Automated Intraventricular Septum Segmentation Using Non-local Spatio-temporal Priors

Mithun Das Gupta[1], Sheshadri Thiruvenkadam[1], Navneeth Subramanian[1], and Satish Govind[2]

[1] John F. Welch Technology Center, GE Global Research, Bangalore, India
[2] Narayana Hrudayalaya, Bangalore, India
{mithun.dasgupta,sheshadri.thiruvenkadam,navneeth.s}@ge.com, drsatishgovind@yahoo.com

Abstract. Automated robust segmentation of intra-ventricular septum (IVS) from B-mode echocardiographic images is an enabler for early quantification of cardiac disease. Segmentation of septum from ultrasound images is very challenging due to variations in intensity/contrast in and around the septum, speckle noise and non-rigid shape variations of the septum boundary. In this work, we effectively address these challenges using an approach that merges novel computer vision ideas with physiological markers present in cardiac scans. Specifically, we contribute towards the following: 1) A novel 1-D active contour segmentation approach that utilizes non-local (NL) temporal cues, 2) Robust initialization of the active contour framework, based on NL-means de-noising, and MRF based clustering that incorporates physiological cues. We validate our claims using cardiac measurement results on ∼30 cardiac scan videos (∼2000 ultrasound frames in total). Our method is fully automatic and near real time (0.1sec/frame) implementation.

1 Introduction

In echo-cardiography, wall and chamber dimensions are used as screening parameters for early indication of cardiac diseases. In this work, we address the interventricular septum thickness (IVSd), which is accepted as a screening parameter for septal hypertrophy and has also shown a correlation to 24 hour ambulatory blood pressure. Unfortunately, the manual measurement of these parameters on echo-cardiograms suffers from large inter and intra observer variability based on the experience and expertise of the cardiologist. An end-to-end IVSd measurement in accordance with the American Society of Echocardiography (ASE) guidelines [1] involves segmentation of the septum, identification of mitral valve tip, and measurement of the thickness orthogonal to septum centerline.

Automation of septum segmentation from ultrasound images faces several challenges. First, the intensity in and around the septum region has a multimodal distribution, precluding simple schemes based on intensity alone. Second,

N. Ayache et al. (Eds.): MICCAI 2012, Part I, LNCS 7510, pp. 683–690, 2012.

the speckle noise inherent in ultrasound images makes edge-based methods unreliable. These methods are sensitive to initialization and are plagued by the near-field haze which leads to low contrast at the upper part of the septum. For any approach to segment the septum across several frames reliably, robustness to noise and initialization is needed.

Region based active contour approaches [2,3] typically offer robustness to initialization and noise. Due to very low contrast at the upper part of the septum, purely region based methods are likely to get trapped in local minima leading to the introduction of shape priors [4]. Given the large inter-patient shape variability and non rigid deformations across frames, it is infeasible to build a shape atlas of the septum. In Subramanian et al. [5], a region based active contour approach with a width prior for the septum is proposed. Constraining the width of the septum, makes the resulting segmentation robust to noise/inhomogeneities induced by near field haze. However, the width constraint is suited for advancing the profiles closer to the septum boundary and fails in cases where the septum boundary has low contrast. Further, [5] relies on a reasonable guess for the contour in the first frame, and in the absence of temporal cues, the segmentation is most likely to drift away in cases of large non-rigid motion between frames.

Cardiac ultrasound data has rich temporal information in terms of continuity of motion of structures which could provide reliable information on the location of the boundary under boundary gaps and low-contrast regions. Hence, it seems natural to model either velocity/acceleration using previous frames [6]. Consequently, we introduce temporal constraints to the segmentation cost functional. We constrain acceleration of boundary points to be similar in a non-local neighborhood around each boundary location. The non-local temporal penalty differentiates our work from the state-of-the-art [5], and is key in robust prediction of the septum location in the absence of strong contrast for the segmentation functional. The second part of our work deals the important aspect of initializing the 1D contours for the segmentation technique. We propose to mitigate the non-trivial problem of initializing the septum boundary using physiological cues. We propose a robust routine based on non-local (NL)-means de-noising, and Markov Random field (MRF) based clustering.

This paper is organized as follows. In Sec. 2, we describe the details of algorithms pertaining to each of the individual pieces. In Sec. 3, we present our results on 32 patients, and our conclusions are listed in Sec. 4.

2 Methods

In a typical echocardiography scan, usually 3 cardiac loops are captured, where one loop is defined as end-diastolic frame to the next end-diastolic frame. Based on the frame rate of the scanner usually about 100 frames are captured in this loop. In this work we look at the parasternal long axis (PLAX) view. The mitral valve tip detection is handled identical to the method proposed by Subramanian et al. [5].

2.1 1D Active Contour Formulation Using Non-local Temporal Priors

Similar to [5], we propose an energy based formulation with a search space of pairs of smooth 1D profiles (representing the top and bottom boundaries of the septum). This representation enables easy access to regional statistics in and around the septum, and model interactions between the top and bottom septum boundaries. Further, the simplified representation makes the approach faster than 2D active contour approaches, making it feasible for real time tracking. The main contribution of this work is to incorporate temporal cues into the above framework to improve robustness to noise, gaps, and large motion.

For image $I : \Omega \rightarrow R$, $\Omega = [a, b] \times [c, d]$, we look for two smooth 1D functions $g, f : [a, b] \rightarrow [c, d]$, whose profiles represent the top and bottom parts of the septum. We denote the septum region between the 1D profiles of f and g, as R^s. We denote the neighborhood region above the septum as R^{up}, i.e. between profiles of g and $g + \Delta$, where Δ is some pre-defined interval. Similarly, the neighborhood below the septum is R^{dn}, between profiles of $f - \Delta$ and f. We make a piece wise constant assumption of intensity in and around a neighborhood of the septum. Consequently, we divide $(a, b) = \bigcup_{i=1}^{K}(a_i, b_i)$ into K disjoint intervals and seek for f, g that give homogeneous distributions in regions $R_i^{up} = R^{up} \cap ([a_i, b_i] \times [c, d])$, $R_i^{dn} = R^{dn} \cap ([a_i, b_i] \times [c, d])$ and $R_i^s = R^s \cap ([a_i, b_i] \times [c, d])$.

The following energy is minimized over the space of smooth 1D functions $f, g : [a, b] \rightarrow [c, d]$ and mean statistics $\mu^{up}, \mu^s, \mu^{dn}$:

$E_{frm}(f, g, \mu^{up}, \mu^s, \mu^{dn}) =$

$$\sum_{i=1}^{K} \Big[\int_{R_i^s} (I - \mu_i^s)^2 dydx + \int_{R_i^{up}} (I - \mu_i^{up})^2 dydx + \int_{R_i^{dn}} (I - \mu_i^{dn})^2 dydx \Big]$$

$$+\lambda_{width} \int_a^b (f + w - g)^2 \, dx + \lambda_{smooth} \int_a^b \left(\sqrt{1 + (f')^2} + \sqrt{1 + (g')^2} \right) dx \quad (1)$$

The data term drives f, g to take piece wise constant values in each of R_i^s, R_i^{up}, R_i^{dn}. The smoothness terms for f and g are governed by parameter $\lambda_{smooth}(0.08)$. The width term constrains the width of R^s to be close to the expected septum width (w) and is balanced by $\lambda_{width}(\sim 1.0\text{cm})$. In most cases, because of strong contrast between the septum and blood pool, the bottom boundary f of the septum is reliably segmented and the width term drives the top boundary g out of local minima closer to the actual boundary. From here, contrast close to the septum boundary takes over and drives the segmentation. In low contrast cases, one would expect the evolution to be dominated by the smoothness and width terms and be drawn to arbitrary minima.

We now augment the above energy with temporal priors computed from previous frames. The septum being an elastic structure exhibits motion that is correlated across different locations, which can be captured through temporal priors resulting in robust prediction of the septum in the absence of strong contrast. Given the large shape variations and complex non-rigid motion of the septum, it seems natural to model either velocity/acceleration using previous

frames similar to [6]. If f^{n-2}, f^{n-1}, f^n are the profiles at frames $n-2, n-1, n$, one could penalize velocity $V^n = f^n - f^{n-1}$ using $\int_a^b (\frac{dV^n}{dx})^2 dx$, or penalize acceleration $A^n = f^{n-2} - 2f^{n-1} + f^n$ using $\int_a^b (\frac{dA^n}{dx})^2 dx$. Here, we look at a non-local penalty for acceleration defined by $\int_a^b \int_a^b w(x,y)(A^n(x) - A^n(y))^2 dxdy$, where w is a weight function for the pair (x, y). The choice of non-local priors is physically intuitive in that velocity interactions between pixels extend beyond local neighborhoods and modeling these non-local interactions might give better robustness to noise/boundary gaps. The Euler Lagrange equations of the above acceleration penalty is $A^n - w * A^n = 0$ or in terms of velocity $V^n - w * V^n = V^{n-1} - w * V^{n-1}$. Thus velocity at each point on the profile f is updated using relative velocities learnt in the previous frame and velocities of non-local neighboring points. As for the choice of w, for simplicity we use $w(x,y) = G_\sigma(|x-y|)$, thus the above equations involve only convolutions and would be fast to compute. Alternatively, one could consider intensity/contrast dependent terms for w to down-select points to learn relative velocities. Note that in the work proposed by Snare et al. [7], the motion prior is local wherein, motion models at neighboring control points do not interact with each other. Thus if a control point falls in a poor contrast/signal dropout location, for a couple of frames, the segmentation will drift away. We modify Eq. 1 to include the temporal term. For each frame n, denote $A^n = f^{n-2} - 2f^{n-1} + f^n$ and $B^n = g^{n-2} - 2g^{n-1} + g^n$. We minimize:

$$E^n_{temp} = E_{frm} + \lambda_T \int_a^b \int_a^b w(x,y)((A^n(x) - A^n(y))^2 + (B^n(x) - B^n(y))^2) dxdy \tag{2}$$

Given profiles from previous time points $f^j, g^j,\ j = n-2, n-1$, E_{frm} is minimized, using descent on the Euler Lagrange equations for Eq. 2, using an explicit finite difference scheme. The following synthetic results show the robustness of the temporal prior. In Fig. 1, the goal is to segment the current frame (III) from two clean previous frames shown in (I), (II). In these experiments, segmentation results are shown in Red, and the ground truth is in Green.

In the noisy (white noise) experiment shown in Fig. 1 (top row), (a) is the noisy frame, (b) is the result without temporal information, (c) is the result using non-local temporal priors (proposed approach). In (d), even for really high levels of noise, a reasonable segmentation is computed. In the second experiment Fig. 1 (bottom row), on segmenting in the presence of gaps, (e), (f) show results without temporal priors. The kink in the lower and upper boundaries is lost, and the missing pieces are completed with straight lines because of the smoothness term. In (g)-(i) using our approach, for increasing size of the gap, the kink is seen to be preserved. The reasoning is that at a location with poor contrast, gap or high noise, a non-local neighbor where the contrast is possibly good, directly contributes to the update at the location using relative velocities learnt from previous frames.

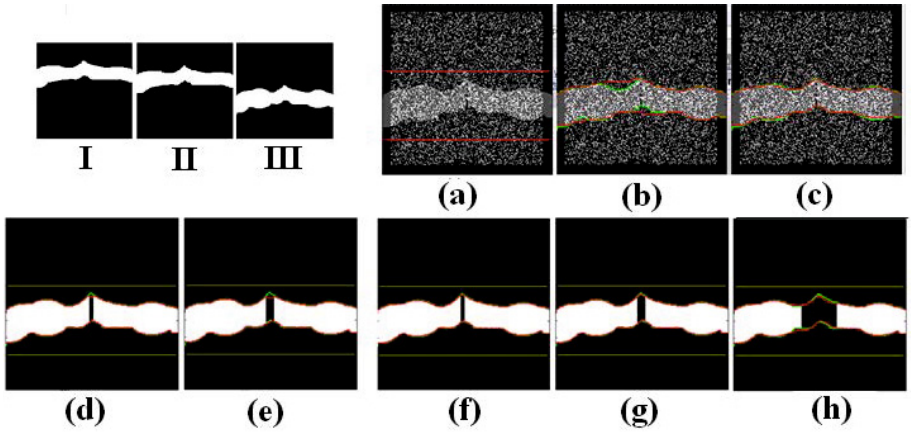

Fig. 1. Synthetic Experiments: (I) Frame n-2 (II) Frame n-1 (III) Current Frame n. Top Row: Noise Robustness: (a) Noisy image with initialization (b) Result w/o temporal prior (c) Result using temporal approach. Bottom Row: Robustness to Gaps: (d)-(e) Results w/o temporal prior (f)-(h) Results using temporal approach. Groundtruth is in green and the algorithm output is shown in red.

2.2 Active Contour Initialization Using Blood Pool Detection

One of the primary drawbacks of active contour based methods [5] for segmentation is the inherent need for correct initialization. The initial contour needs positioned reasonably close to the final contour to guarantee convergence. For completely automated system required to work for large patient populations, initialization has to be driven by physiological cues such that patient variability can be properly captured [8]. The LV blood pool is an "anechoic" region (no oscillating sources in it) and hence it appears completely dark under ultrasound. Based on this key observation, we propose to detect the LV blood pool to infer an estimate for the lower boundary of the septum and initialize the curves. The first step in our approach is a denoising algorithm using Non-Local means, and the second is an MRF based clustering technique to find the maximum width of the blood pool. Once the blood pool is identified, the septum is its immediate neighbor towards the ultrasound probe position. This knowledge can now be used to initialize an active contour method for segmentation.

Neighborhood Based Non-local Smoothing. The key intuition for non-local (NL) means based filtering, as proposed by Buades et al. [9] is that the denoised value at location x is a mean of the values of all points within the image domain whose Gaussian neighborhood is similar to the neighborhood of x. Given a discrete noisy image $\mathbf{v} = \{v(i)|i \in I\}$, the estimated value $NL(\mathbf{v})(i)$ is computed as a weighted average of all the pixels in the image $\mathbf{I}$, given by $NL(\mathbf{v})(i) = \sum_{j \in \mathbf{I}} w(i,j)v(j)$, where the weights $w(i,j)$ quantify the similarity between the pixels i and j and satisfy the conditions $0 \leq w(i,j) \leq 1$ and

$\sum_j w(i,j) = 1$. To decouple the similarity term from the spatial distance term, [9] propose the weighting function to be $w(i,j) = \frac{1}{Z_i} e^{-\frac{sim(i,j)}{h^2}}$, where Z_i is a normalization term such that the weights sum to one and the parameter h controls the spatial decay of the exponential function. Defining a window around pixel i as $\mathcal{N}_i$, the similarity between pixel i and j is defined as the Gaussian weighted similarity of the windows centered at i and j respectively, given by $sim(i,j) = \sum_k e^{-\frac{(\mathcal{N}_i(k) - \mathcal{N}_j(k))^2}{\sigma^2}}$.

MRF Based Clustering. Markov Random Fields (MRF's) have gained tremendous importance since the seminal paper by Geman and Geman [10] that introduced the idea of denoising as a labeling problem and used an MRF model for denoising. We adopt an MRF based clustering technique to project the scanned pixels into a finite label space $\{L : |L| \ll 255\}$ which is the maximum pixel range for 8 bit image data. The observation field Y is fixed and is assumed to be non-interacting. The label field X is evolved with iterations minimizing the following cost function

$$E(L) = \sum_{p \in \mathcal{X}} D_p(L_p) + \sum_{q \in \mathcal{N}_p} V_{p,q}(L_p, L_q) \tag{3}$$

where $\mathcal{N}_p$ represents the neighborhood for a particular node p. The MRF model balances the two cost terms in Eq. 3 to generate the possible label for the target node. The first term constrains the label to be close to the observation. If the cluster center intensities are pre-specified, then this term can be simply evaluated as $I(p) - C_i$ over all labels $i = \{1, 2, \ldots, L\}$, where $I(p)$ is the pixel intensity at location p. For our experiments we typically set $|L| = 5$.

Lateral Cluster Projection. The NL-means filtering and clustering, leaves us with an image where the blood pool is certain to be one of darkest clusters (Fig. 2 (left panel)) in the label map. We compute the radial histogram of the pixels with bins centred at the probe location as our feature. This essentially leads to counting pixels radially for each label class and results in an 1D curve with as many points as the image depth for each bin. To identify the LV blood pool, the two darkest clusters (red and pink colors in Fig. 2 right panel) are considered. The initialization estimate is then, a radial curve at the maximal bin index corresponding to the blood pool cluster.

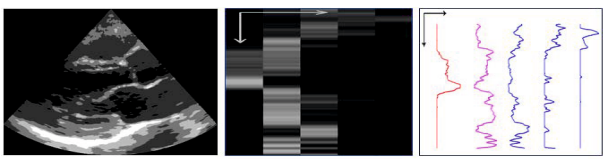

Fig. 2. Left to right: MRF based pixel clusters, label histogram, histogram for individual label classes

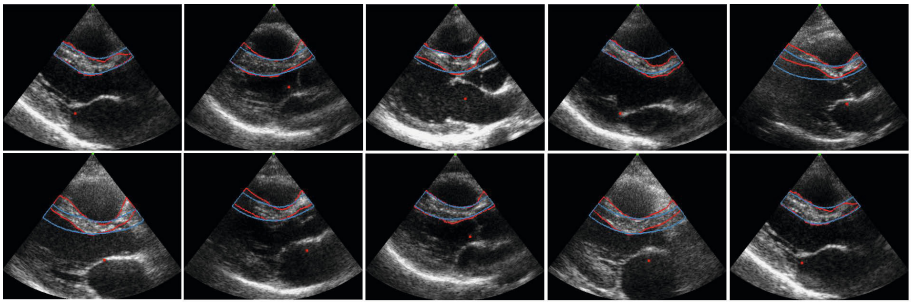

Fig. 3. Comparative Results. Red curve: proposed method. Blue curve: Subramanian et al. [5]. The red dot shows the mitral valve tip.

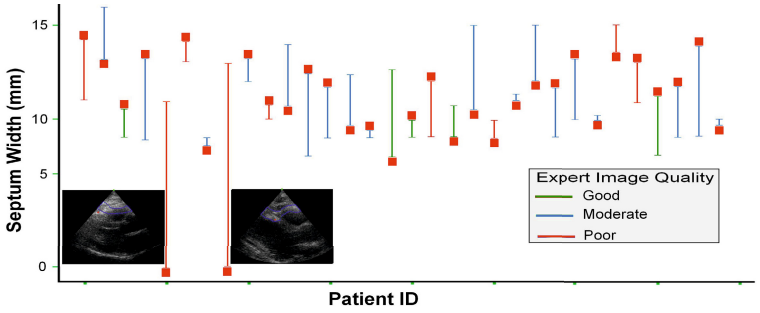

Fig. 4. Quantitative comparisons. Quantitative comparison of automated vs expert measurements on 32 cases. The height of the bar denotes the error in measurement (one end ground truth and the other end algorithm output). Red markers denote the algorithm output. Inset images correspond to the two failure cases with poor image quality.

3 Results

We evaluated our method on B-mode PLAX recordings, representing a total of 32 different patients with has ≈3 cycles/recording. The subjects with varied clinical background, normal chamber dimensions and systolic function underwent routine echocardiography (commercially available Vivid 7, GE) with ECG gating. The patient data used in our validation included normal and hypertrophic patients. We present comparative visual results against the method proposed in [5] for cardiac segmentation and tracking in Fig. 3. Our method runs in near real-time, with an average computation time of 0.1 sec/frame on a 2.6GHZ PC with 2GB RAM.

The performance of the method was quantitatively evaluated by comparing the thickness measurement, as generated by the segmentation, with the cardiologist's measurement. Fig. 4 shows the measurement error for each dataset. We observe that the method performs favorably, with the exception of two cases

rated as having poor image quality. In these cases, both the segmentation and mitral valve estimation fail (Fig. 4). Discarding these two cases, which are extremely poor quality, we report a mean error = 1.8949mm, variance = 3.609mm, maximum error bound = [-1.173mm, 6.0042mm] for our proposed method.

4 Conclusions

We have developed an automatic approach for cardiac segmentation that is robust to the image noise, haze and cardiac motion typical in ultrasound. 1D curve evolutions based on regional statistics and constrained by temporal priors are shown to be well suited for segmentation of rapidly moving cardiac structures. Additionally, the non-trivial step of robust initialization to enhance convergence of active contours is addressed. Our framework was validated on 32 B-mode PLAX recordings and compared favourably with the true boundary on an average of 93% of the cases, making it attractive for clinical application. In the future, we would like to evaluate the performance of the framework to segment and track other cardiac structures, such as, the posterior wall and LV cavity.

References

1. Lang, R.M., et al.: Recommendations for chamber quantification: a report from the American Society of Echocardiography's Guidelines and Standards Committee. J. American Soc. Echocardiography 18, 1440–1463 (2005) [1]
2. Lankton, S., Tannenbaum, A.: Localizing region-based active contours. IEEE Trans. Image Processing 17, 2029–2039 (2008) [2]
3. Chan, T.F., Vese, L.A.: Active contours without edges. IEEE Trans. Image Processing 10, 266–277 (2001) [2]
4. Leventon, M., Grimson, W.L., Faugeras, O.: Statistical shape influence in geodesic active contours. In: CVPR, vol. 1, pp. 316–323 (2000) [2]
5. Subramanian, N., Padfield, D., Thiruvenkadam, S., Narasimhamurthy, A., Frigstad, S.: Automated Interventricular Septum Thickness Measurement from B-Mode Echocardiograms. In: Jiang, T., Navab, N., Pluim, J.P.W., Viergever, M.A. (eds.) MICCAI 2010, Part I. LNCS, vol. 6361, pp. 510–517. Springer, Heidelberg (2010) [2], [3], [4], [7]
6. Niethammer, M., Tannenbaum, A.: Dynamic geodesic snakes for visual tracking. In: IEEE CVPR (2004) [2], [3]
7. Snare, S.R., Mjølstad, O.C., Orderud, F., Dalen, H., Torp, H.: Automated septum thickness measurement–A kalman filter approach. Computer Methods and Programs in Biomedicine (2011) [4]
8. van Stralen, M., et al.: Time Continuous Detection of the Left Ventricular Long Axis and the Mitral Valve Plane in 3-D Echocardiography. Ultrasound in Medicine and Biology 34, 196–207 (2008) [5]
9. Buades, A., Coll, B., Morel, J.M.: A review of image denoising methods, with a new one. In: Multiscale Modeling and Simulation, vol. 4, pp. 490–530 (2006) [5]
10. Geman, S., Geman, D.: Stochastic relaxation, gibbs distributions, and the bayesian restoration of images. IEEE Trans. Pattern Anal. Machine Intelligence 6, 721–741 (1984) [6]