec{X}_t), T_t)}}
\]

where \( f_c(I_t(\vec{X}_t), T_t) \) is the intensity covariance between the image block \( I_t(\vec{X}_t) \) centered at \( \vec{X}_t \) and the speckle template \( T_t \), \( \sigma(I_t(\vec{X}_t)) \) and \( \sigma(T_t) \) are the intensity variance of the image block \( I_t(\vec{X}_t) \) and the speckle template \( T_t \), respectively. To handle the temporal image variation, the speckle template \( T_t \) is also updated online using the image intensities \( I_t(\vec{X}_{t-1}) \) from the previous frame \( t - 1 \).
2.3. Displacement and Velocity Estimation

Given the tracking result \( \vec{X} \) from Section 2.2, displacements and velocities can be computed in the three-dimensional space as shown in Fig. 2. A local heart coordinate system has also been introduced to describe the LV deformation [4]. As illustrated in Fig. 3, the three directions of the heart are defined as longitudinal (meridional) \( D_L \), radial (transmural) \( D_R \), and circumferential \( D_C \). Each point position \( X \) is then projected from the Cartesian coordinate system to the local cardiac coordinate system, \( X' = (X(L), X(R), X(C)) \).

![Fig. 3. Local heart coordinate system. Radial (\( D_R \)): perpendicular to the epicardium, pointing inwards. Longitudinal (\( D_L \)): perpendicular to the radial axis and pointing towards the apex of the ventricle. Circumferential (\( D_C \)): perpendicular to the radial and longitudinal direction and directed anticlockwise around the classical echo short axis image when looking from apex to base.](image)

The longitudinal and radial displacements can then be computed as \( Z_t^{L} = X_t^{L} - X_{t-1}^{L} \) and \( Z_t^{R} = X_t^{R} - X_{t-1}^{R} \), respectively. The circumferential displacements are computed as the rotation angle, \( Z_t^{C} = \arccos(D_{Rt}, D_{Rt-1}) \), where \( , \) denotes the dot product. \( Z_t^{C} \) is defined as positive if the rotation is counter-clockwise viewed from the apex, and negative if clockwise. The velocity can be computed as dividing the displacements by the acquisition time step \( t_f \) of the input 3D+t ultrasound sequence.

3. EXPERIMENTAL RESULTS

In this section, we demonstrate the performance of the proposed automatic detection and tracking method for the left ventricle myocardium. The high frame-rate 3D+t ultrasound sequences were acquired by a Siemens SC2000 system with the average volume size of 200 × 200 × 140 and resolution of 1 mm in the x, y, and z directions. Our proposed method is computationally efficient and the average processing time on a 3.0GHz PC machine is less than 1 second per frame.

To evaluate the accuracy and robustness of our learning-based detection and tracking method, we tested it on a large data set including 503 volumetric ultrasound sequences from human subjects. The data set was split into a training set with 239 sequences and a testing set with the remaining 264 sequences. The training set was used to learn the detectors in section 2.1 and the prior distributions in section 2.2, while the testing set reflected the performance for unseen data. Table 1 reports the error measurements based on the average point distance between our resulting meshes and the ground-truth annotations by experts on the end-diastolic and end-systolic frames. The low error values on both the training and testing data demonstrate the high accuracy performance of our learning-based method on both seen and unseen data.

<table>
<thead>
<tr>
<th>measure(mm)</th>
<th>Training (239 sets)</th>
<th>Testing (264 sets)</th>
</tr>
</thead>
<tbody>
<tr>
<td>mean/std</td>
<td>2.21/1.57</td>
<td>2.68/2.63</td>
</tr>
<tr>
<td>median</td>
<td>1.87</td>
<td>2.20</td>
</tr>
</tbody>
</table>

Table 1. Performance analysis on a large data set including 503 3D+t ultrasound sequences in total. The errors are computed as the average point distance between our resulting meshes and the ground-truth annotations by experts on the end-diastolic and end-systolic frames. The low error values on both the training and testing data demonstrate the high accuracy performance of our learning-based method on both seen and unseen data.

To evaluate the accuracy of our displacement estimation, we compared the results against the expert measurements of the same subject using a validated system, the 2D VVI [16]. 20 studies were collected with both 3D and 2D ultrasound sequences for each subject. The projected displacements were computed by our system based on the 3D sequences, while the corresponding 2D displacements were obtained by the VVI system on the 2D sequences. The correlation scores reported in Table 2 show clearly that the results from our method are consistent with the clinic evaluation. The relatively low scores in the apical region might be caused by the fact that the 2D images can not pass through the true apex consistently during the whole cardiac cycle.

<table>
<thead>
<tr>
<th></th>
<th>Basal</th>
<th>Middle</th>
<th>Apical</th>
</tr>
</thead>
<tbody>
<tr>
<td>Longitudinal</td>
<td>0.89</td>
<td>0.87</td>
<td>0.60</td>
</tr>
<tr>
<td>Radial</td>
<td>0.88</td>
<td>0.83</td>
<td>0.71</td>
</tr>
</tbody>
</table>

Table 2. Example correlation scores of the longitudinal (L) and radial (R) displacements between our method and the expert measurements using 2D VVI. The correlation score is normalized to \([−1, 1]\), with a higher value meaning a stronger agreement. It shows clearly that our estimation is consistent with the clinical evaluation.

Furthermore, since tissue Doppler imaging (TDI) is a popular method to estimate the velocity of the cardiac motion, we projected the estimated velocities along the longitudinal direction and compared them against the expert measurements for the same subjects using TDI. Fig. 4 shows an example from our comparison experiment. Fig. 4(c) plots the longitudinal velocities estimated by our method, while Fig. 4(d) is the expert measurements for the same subject using TDI. The similar amplitude and shape between plots (a) and (b) shows an agreement between two methods. In the future work, we will conduct in-vitro experiments to further validate our method in a clinical setting.

4. CONCLUSION

In this paper, we proposed an automatic framework to estimate the 3D flow of the myocardial motion on high-frame rate volumetric ultrasound data. The advantages of our new
method includes: (1) computing displacements and velocities of the myocardial motion in the three-dimensional space and projecting them to the local heart coordinate system, (2) performing real-time boundary detection to automatically initialize the myocardial boundaries in the first frame, (3) fusing information from multiple cues into a single Bayesian objective function to achieve accurate and robust tracking, and (4) integrating image quality measurements based on image intensities and speckleless scores to handle noise and signal dropouts in ultrasound acquisitions. Comparison experiments on clinical data confirmed these findings in a quantitative manner. The proposed method is efficient and achieves high speed performance of less than 1 second per frame for volumetric ultrasound data.

5. REFERENCES


