

Real-Time Multi-model Tracking of Myocardium in Echocardiography Using Robust Information Fusion

Bogdan Georgescu¹, Xiang Sean Zhou¹, Dorin Comaniciu¹, and Bharat Rao²

¹ Real-Time Vision and Modeling Department, Siemens Corporate Research
755 College Road East, Princeton, NJ 08540, USA
bogdan.georgescu, xiang.zhou, dorin.comaniciu@scr.siemens.com

² Siemens Medical Solutions
51 Valley Stream Parkway, Malvern, PA 19355, USA
bharat.rao@siemens.com

Abstract. Automatic myocardial wall motion tracking in ultrasound images is an important step in analysis of the heart function. Existing methods for Myocardial Wall Tracking are not robust to artifacts induced by signal dropout, significant appearance or gain control changes. We present a unified framework for tracking the myocardium wall motion in real time with uncertainty handling and robust information fusion. Our method is robust in two aspects, firstly robust information fusion is used for combining matching results from multiple appearance models and secondly fusion is performed in the shape space to combine information from measurement and prior knowledge and models. Our approach fully exploits uncertainties from the measurement, shape priors, motion dynamics, and matching process based on multiple appearance models. Experiments illustrate the advantages of our approach validating the theory and showing the potential of very accurate wall motion measurements.

1 Introduction

Accurate analysis of the myocardial wall motion of the left ventricle is crucial for the evaluation of the heart function. This task is difficult due to the fast motion of the heart muscle and respiratory interferences. It is even worse when ultrasound image sequences are used since ultrasound is the noisiest among common medical image modalities such as MRI or CT. Figure 1 illustrates the difficulties of the tracking task due to signal drop-out, poor signal to noise ratio or significant appearance changes.

Several methods have been proposed for myocardial wall tracking. Model-based deformable templates [1,2], Markov random fields [3], optical flow methods [4,5,6], or combinations of above, have been applied for tracking left ventricle (LV) from 2-D image sequences. Jacob et al. provided a brief recent review in [7]. Other related work focuses on the tracking, segmentation, or registration in 3D, 2D+T (spatial + time) or 4-D space [8], [9], [10].

One of the main problems of visual tracking is to maintain a representation of target appearance that is robust enough to cope with inherent changes due to target movement and/or imaging device movement. Methods based on template matching have to adapt the model template in order to successfully track the target. Without adaptation, tracking

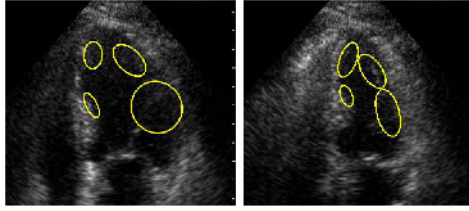


Fig. 1. Echocardiography images with area of acoustic drop-out, low signal to noise ratio and significant appearance changes. Local wall motion estimation has covariances (depicted by the solid ellipses) that reflect heteroscedastic noise.

is reliable only over short periods of time when the appearance does not change significantly. However, in most applications, for long time periods the target appearance undergoes considerable changes in structure. When the model is adapted to the previous frame accumulated motion error and rapid visual changes make the model to drift away from the target. Tracking performance can be improved by maintaining a statistical representation of the model. Using only a normal distribution, where the mean represents the most likely template, will however, not capture the full range of the appearance variability.

It is a common practice to impose model constraints in a shape tracking framework. In most cases, a subspace model is suitable for shape tracking, since the number of modes capturing the major shape variations is limited and usually much smaller than the original number of feature components used to describe the shape. A straightforward treatment is to project tracked shapes into a PCA subspace [1]. However, this approach cannot take advantage of the measurement uncertainty and is therefore not complete: In most real-world scenarios, measurement noise is heteroscedastic in nature (i.e., both anisotropic and inhomogeneous). Alternatively, one could directly incorporate a PCA shape space constraint into a Kalman filter-based tracker. In [11,12] it is suggested to set the system noise covariance matrix to be the covariance of a PCA shape model. However it does not provide a systematic and complete fusion of the model information because, for example, the model mean is discarded and it mixes the uncertainty from system dynamics with the uncertainty from the statistical shape constraint.

In this paper we introduce a unified framework for *fusing motion estimates from multiple appearance models* and *fusing a subspace shape model with the system dynamics and measurements with heteroscedastic noise*. The appearance variability is modeled by maintaining several models over time. This amounts for a nonparametric representation of the probability density function that characterizes the object appearance. Tracking is performed by obtaining independently from each model a motion estimate and its uncertainty through optical flow. A recently proposed robust fusion technique [13] is used to compute the final estimate for each component. The method, named Variable-Bandwidth Density-based Fusion (VBDF), manages the multiple data sources and outliers in the motion estimates. To obtain the final shape estimate we specifically address the issue of heteroscedastic measurement noise and its influence during the fusion with other information sources. When measurement noise is anisotropic and inhomogeneous, joint fusion of all information sources becomes critical for achieving superior performance.

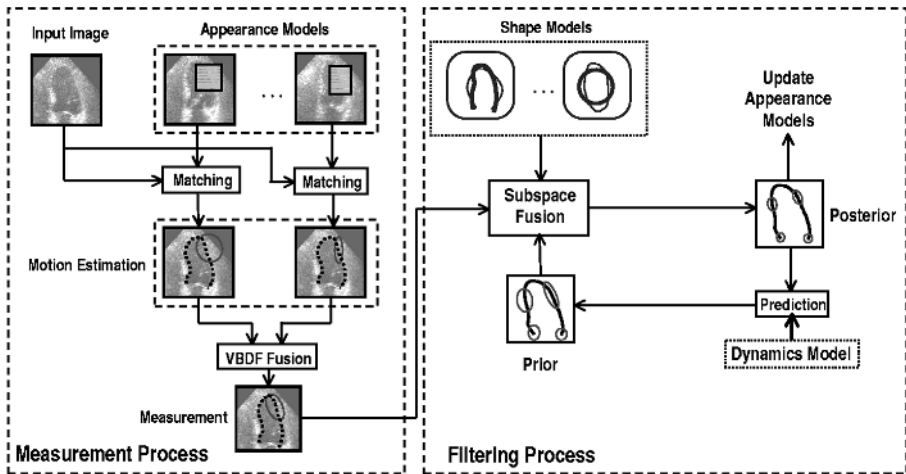


Fig. 2. The block diagram of the robust tracker with the measurement and filtering processes.

In this paper we demonstrate the advantages of the proposed framework for ultrasound heart sequences.

2 Multi-model Tracker with Robust Information Fusion

The diagram of the proposed robust tracking is illustrated in Figure 2. Our approach is robust in two aspects: in the *measurement process*, VBDF fusion is used for combining matching results from multiple appearance models and in the *filtering process*, fusion is performed in the shape space to combine information from measurement, prior knowledge and models while taking advantage of the heteroscedastic nature of the noise.

To model the changes during tracking we propose to maintain several exemplars of the object appearance over time which is equivalent to a nonparametric representation of the appearance distribution. Figure 2 illustrates the *appearance models*, i.e. the current exemplars in the model set, each having associated a set of overlapping components. Throughout this paper, we represent shapes by control or landmark points (components). These points are fitted by splines before shown to the user. A component-based approach is more robust than a global representation, being less sensitive to structural changes thus being able to deal with nonrigid shape deformations.

Each component is processed independently, its location and covariance matrix is estimated in the current image with respect to all of the model templates. For example, one of the components is illustrated by the rectangle in Figure 2 and its location and uncertainty with respect to each model is shown in the motion estimation stage. The VBDF robust fusion procedure is applied to determine the most dominant motion (mode) with the associated uncertainty.

The location of the components in the current frame is further adapted by imposing subspace shape constraints using pre-trained *shape models*. Robust shape tracking is achieved by optimally resolving uncertainties from the system dynamics, heteroscedastic

measurements noise and subspace shape model. By using the estimated confidence in each component location reliable components contribute more to the global shape motion estimation. The current frame is added to the model set if the residual error to the reference appearances is relatively low.

3 Measurement Process

Consider that we have n models M_0, M_1, \dots, M_n . For each image we maintain c components with their location denoted by $\mathbf{x}_{ij}, i = 1 \dots n, j = 1 \dots c$. When a new image is available we estimate the location and the uncertainty for each component and for each model. We adopt for this step the robust optical flow technique proposed in [13] which is also an application of the VBDF technique. The result is the motion estimate $\hat{\mathbf{x}}_{ij}$ for each component and its uncertainty \hat{C}_{ij} . Thus $\hat{\mathbf{x}}_{ij}$ represents the location estimate of component j with respect to model i . The scale of the covariance matrix is also estimated from the matching residual errors.

The VBDF estimator is based on nonparametric density estimation with adaptive kernel bandwidths [13]. The choice of the VBDF estimator is motivated by its good performance in the presence of outliers in the input data when compared to previous methods such as Covariance Intersection or BLUE estimation assuming single source, statistically independent data. The VBDF estimator is defined as the *location of the most significant mode* of a density function. The mode computation is based on the variable-bandwidth mean shift technique in a multiscale optimization framework.

Let $\hat{\mathbf{x}}_i \in \mathbb{R}^d, i = 1 \dots n$ be the available d -dimensional estimates, each having an associated uncertainty given by the covariance matrix \hat{C}_i (we drop the component index j for now). A bandwidth matrix $\hat{H}_i = \hat{C}_i + \alpha^2 \mathbf{I}$ is associated with each point $\hat{\mathbf{x}}_i$, where \mathbf{I} is the identity matrix and the parameter α determines the scale of the analysis. The sample point density estimator at location $\hat{\mathbf{x}}$ is defined by

$$\hat{f}(\mathbf{x}) = \frac{1}{n(2\pi)^{d/2}} \sum_{i=1}^n \exp\left(-\frac{1}{2}(\mathbf{x} - \hat{\mathbf{x}}_i)^\top \hat{H}_i^{-1}(\mathbf{x} - \hat{\mathbf{x}}_i)\right). \tag{1}$$

The variable bandwidth mean shift vector at location \mathbf{x} is given by

$$\mathbf{m}(\mathbf{x}) = H_h(\mathbf{x}) \sum_{i=1}^n \omega_i(\mathbf{x}) \hat{H}_i^{-1} \hat{\mathbf{x}}_i - \mathbf{x} \quad \text{where} \quad H_h(\mathbf{x}) = \left(\sum_{i=1}^n \omega_i(\mathbf{x}) \hat{H}_i^{-1} \right)^{-1}. \tag{2}$$

H_h represents the harmonic mean of the bandwidth matrices weighted by the data-dependent weights $\omega_i(\mathbf{x})$ computed at the current location \mathbf{x}

$$\omega_i(\mathbf{x}) = \frac{\frac{1}{|\hat{H}_i|^{1/2}} \exp\left(-\frac{1}{2}(\mathbf{x} - \hat{\mathbf{x}}_i)^\top \hat{H}_i^{-1}(\mathbf{x} - \hat{\mathbf{x}}_i)\right)}{\sum_{i=1}^n \frac{1}{|\hat{H}_i|^{1/2}} \exp\left(-\frac{1}{2}(\mathbf{x} - \hat{\mathbf{x}}_i)^\top \hat{H}_i^{-1}(\mathbf{x} - \hat{\mathbf{x}}_i)\right)}. \tag{3}$$

Updating iteratively the current location using the mean shift vector yields a hill-climbing procedure which converges to a stationary point of the underlying density. The

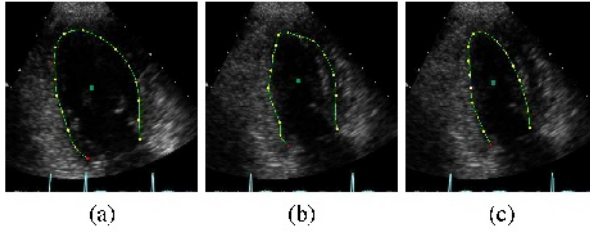


Fig. 3. Multiple models versus single model. (a) initial contour; (b) 17th contour tracked using a single appearance model (c) 17th contour tracked using multiple appearance models.

VBDF estimator finds the most important mode by iteratively applying the mean shift procedure at several scales. It starts from a large scale by choosing the parameter α large with respect to the spread of the points. In this case the density surface is unimodal and the determined mode will correspond to the globally densest region. The procedure is repeated while reducing the value of the parameter α and starting the the mean shift iterations from the mode determined at the previous scale. In the final step the bandwidth matrix associated to each point is equal to the covariance matrix, ($\hat{H}_i = \hat{C}_i$).

The VBDF estimator is a powerful tool for information fusion with the ability to deal with multiple source models. This is important for motion estimation as points in a local neighborhood may exhibit multiple motions. The most significant mode corresponds to the most relevant motion. The VBDF robust fusion technique is applied to determine the most relevant location $\hat{\mathbf{x}}_j$ for component j in the current frame. The mode tracking across scales results in

$$\hat{\mathbf{x}}_j = C(\hat{\mathbf{x}}_j) \sum_{i=1}^n \omega_i(\hat{\mathbf{x}}_j) \hat{C}_{ij}^{-1} \hat{\mathbf{x}}_{ij} \quad \text{and} \quad C(\hat{\mathbf{x}}_j) = \left(\sum_{i=1}^n \omega_i(\hat{\mathbf{x}}_j) \hat{C}_{ij}^{-1} \right)^{-1}. \quad (4)$$

Figure 3 shows the advantage of using multiple appearance models. The initial frame with the associated contour is shown in Figure 3a. Using a single model yields an incorrect tracking results (Figure 3b) and the multiple model approach correctly copes with the appearance changes (Figure 3c).

4 Filtering Process

The analysis is based on vectors formed by concatenating the coordinates of all control points [7,1]. A typical tracking framework fuses information from the prediction defined by a dynamic process and from noisy measurements. For shape tracking additional *global* constraints are necessary to stabilize the overall shape in a feasible range.

Let us now turn our attention to the problem of information fusion with one of the sources in a subspace. Given two noisy measurements of the same n -dimensional variable \mathbf{x} , each characterized by a multidimensional Gaussian distribution, $\mathcal{N}(\mathbf{x}_1, C_1)$ and $\mathcal{N}(\mathbf{x}_2, C_2)$, the maximum likelihood estimate of \mathbf{x} is the point with the minimal sum of Mahalanobis distances to the two centroids. Now, assume that one of the Gaussians

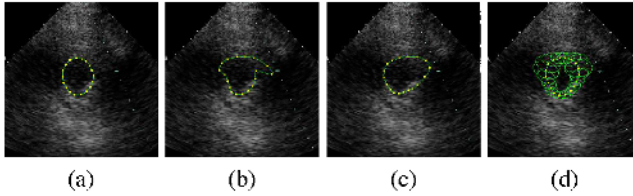


Fig. 4. Orthogonal projection versus our proposed fusion approach. (a) expert-drawn contour; (b) un-constrained flow results; (c) constrained flow using orthogonal projection; (d) contour obtained by our fusion framework with uncertainty ellipses.

is in a subspace of dimension p , e.g., C_2 is singular. With the singular value decomposition of $C_2 = UAU^T$, where $U = [\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_n]$, with \mathbf{u}_i 's orthonormal and $A = \text{diag}\{\lambda_1, \lambda_2, \dots, \lambda_p, 0, \dots, 0\}$. The distance to be minimized becomes:

$$d^2 = (\mathbf{U}_p \mathbf{y} - \mathbf{x}_1)^T C_1^{-1} (\mathbf{U}_p \mathbf{y} - \mathbf{x}_1) + (\mathbf{U}_p \mathbf{y} - \mathbf{x}_2)^T C_2^+ (\mathbf{U}_p \mathbf{y} - \mathbf{x}_2) \quad (5)$$

where $\mathbf{U}_p = [\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_p]$ represents the subspace basis and \mathbf{y} the value in this subspace. Taking derivative with respect to \mathbf{y} yields the fusion estimator for the subspace:

$$\mathbf{y}^* = C_{\mathbf{y}^*} U_p^T (C_1^{-1} \mathbf{x}_1 + C_2^+ \mathbf{x}_2) \quad \text{where} \quad C_{\mathbf{y}^*} = [U_p^T (C_1^{-1} + C_2^+) U_p]^{-1}. \quad (6)$$

Equivalent expressions can be obtained in the original space:

$$\mathbf{x}^* = U_p \mathbf{y}^* = C_{\mathbf{x}^*} (C_1^{-1} \mathbf{x}_1 + C_2^+ \mathbf{x}_2) \quad \text{where} \quad C_{\mathbf{x}^*} = U_p C_{\mathbf{y}^*} U_p^T. \quad (7)$$

It can be shown that $C_{\mathbf{x}^*}$ and $C_{\mathbf{y}^*}$ are the covariance matrices for \mathbf{x}^* and \mathbf{y}^* (see [14]).

To complete the shape tracking method, the subspace fusion with shape models is integrated into a Kalman filtering framework [15]. Kalman filtering with subspace constraints provides a unified fusion of the system dynamics, a subspace model, and measurement noise information. For details please see [14].

For endocardium tracking what we theoretically need is *the statistical shape model of the current heart* instead of a generic heart. Therefore we apply a strongly-adapted-PCA (SA-PCA) model by assuming that the PCA model and the initialized contour *jointly* represent the variations of the current case [14]. With SA-PCA, our framework now incorporates four information sources: the system dynamic, measurement, subspace model, and the initial contour.

An example is shown in Figure 4 for comparison between our approach and orthogonal projection. The fusion will not correct the error completely, but note that this correction step is accumulative so that the overall effect at a later frame in a long sequence can be very significant.

5 Experiments

In this section we will apply and evaluate the new framework to track heart contours using very noisy echocardiography data. The tracker was implemented in C++ and is running

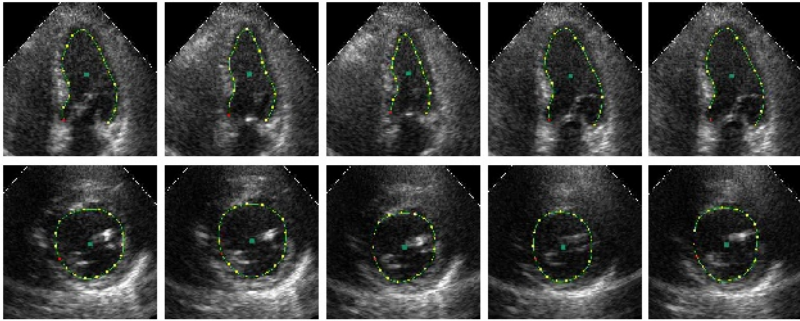


Fig. 5. Two tracking examples in rows, with 5 snapshots per sequence.

at about 20 frames per second on a single 2GHz Pentium 4 PC. Our data were selected by a cardiologist to represent normals as well as various types of cardiomyopathies, with sequences varying in length from 18 frames to 90 frames. Both training and test data were traced by experts, and confirmed by one cardiologist. We used both apical two- or four-chamber views (open contour with 17 control points) and parasternal short axis views (closed contour with 18 control points) for training and testing. PCA is performed and the original dimensionality of 34 and 36 is reduced to 7 and 8, respectively. For the appearance models we maintain 20 templates to capture the appearance variability. For systematic evaluation, we use a set of 32 echocardiogram sequences outside of the training set for testing, with 18 parasternal short-axis views and 14 apical two- or four-chamber views, all with expert-annotated ground-truth contours.

Figure 5 shows snapshots from two tracked sequences. Notice that the endocardium is not always on the strongest edge. Sometimes it manifests itself only by a faint line; sometimes it is completely invisible or buried in heavy noise; sometimes it will cut through the root of the papillary muscles where no edge is present. To compare performance of different methods, we used the Mean Sum of Squared Distance (MSSD) (*cf.* [16]) and a Mean Absolute Distance (MAD) (*cf.* [17]). Our proposed method (“Proposed”) is compared with a tracking algorithm without shape constraint (“Flow”) or with the same tracker with orthogonal PCA shape space constraints (“FlowShapeSpace”). Figure 6 and Table 1 show the comparison using the two distance measures. Our proposed method (“Proposed”) significantly outperforms others, with lower average distances and lower standard deviations for such distances.

It should be noted that our results are not indicative for *border localization* accuracies, but rather for *motion tracking* performances given an initial contour. We have set our goal to track control points on the endocardium, with anisotropic confidence estimated at each point at any given time step by using multiple appearance models, and exploit this information when consulting a prior shape model as a constraint. Our framework is general and can be applied to other modalities. Future potential applications include tracking in MR, perfusion, and extensions to tracking in 3-D or 4-D.

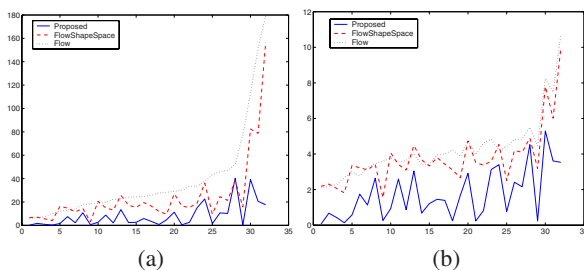


Fig. 6. Mean distances ((a) $MSSD_i$, (b) MAD_i) between tracked points and the ground truth.

Table 1. Error analysis (“Most Difficult Cases” are the last three cases in Figure 6a).

Methods	All Cases				Most Difficult Cases			
	$MSSD$	$\bar{\sigma}_{MSSD}$	MAD	$\bar{\sigma}_{MAD}$	$MSSD$	$\bar{\sigma}_{MSSD}$	MAD	$\bar{\sigma}_{MAD}$
Flow	38.1	82.9	4.3	3.6	147.9	325.0	8.8	8.2
FlowShapeSpace	24.7	35.5	3.8	2.4	106.0	181.2	7.9	6.3
Proposed	8.3	14.3	1.7	1.6	25.8	34.8	4.1	2.8

Acknowledgment. We would like to thank Dr. Alan Katz from St. Francis Hospital for fruitful interactions and guidance. We are grateful for the inspiring discussions and generous support from Alok Gupta and Sriram Krishnan of Siemens Medical Solutions.

References

1. Cootes, T., Taylor, C.: Statistical models of appearance for medical image analysis and computer vision. In: Proc. SPIE Medical Imaging. (2001) 236–248
2. Chalana, V., Linker, D.T., Haynor, D.R., Kim, Y.: A multiple active contour model for cardiac boundary detection on echocardiographic sequences. *IEEE Trans. Medical Imaging* **15** (1996) 290–298
3. Mignotte, M., Meunier, J., Tardif, J.C.: Endocardial boundary estimation and tracking in echocardiographic images using deformable templates and markov random fields. *Pattern Analysis and Applications* **4** (2001) 256–271
4. Mailloux, G.E., Langlois, F., Simard, P.Y., Bertrand, M.: Restoration of the velocity field of the heart from two-dimensional echocardiograms. *IEEE Trans. Medical Imaging* **8** (1989) 143–153
5. Adam, D., Hareuveni, O., Sideman, S.: Semiautomated border tracking of cine echocardiographic ventricular images. *IEEE Trans. Medical Imaging* **6** (1987) 266–271
6. Baraldi, P., Sarti, A., Lamberti, C., Prandini, A., Sgal-lari, F.: Semiautomated border tracking of cine echocardiographic ventricular images. *IEEE Trans. Biomedical Eng.* (1986) 259–272
7. Jacob, G., Noble, J., Behrenbruch, C., Kelion, A., Banning, A.: A shape-space-based approach to tracking myocardial borders and quantifying regional left-ventricular function applied in echocardiography. *IEEE Trans. Medical Imaging* **21** (2002) 226–238
8. Roche, A., Pennec, X., Malandain, G., Ayache, N.: Rigid registration of 3d ultrasound with mr images: a new approach combining intensity and gradient information. *IEEE Trans. Medical Imaging* **20** (2001) 1038–1049

9. Montillo, A., Metaxas, D., Axel, L.: Automated segmentation of the left and right ventricles in 4d cardiac spamm images. In: Proc. of Medical. Image Computing and Computer Assisted Intervention (MICCAI), Tokyo, Japan. (2002) 620–633
10. Hellier, P., Barillot, C.: Coupling dense and landmark-based approaches for non rigid registration. *IEEE Trans. Medical Imaging* **22** (2003) 217–227
11. Jacob, G., Noble, A., Blake, A.: Robust contour tracking in echocardiographic sequence. In: Proc. Intl. Conf. on Computer Vision, Bombay, India. (1998) 408–413
12. Blake, A., Isard, M., Reynard, D.: Learning to track the visual motion of contours. *Artificial Intelligence* **78** (1995) 101–133
13. Comaniciu, D.: Nonparametric information fusion for motion estimation. In: Proc. IEEE Conf. on Computer Vision and Pattern Recognition, Madison, Wisconsin. (2003) 59–66
14. Comaniciu, D., Zhou, X.S., Krishnan, S.: Robust real-time myocardial border tracking for echocardiography: An information fusion approach. In: *IEEE Trans. Medical Imaging*. (2004) to appear
15. Grewal, M.S., Andrews, A.P.: *Kalman Filtering: Theory and Practice*. Prentice Hall (1993)
16. Akgul, Y., Kambhamettu, C.: A coarse-to-fine deformable contour optimization framework. *IEEE Trans. Pattern Anal. Machine Intell.* **25** (2003) 174–186
17. Mikić, I., Krucinski, S., Thomas, J.D.: Segmentation and tracking in echocardiographic sequences: Active contours guided by optical flow estimates. *IEEE Trans. Medical Imaging* **17** (1998) 274–284