Tracking Endocardial Boundary and Motion via Graph Cut Distribution Matching and Multiple Model Filtering

Kumaradevan Punithakumar[1], Ismail Ben Ayed[1], Ali Islam[2], Ian Ross[3] and Shuo Li[1,4]

[1] GE Healthcare, London, Ontario, Canada
[2] St. Joseph's Health Care, London, Ontario, Canada
[3] London Health Science Center, London, Ontario, Canada
[4] University of Western Ontario

Abstract. Tracking the left ventricular (LV) endocardial boundary and motion from cardiac magnetic resonance (MR) images is difficult because of low contrast and photometric similarities between the heart wall and papillary muscles within the LV cavity. This study investigates the problem via *Graph Cut Distribution Matching (GCDM)* and *Interacting Multiple Model (IMM) smoothing*. GCDM yields initial frame segmentations by keeping the same photometric/geometric distribution of the cavity over cardiac cycles, whereas IMM constrains the results with prior knowledge of temporal consistency. Incorporation of prior knowledge that characterizes the dynamic behavior of the LV enhances the accuracy of both motion estimation and segmentation. However, accurately characterizing the behavior using a single Markovian model is not sufficient due to substantial variability in heart motion. Moreover, dynamic behaviors of normal and abnormal hearts are very different. This study introduces multiple models, each corresponding to a different phase of the LV dynamics. The IMM, an effective estimation algorithm for Markovian switching systems, yields the state estimate of endocardial points as well as the model probability that indicates the most-likely model. The proposed method is evaluated quantitatively by comparison with independent manual segmentations over 2280 images acquired from 20 subjects, which demonstrated competitive results in comparisons with a recent method.

1 Introduction

Tracking myocardial boundary and motion plays a leading role in the diagnosis of cardiovascular diseases. It allows analyzing and quantifying myocardial motion [1]. Magnetic Resonance (MR) sequences are widely used for analyzing cardiac function, and provide a large number of images[1]. Therefore, tracking based on manual delineation of the Left Ventricular (LV) boundary in all these images is

[1] Typically, the number of images per subject is equal to 200.

H. Zha, R.-i. Taniguchi, and S. Maybank (Eds.): ACCV 2009, Part III, LNCS 5996, pp. 172–182, 2010.

prohibitively time consuming, and automating the process has been the subject of an intense research effort recently [2,3]. This problem is difficult due to the low contrast and photometric similarities between connected cardiac regions - for instance, the papillary muscles within the cavity and heart wall have approximately the same intensity. Therefore, standard segmentation methods based solely on intensity information cannot yield accurate tracking. To overcome this difficulty, most of existing methods constrain the solution with prior geometric properties, such as the shape of the LV cavity learned *a priori* from a finite training set [4]. Unfortunately, such training information is not sufficiently reliable to recover the substantial variability between subjects [1]. Furthermore, these methods do not account for *temporal consistency* of cardiac motion.

The system proposed in this study consists of two complementary steps. The first step, referred to as Graph Cut Distribution Matching (GCDM), yields initial segmentation of the LV cavity within each frame by keeping the same photometric/geometric distribution of the cavity over cardiac cycles. This is done by minimizing a distribution-matching energy which measures the similarity between a given segmentation of the first frame and the unknown segmentation of the current frame. Based on global distribution information learned from the first frame in the current data, GCDM overcomes some of the difficulties inherent to cardiac images without resorting to *a priori* training.

The second step, referred to as Interacting Multiple Model (IMM) smoothing, constrains the segmentation results with prior knowledge of temporal consistency via *multiple models*. Incorporation of such prior knowledge, which characterizes the dynamic behavior of the LV motion, enhances the accuracy of both segmentation and tracking. Particularly, a cyclic temporal model is well suited for periodic cardiac motion [3,5]. However, due to the substantial variability in the dynamics of the LV of a normal heart, accurate representation of the motion with a single Markovian model is not sufficient. Moreover, the dynamics of normal and abnormal hearts are very different. Therefore, the LV dynamics can be viewed as a *Markovian switching system*, which has both *continuous* (noise) and *discrete* (model) uncertainties. For such systems, the IMM is an effective solution. It yields the state estimate as well as the model probability indicating the most-likely model. Furthermore, in IMM filtering, the state estimates are updated using only the past observations. However, if a delay in estimation can be tolerated, the results could be drastically improved using future measurements. As such, IMM smoothing [6] can be further exploited for our problem.

The proposed method is evaluated quantitatively by comparison with independent manual segmentations over 2280 images acquired from 20 subjects, which demonstrated competitive results in comparisons with a recent method.

2 Graph Cut Distribution Matching

Consider a MR cardiac sequence containing N frames[2] $\mathbf{I}_p^n = \mathbf{I}^n(p) : \mathcal{P} \subset \mathbb{R}^2 \rightarrow \mathcal{I}$, $n \in [1..N]$, with $\mathcal{P}$ the positional array and $\mathcal{I}$ the space of photometric variables.

[2] The number of frames N is typically equal to 20 or 25.

For each frame $n \in [2..N]$, this first stage consists of dividing $\mathcal{P}$ into two regions– the *heart cavity* and the its complement in $\mathcal{P}$–according to *photometric* and *geometric* criteria. We state the problem as the minimization of a discrete cost function with respect to a binary variable (labeling), $\mathcal{L}^n(p) : \mathcal{P} \rightarrow \{0, 1\}$, which defines a variable partition of $\mathcal{P}$: the *heart cavity* $\mathbf{C}^n$ corresponding to region $\{p \in \mathcal{P}/\mathcal{L}^n(p) = 1\}$ and its complement, the *background* $\mathbf{B}^n$ corresponding to region $\{p \in \mathcal{P}/\mathcal{L}^n(p) = 0\}$. The optimal labeling is obtained by minimizing an energy containing two kernel density matching terms, an intensity matching term and a distance matching term. To introduce the energy, we first consider the following definitions for any labeling $\mathcal{L} : \mathcal{P} \rightarrow \{0, 1\}$, any image $\mathbf{I} : \mathcal{P} \subset \mathbb{R}^2 \rightarrow \mathcal{I}$, and any space of variables $\mathcal{I}$.

- $\mathbf{P}^{\mathcal{I}}_{\mathcal{L},\mathbf{I}}$ is the Kernel Density Estimate (KDE) of the distribution of image data $\mathbf{I}$ within region $\mathbf{R}_{\mathcal{L}} = \{p \in \mathcal{P}/\mathcal{L}(p) = 1\}$

$$\forall i \in \mathcal{I}, \quad \mathbf{P}^{\mathcal{I}}_{\mathcal{L},\mathbf{I}}(i) = \frac{\sum_{p \in \mathbf{R}_{\mathcal{L}}} K(i - \mathbf{I}_p)}{\mathbf{A}_{\mathcal{L}}}, \quad \text{with} \quad K(y) = \frac{1}{\sqrt{2\pi\sigma^2}} exp^{-\frac{y^2}{2\sigma^2}}, \quad (1)$$

with $\mathbf{A}_{\mathcal{L}}$ is the number of pixels within $\mathbf{R}_{\mathcal{L}}$: $\mathbf{A}_{\mathcal{L}} = \sum_{\mathbf{R}_{\mathcal{L}}} 1$, and σ is the width of the Gaussian kernel. Note that choosing K equal to the Dirac function yields the histogram.

- $\mathcal{B}(f, g)$ is the *Bhattacharyya* coefficient[3] measuring the amount of overlap (similarity) between two distributions f and g: $\mathcal{B}(f, g) = \sum_{i \in \mathcal{I}} \sqrt{f(i)g(i)}$.

We assume that a segmentation of frame $\mathbf{I}^1$, i.e., a labeling $\mathcal{L}^1$ defining a partition $\{\mathbf{C}^1, \mathbf{B}^1\}$, is given. Using prior information from this frame, the photometric and geometric model distributions of the cavity are learned, and used in the following distribution matching constraints to segment subsequent frames.

Photometric constraint. Given the learned model of intensity, which we denote $\mathbf{M}^{\mathcal{I}} = \mathbf{P}^{\mathcal{I}}_{\mathcal{L}^1,\mathbf{I}^1}$, the purpose of this term is to find for each subsequent frame $\mathbf{I}^n$ a region $\mathbf{C}^n$ whose intensity distribution most closely matches $\mathbf{M}^{\mathcal{I}}$. To this end, we minimizes the following intensity matching function with respect to $\mathcal{L}$:

$$\mathcal{B}^{\mathcal{I}}(\mathcal{L}, \mathbf{I}^n) = -\mathcal{B}(\mathbf{P}^{\mathcal{I}}_{\mathcal{L},\mathbf{I}^n}, \mathbf{M}^{\mathcal{I}}) = -\sum_{i \in \mathcal{I}} \sqrt{\mathbf{P}^{\mathcal{I}}_{\mathcal{L},\mathbf{I}^n}(i)\mathbf{M}^{\mathcal{I}}(i)} \quad (2)$$

Geometric constraint. The purpose of this term is to constrain the segmentation with prior geometric information (shape, scale, and position of the cavity) obtained from the learning frame. Let c be the centroid of cavity $\mathbf{C}^1$ in the learning frame and $\mathbf{D}(p) = \frac{\|p-c\|}{N_{\mathbf{D}}} : \mathcal{P} \rightarrow \mathcal{D}$ a *distance image* measuring at each point $p \in \mathcal{P}$ the normalized distance between p and c, with $\mathcal{D}$ the space of distance variables and $N_{\mathbf{D}}$ a normalization constant. Let $\mathbf{M}^{\mathcal{D}} = \mathbf{P}^{\mathcal{D}}_{\mathcal{L}^1,\mathbf{D}}$ the model distribution of distances within the cavity in the learning frame. We seek a region $\mathbf{C}^n$ whose distance distribution most closely matches $\mathbf{M}^{\mathcal{D}}$ by minimizing:

$$\mathcal{B}^{\mathcal{D}}(\mathcal{L}, \mathbf{D}) = -\mathcal{B}(\mathbf{P}^{\mathcal{D}}_{\mathcal{L},\mathbf{D}}, \mathbf{M}^{\mathcal{D}}) = -\sum_{d \in \mathcal{D}} \sqrt{\mathbf{P}^{\mathcal{D}}_{\mathcal{L},\mathbf{D}}(d)\mathbf{M}^{\mathcal{D}}(d)} \quad (3)$$

[3] Note that the values of $\mathcal{B}$ are always in $[0, 1]$, where 0 indicates that there is no overlap, and 1 indicates a perfect match between the distributions.

Note that this geometric prior is invariant to rotation, and embeds *implicitly* uncertainties with respect to scale via the kernel width σ in (1). The higher σ, the more scale variations allowed. In our experiments, $\sigma = 2$ was sufficient to handle effectively variations in the scale of the cavity. This geometric prior relaxes (1) complex learning/modeling of geometric characteristics and the need of a training set and (2) *explicit* optimization with respect to geometric transformations.

The discrete energy function. The discrete energy function we minimize contains the photometric and geometric matching terms as well as a regularization term for smooth segmentation boundaries. For each $n \in [2..N]$, the first stage of our algorithm computes the optimal labeling $\mathcal{L}^n_{opt}$ minimizing:

$$\mathcal{F}(\mathcal{L}, \mathbf{I}^n) = \mathcal{B}^{\mathcal{I}}(\mathcal{L}, \mathbf{I}^n) + \mathcal{B}^{\mathcal{D}}(\mathcal{L}, \mathbf{D}) + \lambda \mathbf{S}(\mathcal{L}) \tag{4}$$

where $\mathbf{S}(\mathcal{L})$ is related to the length of the partition boundary given by [7]:

$$\mathbf{S}(\mathcal{L}) = \sum_{\{p,q\} \in \mathcal{N}} \frac{1}{\|p - q\|} \delta_{\mathcal{L}_p \neq \mathcal{L}_q}, \quad \text{with} \quad \delta_{x \neq y} = \begin{cases} 1 & \text{if} \quad x \neq y \\ 0 & \text{if} \quad x = y \end{cases}, \tag{5}$$

and $\mathcal{N}$ is a neighborhood system containing all unordered pairs $\{p, q\}$ of neighboring elements of $\mathcal{P}$. λ is a positive constant that balances the relative contribution of $\mathbf{S}$.

Graph cut optimization. The distribution matching terms in $\mathcal{F}(\mathcal{L}, \mathbf{I}^n)$ do not afford an analytical form amenable to graph cut optimization. The ensuing problem is *NP-hard*. Furthermore, gradient-based optimization procedures are computationally very expensive and difficult to apply. To overcome this problem, we compute a first-order approximation of the Bhattacharyya measures in $\mathcal{F}(\mathcal{L}, \mathbf{I}^n)$ by introducing an auxiliary[4] labeling which corresponds to an arbitrary, fixed partition. For any labeling $\mathcal{L}$, the intensity matching term minus a constant reads:

$$\mathcal{B}^{\mathcal{I}}(\mathcal{L}, \mathbf{I}^n) - \underbrace{\mathcal{B}^{\mathcal{I}}(\mathcal{L}^a, \mathbf{I}^n)}_{Constant} \approx \underbrace{\sum_{p \in \mathcal{P}} \delta \mathcal{B}^{\mathcal{I}}_{p, \mathcal{L}^a, \mathcal{L}}}_{Variations\ of\ \mathcal{B}^I} = -\frac{1}{2} \sum_{p \in \mathcal{P}} \sum_{i \in \mathcal{I}} \sqrt{\frac{\mathbf{M}^{\mathcal{I}}(i)}{\mathbf{P}^{\mathcal{I}}_{\mathcal{L}^a, \mathbf{I}^n}(i)}} \delta \mathbf{P}^{\mathcal{I}}_{p, \mathcal{L}^a, \mathcal{L}}(i), \tag{6}$$

where $\delta \mathcal{B}^{\mathcal{I}}_{p, \mathcal{L}^a, \mathcal{L}}$ and $\delta \mathbf{P}^{\mathcal{I}}_{p, \mathcal{L}^a, \mathcal{L}}(i))$ are the elementary variations of, respectively, $\mathcal{B}^{\mathcal{I}}(\mathcal{L}^a, \mathbf{I}^n)$ and $\mathbf{P}^{\mathcal{I}}_{\mathcal{L}^a, \mathbf{I}^n}(i)$, each corresponding to changing the label of pixel p from $\mathcal{L}^a(p)$ to $\mathcal{L}(p)$. Elementary variation $\delta \mathcal{B}^{\mathcal{I}}_{p, \mathcal{L}^a, \mathcal{L}}$ is computed in the rightmost equality of (6) with the first-order expansion of the Bhattacharyya measure $\mathcal{B}^{\mathcal{I}}(\mathcal{L}, \mathbf{I}^n)$. Then, we compute elementary variations $\delta \mathbf{P}^{\mathcal{I}}_{p, \mathcal{L}^a, \mathcal{L}}(i)$, $i \in \mathcal{I}$ using the the kernel density estimate in (1), which yields after some manipulations:

$$\delta \mathbf{P}^{\mathcal{I}}_{p, \mathcal{L}^a, \mathcal{L}}(i) = \begin{cases} \delta_{\mathcal{L}^a(p) \neq 1} \frac{K(i - \mathbf{I}^n_p) - \mathbf{P}^{\mathcal{I}}_{\mathcal{L}^a, \mathbf{I}^n}(i)}{\mathbf{A}_{\mathcal{L}^a} + 1} & \text{if} \quad \mathcal{L}(p) = \mathbf{1} \\ \delta_{\mathcal{L}^a(p) \neq 0} \frac{\mathbf{P}^{\mathcal{I}}_{\mathcal{L}^a, \mathbf{I}^n}(i) - K(i - \mathbf{I}^n_p)}{\mathbf{A}_{\mathcal{L}^a} - 1} & \text{if} \quad \mathcal{L}(p) = \mathbf{0} \end{cases} \tag{7}$$

[4] Note that $\mathcal{L}^a$ is an arbitrary, fixed labeling which can be obtained from a given segmentation of the first frame.

where $\delta_{x \neq y}$ given by (5). Finally, using (7) in (6) and after some manipulations, the intensity matching term reads as the sum of unary penalties plus a constant:

$$\mathcal{B}^{\mathcal{I}}(\mathcal{L}, \mathbf{I}^n) \approx constant + \sum_{p \in \mathcal{P}} \mathbf{b}^{\mathcal{I}}_{p,\mathbf{I}^n}(\mathcal{L}(p)), \tag{8}$$

with $\mathbf{b}^{\mathcal{I}}_{p,\mathbf{I}}$ given, for any image $\mathbf{I} : \mathcal{P} \subset \mathbb{R}^2 \rightarrow \mathcal{I}$ and any space of variables $\mathcal{I}$, by

$$\begin{aligned} \mathbf{b}^{\mathcal{I}}_{p,\mathbf{I}}(\mathbf{1}) &= \frac{\delta_{\mathcal{L}^a(p) \neq \mathbf{1}}}{2(\mathbf{A}_{\mathcal{L}^a} + 1)} (\mathcal{B}^{\mathcal{I}}(\mathcal{L}^a, \mathbf{I}) - \sum_{i \in \mathcal{I}} K(i - \mathbf{I}_p) \sqrt{\frac{\mathbf{M}^{\mathcal{I}}(i)}{\mathbf{P}^{\mathcal{I}}_{\mathcal{L}^a,\mathbf{I}}(i)}}) \\ \mathbf{b}^{\mathcal{I}}_{p,\mathbf{I}}(\mathbf{0}) &= \frac{\delta_{\mathcal{L}^a_p \neq \mathbf{0}}}{2(\mathbf{A}_{\mathcal{L}^a} - 1)} (\sum_{i \in \mathcal{I}} K(i - \mathbf{I}_p) \sqrt{\frac{\mathbf{M}^{\mathcal{I}}(i)}{\mathbf{P}^{\mathcal{I}}_{\mathcal{L}^a,\mathbf{I}}(i),}} - \mathcal{B}^{\mathcal{I}}(\mathcal{L}^a, \mathbf{I})) \end{aligned} \tag{9}$$

Using a similar computation for the distance matching term, adopting the same notation in (9) for distance image $\mathbf{D}$, and ignoring the constants, the problem reduces to optimizing the following sum of unary and pairwise (submodular) penalties:

$$\mathcal{L}^{opt} = \arg \min_{\mathcal{L}:\mathcal{P} \rightarrow \{0,1\}} \sum_{p \in \mathcal{P}} \{\mathbf{b}^{\mathcal{I}}_{p,\mathbf{I}^n}(\mathcal{L}(p)) + \mathbf{b}^{\mathcal{D}}_{p,\mathbf{D}}(\mathcal{L}(p))\} + \lambda \mathbf{S}(\mathcal{L}) \tag{10}$$

In combinatorial optimization, a global optimum of the sum of unary and pairwise (submodular) penalties can be computed efficiently in low-order polynomial time by solving an equivalent max-flow problem [8]. In our case, it suffices to build a weighted graph $\mathcal{G} = \langle \mathbf{N}, \mathbf{E} \rangle$, where $\mathbf{N}$ is the set of nodes and $\mathbf{E}$ the set of edges connecting these nodes. $\mathbf{N}$ contains a node for each pixel $p \in \mathcal{P}$ and two additional terminal nodes, one representing the foreground region (i.e., the cavity), denoted $\mathbf{T_F}$, and the other representing the background, denoted $\mathbf{T_B}$. Let $\mathbf{w}_{p,q}$ be the weight of the edge connecting neighboring pixels $\{p, q\}$ in $\mathcal{N}$, and $\{\mathbf{w}_{p,\mathbf{T_F}}, \mathbf{w}_{p,\mathbf{T_B}}\}$ the weights of the edges connecting each pixel p to each of the terminals. By setting the edge weights as follows:

$$\mathbf{w}_{p,\mathbf{T_F}} = \mathbf{b}^{\mathcal{I}}_{p,\mathbf{I}^n}(\mathbf{0}) + \mathbf{b}^{\mathcal{D}}_{p,\mathbf{D}}(\mathbf{0}); \quad \mathbf{w}_{p,\mathbf{T_B}} = \mathbf{b}^{\mathcal{I}}_{p,\mathbf{I}^n}(\mathbf{1}) + \mathbf{b}^{\mathcal{D}}_{p,\mathbf{D}}(\mathbf{1}); \quad \mathbf{w}_{p,q} = \frac{\lambda}{\|p - q\|},$$

we compute, using the max-flow algorithm of Boykov and Kolmogorov [8], a minimum cut $\mathcal{C}^n_{opt}$ of $\mathcal{G}$, i.e., a subset of edges in $\mathbf{E}$ whose removal divides the graph into two disconnected subgraphs, each containing a terminal node, and whose sum of edge weights is minimal. This minimum cut, which assigns each node (pixel) p in $\mathcal{P}$ to one of the two terminals, induces an optimal labeling $\mathcal{L}^n_{opt}$ ($\mathcal{L}^n_{opt}(p) = 1$ if p is connected to $\mathbf{T_F}$ and $\mathcal{L}^n_{opt}(p) = 0$ if p is connected to $\mathbf{T_B}$), which minimizes globally the approximation in (10).

3 Dynamic Model for Temporal Periodicity

Let (x, y) be a Cartesian point on the boundary between the segmentation regions obtained with graph cut distribution matching $\mathbf{R}_{\mathcal{L}^n_{opt}} = \{p \in \mathcal{P} / \mathcal{L}^n_{opt}(p) = 1\}$

and $\mathbf{R}^c_{\mathcal{L}^n_{opt}} = \{p \in \mathcal{P}/P\mathcal{L}^n_{opt}(p) = 0\}$. Consider the state vector $\xi = [\bar{\mathsf{x}}\, \mathsf{x}\, \dot{\mathsf{x}}]^T$ that describes the dynamics of the point in x-coordinate direction, where $\dot{\mathsf{x}}$ and $\bar{\mathsf{x}}$ denote, respectively, velocity and the mean position over cardiac cycle. We assume the heart motion is periodic. A *continuous state-space* model that describes the cyclic motion of the point is given by,

$$\dot{\xi}(t) = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ \omega^2 & -\omega^2 & 0 \end{bmatrix} \xi(t) + \begin{bmatrix} 1 & 0 \\ 0 & 0 \\ 0 & 1 \end{bmatrix} w(t) = A(\omega)\xi(t) + Bw(t) \tag{11}$$

where ω is the angular frequency, and $w(t)$ the white noise that accounts for approximating the unpredictable modeling errors arising in LV motion. Model (11) is linear for a given ω and can be viewed as an approximation of the temporal periodic model used in [5] where the higher-order terms of the Fourier expansion were neglected. A bank of models can be effectively used in parallel to closely match the changing dynamics of boundary points as discussed in Section 4. The *discrete-time equivalent* of (11) can be derived as

$$\xi_{k+1} = \begin{bmatrix} 1 & 0 & 0 \\ 1 - \cos(\omega T) & \cos(\omega T) & \frac{1}{\omega}\sin(\omega T) \\ \omega \sin(\omega T) & -\omega \sin(\omega T) & \cos(\omega T) \end{bmatrix} \xi_k + w_k = F(\omega)\xi_k + w_k \tag{12}$$

where w_k is the process noise of the discrete-model. We can consider the state vector $x = [\bar{\mathsf{x}}\, \mathsf{x}\, \dot{\mathsf{x}}\, \bar{\mathsf{y}}\, \mathsf{y}\, \dot{\mathsf{y}}]^T$ that describes the dynamics in x-y plane. The discrete state-space model in x-y plane is given by

$$\mathbf{x}_{k+1} = \begin{bmatrix} F(\omega) & \mathbf{0}_{3\times 3} \\ \mathbf{0}_{3\times 3} & F(\omega) \end{bmatrix} \mathbf{x}_k + v_k = F_k \mathbf{x}_k + v_k \tag{13}$$

The single Markovian model in (13) is insufficient to describe the LV dynamics due to the following reasons: (1) The angular frequency that characterizes the motion of a LV point for normal subjects changes over time. (2) The dynamics of LV motion differ significantly in systolic and diastolic phases of heart beat. (3) The LV dynamics of abnormal subjects differ significantly from those of normal subjects. Therefore, the LV dynamics is a *hybrid system* – a system which has both *continuous* (noise) and *discrete* (model) uncertainties – and, as such, it requires an interacting multiple model (IMM) approach. In the context of tracking maneuvering targets [9]. IMM estimation is shown to be very effective in the cases of hybrid systems. In the next section, we devise IMM to track the motion of the LV.

4 Interacting Multiple Model Algorithm

Let the system consists of n discrete set of models denoted by $M = \{M^1, \ldots, M^n\}$. Let $\mu_0^j = P\{M_0^j\}$ be the prior probability of model M^j, and $p_{ij} = P\{M_k^j | M_{k-1}^i\}$ be the probability of switching model from i to model j, with M_k^j being the

model M^j at time step k. The system equations corresponding to M_k^j is given by: $x_k = F_k^j x_{k-1} + w_{k-1}^j$ and $z_k = H_k^j x_k + v_k^j$. The one cycle recursion of the IMM filter can be summarized as follows.

Interaction. The mixing probabilities $\mu_k^{i|j}$ for each model M^i and M^j are calculated as follows. $\bar{c}_j = \sum_{i=1}^n p_{ij}\mu_{k-1}^i$ and $\mu_k^{i|j} = \frac{1}{\bar{c}_j}\sum_{i=1}^n p_{ij}\mu_{k-1}^i$ where μ_{k-1}^i is the model probability. The inputs to each filter are calculated by

$$m_{k-1}^{0j} = \sum_{i=1}^n \mu_k^{i|j} m_{k-1}^i \; and \; P_{k-1}^{0j} = \sum_{i=1}^n \mu_k^{i|j}[P_{k-1}^i + (m_{k-1}^i - m_{k-1}^{0j})(m_{k-1}^i - m_{k-1}^{0j})^T].$$

Filtering. Kalman filter [9] is used for mode-conditioned state estimates:

$$[m_k^{-,i}, P_k^{-,i}] = \mathrm{KF}_p(m_{k-1}^{0j}, P_{k-1}^{0j}, F(\omega^i), Q_k^i) \; and \; [m_k^i, P_k^i] = \mathrm{KF}_u(m_k^{-,i}, P_k^{-,i}, z_k, H_k^i, R_k^i)$$

where KF_p and KF_u denote prediction and update equations of Kalman filter, respectively. The probability of model M_k^j being correct (mode probability) is computed as a function of the likelihoods of the other filters: $\mu_k^i = \frac{\Lambda_k^i \bar{c}_i}{\sum_{i=1}^n \Lambda_k^i \bar{c}_i}$ where likelihood of model M^i is given by $\Lambda_k^i = \mathcal{N}(v_k^i; 0, S_k^i)$ where v_k^i is the measurement residual and S_k^i innovation covariance for model M^i in the Kalman filter update step.

Mixing. The estimate of the IMM algorithm is calculated by combining individual mode-conditioned filter estimates using mode probabilities as follows: $m_k = \sum_{i=1}^n \mu_k^i m_k^i$ and $P_k = \sum_{i=1}^n \mu_k^i \left[P_k^i + (m_k^i - m_k)(m_k^i - m_k)^T\right]$.

4.1 Fixed-Interval IMM-Smoother

If an estimation delay can be tolerated, the performance of the filtering algorithm can be improved drastically by smoothing. In cardiac images, the delay in estimation is not significant as the imaging frequency is relatively high. There are several variations of smoothing available [10]. Here, we use a *fixed-interval smoothing*, which is the most common type. The optimal solution for fixed-interval smoothing is to fuse the posterior distributions obtained by two optimal IMM estimators, one running forward and the other backward using an equivalent reverse-time Markov model. However, obtaining the equivalent reverse-time model and the optimal forward/backward IMM estimators are difficult. The approximate fixed-interval smoother [6], which uses simpler fusion technique and an approximation of the required backward IMM algorithm directly from original Markov switching system with white Gaussian noise, is used to resolve the problem.

5 Experiments

The proposed method was applied to 120 short-axis sequences of cardiac cine MR images, with a temporal resolution of 20 frames/cardiac cycle, acquired from 20 subject: the endocardial boundary was tracked in a total of 2280 images including apical, mid-cavity and basal slices, and the results were evaluated

quantitatively by comparisons with the manual segmentations performed independently by a medical professional. The results were also compared with the recent LV boundary tracking method in [2], using the same data.

Parameter settings. The regularization and kernel width parameters were unchanged for all the datasets in GCDM: α set equal to 0.15, the kernel width σ to 2 for distance distributions, and to 10 for intensity distributions. Four dynamic models were used in the IMM (the values were measured in squared pixels and $w_0 = 2\pi/(\text{heart period})$): (1) $\omega = \omega_0/2$, $q_1 = 0.02$, $q_2 = 0.1$, $R_k = 0.5$ (2) $\omega = \omega_0/2$, $q_1 = 0.2$, $q_2 = 1$, $R_k = 8$ (3) $\omega = 2\omega_0$, $q_1 = 0.02$, $q_2 = 0.1$, $R_k = 0.5$ (4) $\omega = 2\omega_0$, $q_1 = 0.2$, $q_2 = 1$, $R_k = 8$. The filters were initialized by *two-point differencing* [9].

Quantitative performance evaluation. We used two criteria to evaluate the performances of the algorithms.

- **Root mean squared error:** The *Root Mean Squared Error* ($RMSE$) is computed using *symmetric nearest neighbor correspondences* between manual and automatic LV boundaries using 24 equally-spaced points along the boundary. The $RMSE$ over N number of points is given by: $RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(\hat{\mathrm{x}}_i - \tilde{\mathrm{x}}_i)^2 + (\hat{\mathrm{y}}_i - \tilde{\mathrm{y}}_i)^2}$, where $(\hat{\mathrm{x}}_i,\ \hat{\mathrm{y}}_i)$ is a point on the automatic boundary and $(\tilde{\mathrm{x}}_i,\ \tilde{\mathrm{y}}_i)$ is the corresponding point on the manual boundary. Table 1 reports the $RMSE$ for the proposed method and [2] averaged over all the dataset. The proposed method yielded an $RMSE$ of 2.4 pixels, whereas the method in [2] yielded 3.1. The average $RMSE$ plotted against the time step is shown in Fig. 2(a). The proposed algorithm yielded a lower $RMSE$ compared to [2] and, therefore, a higher conformity to the manual segmentation.
- **Dice metric:** We computed the *Dice Metric* (DM), a common measure of similarity between manual and automatic segmentation [2]. The DM is given by: $DM = \frac{2\mathbf{V_{am}}}{\mathbf{V_a}+\mathbf{V_m}}$, where $\mathbf{V_a}$, $\mathbf{V_m}$ and $\mathbf{V_{am}}$ are the volumes of, respectively, the automatically segmented cavity, the corresponding hand-labeled cavity, and the intersection between them. Note that DM is always between 0 and 1, where 1 means a perfect match. The proposed method yielded a DM equal to 0.915 ± 0.002, whereas the method in [2] yielded 0.884 ± 0.008, for all the data analyzed (refer to Table 1 where DM is expressed as mean $\pm$ standard deviation). We also evaluated the algorithm using the *reliability function* of the obtained Dice metrics, defined for each $d \in [0, 1]$ as the probability of obtaining DM higher than d over all volumes: $\mathcal{R}(d) = Pr(DM > d) =$(number of volumes segmented with DM higher than d)/(total number of volumes). In

Table 1. The $RMSE$ and DM statistics for the proposed method (GCDM-IMM) and method in [2]

Performance measure	**RMSE (pixels)**	**DM**
GCDM-IMM	2.4	0.915 ± 0.002
Method in [2]	3.1	0.884 ± 0.008

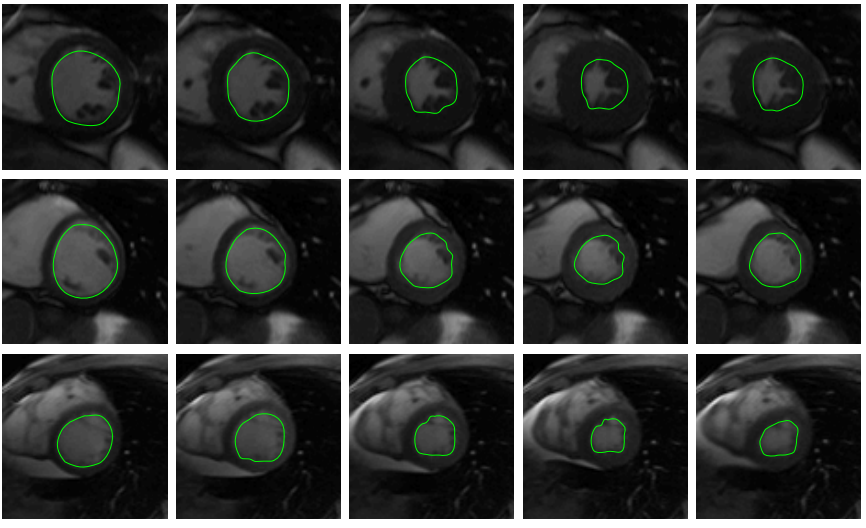
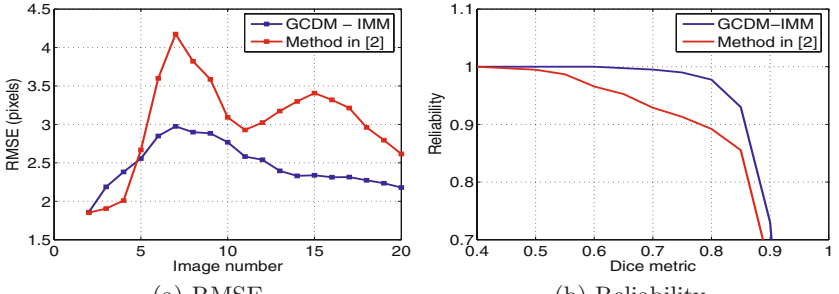


Fig. 1. Representative examples of the LV boundary tracking using the proposed method: mid-cavity (1st row), basal (2nd row) and apical (3rd row) frames. The first row depicts typical examples where the proposed method included accurately the papillary muscles inside the target cavity, although these have an intensity profile similar to the surrounding myocardium.

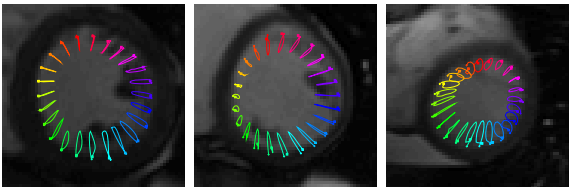


(a) RMSE (b) Reliability

Fig. 2. Comparison between automatic and manual segmentations of 2280 images, for both the proposed method (GCDM-IMM) and the method in [2]

comparison to method [2], the proposed algorithm led to a higher reliability curve, as depicted in Fig. 2(b).

Visual inspection. In figures 3 and 1, we give a representative sample of the results for 3 subjects. Fig. 3 shows the trajectory of LV points estimated using the proposed GCDM-IMM method. The first row in Fig. 1 depicts typical examples where the proposed method included accurately the papillary muscles inside the target cavity, although these have an intensity profile similar to the surrounding myocardium.



(a) mid-cavity (b) basal (c) apical

Fig. 3. Trajectory of LV endocardial boundary points estimated using the proposed method

6 Conclusions

This study investigates the problem of tracking endocardial boundary and motion via Graph Cut Distribution Matching (GCDM) and Interacting Multiple Model (IMM) smoothing. GCDM yields initial frame segmentations by keeping the same photometric/geometric distribution of the cavity over cardiac cycles, whereas IMM constrains the results with prior knowledge of temporal consistency. The proposed method is evaluated quantitatively using root mean squared error and Dice metric, by comparison with independent manual segmentations over 2280 images acquired from 20 subjects, which demonstrated significantly better results as compared to a recent method [2].

References

1. Jolly, M.P.: Automatic recovery of the left ventricular blood pool in cardiac cine MR images. In: Metaxas, D., Axel, L., Fichtinger, G., Székely, G. (eds.) MICCAI 2008, Part I. LNCS, vol. 5241, pp. 110–118. Springer, Heidelberg (2008)
2. Ben Ayed, I., Lu, Y., Li, S., Ross, I.: Left ventricle tracking using overlap priors. In: Metaxas, D., Axel, L., Fichtinger, G., Székely, G. (eds.) MICCAI 2008, Part I. LNCS, vol. 5241, pp. 1025–1033. Springer, Heidelberg (2008)
3. Spottiswoode, B., Zhong, X., Hess, A., Kramer, C., Meintjes, E., Mayosi, B., Epstein, F.: Tracking myocardial motion from cine DENSE images using spatiotemporal phase unwrapping and temporal fitting. IEEE Transactions on Medical Imaging 26(1), 15–30 (2007)
4. Andreopoulos, A., Tsotsos, J.K.: Efficient and generalizable statistical models of shape and appearance for analysis of cardiac MRI. Medical Image Analysis 12(3), 335–357 (2008)
5. McEachen, J., Nehorai, A., Duncan, J.: Multiframe temporal estimation of cardiac nonrigid motion. IEEE Transactions on Image Processing 9(4), 651–665 (2000)
6. Helmick, R., Blair, W., Hoffman, S.: Fixed-interval smoothing for Markovian switching systems. IEEE Transactions on Information Theory 41(6), 1845–1855 (1995)
7. Boykov, Y., Kolmogorov, V.: Computing geodesics and minimal surfaces via graph cuts. In: Proceedings of the Ninth IEEE International Conference on Computer Vision, vol. 1, pp. 26–33 (2003)

8. Boykov, Y., Kolmogorov, V.: An experimental comparison of min-cut/max- flow algorithms for energy minimization in vision. IEEE Transactions on Pattern Analysis and Machine Intelligence 26(9), 1124–1137 (2004)
9. Bar-Shalom, Y., Kirubarajan, T., Li, X.R.: Estimation with Applications to Tracking and Navigation. John Wiley & Sons, Inc., New York (2002)
10. Rong Li, X., Jilkov, V.: Survey of maneuvering target tracking. Part V: Multiple-model methods. IEEE Transactions on Aerospace and Electronic Systems 41(4), 1255–1321 (2005)