A Unified Framework for Uncertainty Propagation in Automatic Shape Tracking

X. S. Zhou, D. Comaniciu, B. Xie, R. Cruceanu,

Siemens Corporate Research, Inc.
Princeton, NJ 08536

A. Gupta

Siemens Medical Solutions USA, Inc.
Malvern, PA 19355

Abstract

Uncertainty handling plays an important role during shape tracking. We have recently shown that the fusion of measurement information with system dynamics and shape priors greatly improves the tracking performance for very noisy images such as ultrasound sequences [22]. Nevertheless, this approach required user initialization of the tracking process. This paper solves the automatic initialization problem by performing boosted shape detection as a generic measurement process and integrating it in our tracking framework. We show how to propagate the local detection uncertainties of multiple shape candidates during shape alignment, fusion with the predicted shape prior, and fusion with subspace constraints. As a result, we treat all sources of information in a unified way and derive the posterior shape model as the shape with the maximum likelihood. Our framework is applied for the automatic tracking of endocardium in ultrasound sequences of the human heart. Reliable detection and robust tracking results are achieved when compared to existing approaches and inter-expert variations.

1. Introduction

Measurement uncertainty plays an important role during shape tracking. It has been shown that the fusion of such information with system dynamics and shape priors greatly improves the tracking performance especially for very noisy images [22, 23]. In this paper, we aim at solving the automatic initialization problem through detection, and present a framework to incorporate information from all sources involved in the detection and tracking process.

Recent studies show that component-based object detectors can deal with large variations in pose and illumination, and are more robust under occlusions and heteroscedastic noise [19, 21, 9, 2, 6]. Analogously, for our motivating application of echocardiogram (ultrasound heart sequences) analysis [22], local appearance of the same anatomical structure (e.g., the septum) is similar across patients, while the *configuration* or *shape* of the heart can be dramatically different due to viewing angles or disease conditions, etc. Therefore, we propose to apply individually trained component detectors to exploit relatively stable local appearances, while using global shape models to constrain the component fusion process.

We estimate uncertainties in component detection in the form of covariance matrices, to take full advantage of the heteroscedastic [12, 10, 14] nature of the underlying anatomic structure and of the noise. A key step of component fusion with uncertainty is the constraint from a prior statistical model. The ideal formulation is to simultaneously optimize the component locations and the invariant transform that aligns the detection and the model. But close-form solutions are difficult to obtain and do not exist even for simple transforms. We use a sequential approach where two optimization processes propagate the uncertainties and provide us with the maximum likelihood solutions.

For capturing local appearance variations, classical solutions (e.g., Active Appearance Model [5]) rely on the Gaussian assumption. Recently, this assumption has been relaxed through the use of nonlinear learning machines such as SVM [19] or Boosting [7]. It has been demonstrated that AdaBoost-based object detectors are fast and effective for face and people [20] detection. In this paper, we use Adaboosting for component detection.

The state-of-the-art solutions such as those from Blake and Isard [4], Cootes, Taylor, and Cristinacce [5, 6], Jacob et al. [11], and Mitchell et al. [17], did not address the uncertainty issue for the detection and tracking of shapes beyond Kalman. We estimate and propagate detection and measurement uncertainties, and provide a unified view on different information sources involved in the shape detection and tracking process. As compared with existing pure detection schemes (e.g., [20, 6]) we incorporate tracking naturally within a joint fusion framework: the uncertainty of local detectors, the subspace statistical model, and the dynamic prediction are jointly considered during shape alignment, regularization, and tracking. Our contributions include: 1. a unified framework for optimally fusing uncertainties from local detection, motion dynamics, and subspace shape model during automatic shape detection and tracking; 2. boosted component detectors for left ventricle border localization in echocardiography sequences.

Next we state the problem and discuss boosted component detection. In Section 4 we present our handling of uncertainties in a unified detection and tracking framework. Experimental evaluations are presented in Section 5.

2. Problem Statement

We first define some terms used in this paper and our scope of discussion.

2.1. Pre-shape and Shape Space

We are concerned with sets of k *labeled* points in a 2-D Euclidean space where $k \geq 2$, and a set of invariant transforms. A set of k points will be called a *pre-shape* [13]. Any two pre-shapes will be regarded as having the same shape if either of them can be transformed into the other. With a common reference, the assemblage of all possible shapes forms the *shape space*[1].

By *shape tracking* instead of *contour tracking*, we stress that we detect and maintain point labels throughout the detection and tracking steps. This facilitates alignment and structural analysis, and tangential motion estimation.

2.2. Pre-shape Candidate Detection

Given an input image sequence, e.g., an ultrasound sequence of the heart in our case, the goal is to detect and track some underlying structures, such as the endocardial border, based on image appearance and prior knowledge regarding the shape. We adopt a component-based detection and tracking scheme in which we produce detection maps for local components separately. A "component" is one of the k points of a pre-shape. Mode detection is applied in the detection map with covariances estimated around each peak location. Each unique combination of detected components becomes a Gaussian in the $2k$-dimensional pre-shape space.

2.3. Model-Guided Optimization

Prior knowledge about geometry or configurations is represented by a pre-trained *shape model* using ground-truth data. The shape model, built in the shape space as defined above, captures variations among *pre-shapes modulo some invariant transforms* [13], such as similarity transforms. Principal component analysis (PCA), its kernel version, or independent component analysis (ICA) can be applied for finding the lower-dimensional subspace containing meaningful variations. We adopt PCA mostly because it enables concise analytical solutions in our framework.

The *transformation-free* shape model is used to evaluate and optimize among multiple detection candidates, based on a maximum likelihood formulation. The challenge is to simultaneously obtain the maximum likelihood point locations and the associated optimal transform, taking into account uncertainties in detection.

2.4. Tasks and Emphasis

Our tasks include: building a shape model using training data; detecting the candidate pre-shapes; pruning and optimization of the pre-shapes in the shape space guided by the model and by the system dynamics (during tracking).

The emphasis of this paper is on the estimation and propagation of uncertainties in local detection, and their influences over or interactions with: 1. the invariant transformation into the shape space (Section 4.1); 2. the shape space model constraint (Section 4.2); and 3. the tracking of the overall shape (Section 4.3).

3. Boosting for Component Detection

Boosting techniques [7] have been successfully applied in face and people detection [20]. The advantage of boosting as oppose to traditional Gaussian appearance models is that it can deal with complex distributions such as multi-modal distribution, which is common for our application. Boosting is also much faster than other non-linear alternatives such as kernel support vector machines [9].

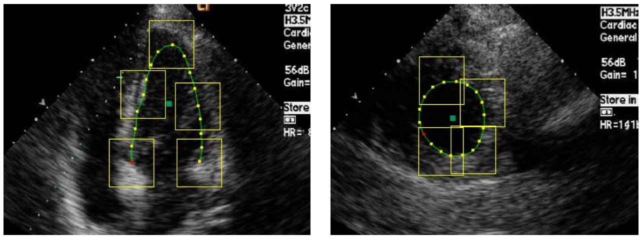
In this work, we apply AdaBoosting for local component detection, training one detector per point on the pre-shape. Unlike existing work where a "winner-takes-all" strategy is applied on component localization [9], we perform mode finding on each detection map and retain multiple modes, each with an estimated full covariance matrix characterizing the anisotropic uncertainty of that candidate location. The covariance matrix is obtained as a function of the Fisher information matrix or Hessian matrix estimated in the neighborhood of the peak location [12] (see Figure 6(d)$\sim$(f)).

One difficulty involving ultrasound images is the local signal drop-outs due to the directionality of the ultrasound beam [18] (Figure 1(b)), 6(a), 7(a) and (d)). A component-based scheme provides us with the flexibility of screening out training image patches from the drop-out regions. To do this, we simply delete training patches with very small trace for the Hessian matrix estimated in the patch (Figure 1).

The size of a training patch is adapted with respect to the heart size, which is measured by the average distance among control points.

With probabilistic component detection, we are faced with the issue of *pre-shape candidate evaluation*, and the problem of *model-guided shape constraining*. Even if we only have one candidate pre-shape, it would still be "fuzzy" with uncertainties in its point locations. The treatment of

[1] This definition of shape space is analogous to Kendall et al.'s definition [13], and are in agreement with that of Cootes and Taylor [5], but somewhat different from that of Blake and Isard ([4], p. 74) which is defined as linear subspace of the pre-shape space.



(a) (b)

Figure 1: Training and detection windows for AdaBoosting. (a) apical view: an open contour with 17 points; (b) parasternal short axis view: a closed contour with 18 points. non-informative patches, such as those from the drop-out region in (b), are screened out (see text).

such uncertainty in the context of other constraints is the focus of the next section.

4. Integrated Detection and Tracking

Given a candidate pre-shape denoted by $\mathcal{N}(\mathbf{x}, \mathbf{C_x})$, a multidimensional Gaussian distribution, with mean $\mathbf{x}$ and covariance $\mathbf{C_x}$, the first step is to find among the sample pre-shapes $\mathbf{x^o}$ the one that has maximum likelihood of being generated jointly by $\mathcal{N}(\mathbf{x}, \mathbf{C_x})$, the shape model $\mathcal{N}(\mathbf{m}, \mathbf{C_m})$, and the predicted shape $\mathcal{N}(\mathbf{x_-}, \mathbf{C_{x_-}})$ from previous time step, under an optimal invariant transform.

An equivalent formulation is to find $\mathbf{x}^*$ to minimize the sum of Mahalanobis distances in the pre-shape space and the transformed shape space, i.e.,

$$\mathbf{x}^* = \underset{\{\mathcal{T},\mathbf{x^o}\}}{\text{argmin}}\ d^2, \tag{1}$$

$$\begin{aligned} d^2 =& (\mathbf{x^{o\prime}} - \mathbf{m})^T \mathbf{C_m}^{-1} (\mathbf{x^{o\prime}} - \mathbf{m}) + (\mathbf{x^o} - \mathbf{x})^T \mathbf{C_x}^{-1} \\ & (\mathbf{x^o} - \mathbf{x}) + (\mathbf{x^o} - \mathbf{x_-})^T \mathbf{C_{x_-}}^{-1} (\mathbf{x^o} - \mathbf{x_-}), \end{aligned} \tag{2}$$

where $\mathbf{x^{o\prime}} = \mathcal{T}(\mathbf{x^o})$ with $\mathcal{T}$ being the invariant transform.

With multiple candidate pre-shapes, the one producing the highest likelihood, considering also the likelihood value in the detection map, wins at the decision time.

Eq. (2) requires simultaneous optimization over the location and the transform, and does not have a close-form solution even for simple transforms such as the similarity transform permitting only translation, rotation, and scaling[2]. The global optimal can be sought numerically through iterations, but the computation can be too expensive.

A more interesting issue arises when $\mathbf{C_m}$ is singular, i.e., the model resides in a subspace, the formulation breaks down and existing approximation approach (e.g., [2]) will not apply. One might try regularization on the covariance matrix $\mathbf{C_m}$ as suggested by Friedman [8], but this process adds back the noise dimensions which were deliberately removed, and it does not serve the purpose of this problem.

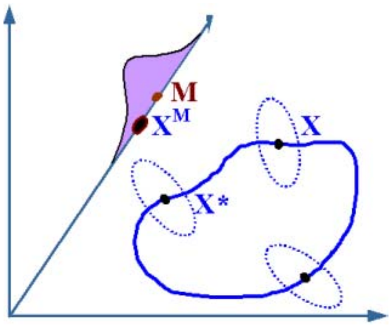


Figure 2: Invariance manifold for shape alignment. Under invariant transforms, the pre-shape $\mathbf{X}$ traverses a manifold, $\mathcal{C}$, illustrated by the thick curve. In general, $\mathcal{C}$ will not intersect the shape model subspace $\mathcal{F}$ (the slanted axis containing the model centroid $\mathbf{M}$).

The difficulty stems from the fact that the *manifold* (i.e., the *shape*) spanned by an arbitrary pre-shape through all possible transforms does not intersect the shape subspace in general, especially when the subspace dimension is relatively small. In our case, the shape sub-space have dimensions from 6 to 12, while the full Euclidean space has dimension $\geq$ 34. Figure 2 illustrates conceptually this relationship, with the thick curve depicting the manifold spanned by a pre-shape vector $\mathbf{X}$, and the slanted axis and a one dimensional Gaussian distribution representing the subspace model. The prediction is omitted here, or you may regard $\mathbf{X}$ as the fusion result of the detection and prediction (we will come back to this point in the sequel).

We propose a two-step optimization scheme as an alternative solution, with close-form solutions for both steps. This scheme can be easily explained using Figure 2: The first step is to go from $\mathbf{X}$ to $\mathbf{X}^*$, or in other words, to find the optimal transform from $\mathbf{X}$ to $\mathbf{M}$, using information in $\mathbf{C_x}$. The second step is to go from $\mathbf{X}^*$ to $\mathbf{X^M}$, using additional information from $\mathbf{C_M}$. We will call the first step *the alignment step*, and second *the constraining step*.

4.1. Shape Alignment with Uncertainty

The goal of this step is to consider the component uncertainties during the transform of the pre-shape and its covariance matrix toward the model. We minimize

$$d^2 = (\mathbf{m} - \mathbf{x}')^T \mathbf{C}_\mathbf{x}'^{-1} (\mathbf{m} - \mathbf{x}') \tag{3}$$

where $\mathbf{x}' = \mathcal{T}(\mathbf{x})$ and $\mathbf{C}'_\mathbf{x} = \mathcal{T}(\mathbf{C_x})$. To simplify notations, we have assumed that the prediction $\mathcal{N}(\mathbf{x_-}, \mathbf{C_{x_-}})$ has been fused into $\mathcal{N}(\mathbf{x}, \mathbf{C_x})$—we will discuss this step later.

When $\mathcal{T}$ is the similarity transform, we have:

$$\mathbf{x}' = \mathbf{R}\mathbf{x} + \mathbf{t}, \tag{4}$$

[2] Cootes and Taylor [5] showed that for a simpler case where only one weight matrix is considered, close-form solution exists.

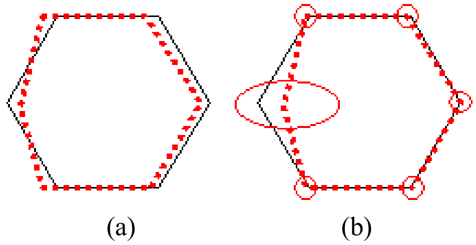


Figure 3: Shape alignment. (a) without considering uncertainties in localization; (b) with heteroscedastic [12, 10, 14] uncertainties. The ellipses depicts the covariance on point locations, representing information in a block-diagonal $\mathbf{C}_{\mathbf{x}}$. A full $\mathbf{C}_{\mathbf{x}}$ is not easy to visualize.

where $\mathbf{t}$ is the translation vector with two free parameters and $\mathbf{R}$ is a block diagonal matrix with each block being

$$R_i = \begin{pmatrix} a & -b \\ b & a \end{pmatrix} \tag{5}$$

With straight algebra we can rewrite Eq.(3) as follows:

$$\begin{aligned} d^2 =& (\mathbf{R}^{-1}(\mathbf{m}-\mathbf{t})-\mathbf{x})^T \mathbf{C}_{\mathbf{x}}{}^{-1}(\mathbf{R}^{-1}(\mathbf{m}-\mathbf{t})-\mathbf{x}) \\ =& (\mathcal{T}^{-1}(\mathbf{m})-\mathbf{x})^T \mathbf{C}_{\mathbf{x}}{}^{-1}(\mathcal{T}^{-1}(\mathbf{m})-\mathbf{x}) \end{aligned} \tag{6}$$

By taking derivative with respect to the four free parameters in $\mathbf{R}$ and $\mathbf{t}$, a close-form solution can be obtained. The details are omitted for space but one can consult the solution in ([5], p. 102), with an additional step to get back $\mathcal{T}$ from $\mathcal{T}^{-1}$. Figure 3 illustrate shape alignment with and without considering uncertainties in point locations. The intuition is to trust more the points with higher confidence.

4.2. Model Constraining with Uncertainty

With the pre-shape aligned with the model, we seek the shape with maximum likelihood of being generated by the two competing information sources, namely, the aligned detection/prediction versus the (subspace) model.

With a full-space model, the formulation is directly related to information fusion with Gaussian sources, or BLUE (best linear unbiased estimator) [3, 15].

4.2.1 BLUE

Given two noisy measurements of the same n-dimensional variable $\mathbf{x}$, each characterized by a multidimensional Gaussian distribution, $\mathcal{N}(\mathbf{x}_1, \mathbf{C}_1)$ and $\mathcal{N}(\mathbf{x}_2, \mathbf{C}_2)$, the maximum likelihood estimate of $\mathbf{x}$ is the point with the minimal sum of Mahalanobis distances, $\mathbf{D}^2(\mathbf{x}, \mathbf{x}_i, \mathbf{C}_i), i = 1, 2$, to the two centroids, i.e., $\mathbf{x}^* = \mathit{argmin}\ d^2$ with

$$\begin{aligned} d^2 =& \mathbf{D}^2(\mathbf{x}, \mathbf{x}_1, \mathbf{C}_1) + \mathbf{D}^2(\mathbf{x}, \mathbf{x}_2, \mathbf{C}_2) \\ =& (\mathbf{x}-\mathbf{x}_1)^T \mathbf{C}_1^{-1}(\mathbf{x}-\mathbf{x}_1) + (\mathbf{x}-\mathbf{x}_2)^T \mathbf{C}_2^{-1}(\mathbf{x}-\mathbf{x}_2) \end{aligned} \tag{7}$$

Taking derivative with respect to $\mathbf{x}$ and setting it to zero, we get the best linear unbiased estimate (BLUE) of $\mathbf{x}$ ([3, 15]):

$$\mathbf{x}^* = \mathbf{C}(\mathbf{C}_1^{-1}\mathbf{x}_1 + \mathbf{C}_2^{-1}\mathbf{x}_2) \tag{8}$$

$$\mathbf{C} = (\mathbf{C}_1^{-1} + \mathbf{C}_2^{-1})^{-1} \tag{9}$$

However, when the model resides in a subspace, the original fusion formulation needs to be modified as follows:

4.2.2 Subspace BLUE

Assume that, without loss of generality, $\mathbf{C}_2$ is singular. With the singular value decomposition of $\mathbf{C}_2 = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^T$, where $\mathbf{U} = [\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_n]$, with $\mathbf{u}_i$'s orthonormal and $\mathbf{\Lambda} = \mathit{diag}\{\lambda_1, \lambda_2, \ldots, \lambda_p, 0, \ldots, 0\}$, we rewrite Mahalanobis distance to $\mathbf{x}_2$ in Eq. (7) in the canonical form:

$$\begin{aligned} \mathbf{D}^2(\mathbf{x}, \mathbf{x}_2, \mathbf{C}_2) =& (\mathbf{x}-\mathbf{x}_2)^T \mathbf{C}_2^{-1}(\mathbf{x}-\mathbf{x}_2) \\ =& \sum_{i=1}^{n} \lambda_i^{-1} [\mathbf{U}^T(\mathbf{x}-\mathbf{x}_2)]^2 \end{aligned} \tag{10}$$

When λ_i tends to 0, $\mathbf{D}^2(\mathbf{x}, \mathbf{x}_2, \mathbf{C}_2)$ goes to infinity, unless $\mathbf{U}_0^T\mathbf{x} = 0$, where $\mathbf{U}_0 = [\mathbf{u}_{p+1}, \mathbf{u}_{p+2}, \ldots, \mathbf{u}_n]$. Here we have assumed, without loss of generality, that the subspace passes through the origin of the original space. Since $\mathbf{x}_2$ resides in the subspace, $\mathbf{U}_0{}^T\mathbf{x}_2 = \mathbf{0}$.

Because $\mathbf{U}_0{}^T\mathbf{x} = \mathbf{0}$, Eq. (7) now becomes:

$$\begin{aligned} d^2 =& (\mathbf{U}_p\mathbf{y}-\mathbf{x}_1)^T \mathbf{C}_1^{-1}(\mathbf{U}_p\mathbf{y}-\mathbf{x}_1) + \\ & (\mathbf{U}_p\mathbf{y}-\mathbf{x}_2)^T \mathbf{C}_2^{+}(\mathbf{U}_p\mathbf{y}-\mathbf{x}_2) \end{aligned} \tag{11}$$

where $\mathbf{y}$ is a $1 \times p$ vector.

Taking derivative with respect to $\mathbf{y}$ yields the fusion estimator for the subspace:

$$\mathbf{y}^* = \mathbf{C}_{\mathbf{y}^*}\mathbf{U}_p^T(\mathbf{C}_1^{-1}\mathbf{x}_1 + \mathbf{C}_2^{+}\mathbf{x}_2), \tag{12}$$

$$\mathbf{C}_{\mathbf{y}^*} = [\mathbf{U}_p^T(\mathbf{C}_1^{-1} + \mathbf{C}_2^{+})\mathbf{U}_p]^{-1}, \tag{13}$$

with equivalent expressions in the original space:

$$\mathbf{x}^* = \mathbf{U}_p\mathbf{y}^* = \mathbf{C}_{\mathbf{x}^*}(\mathbf{C}_1^{-1}\mathbf{x}_1 + \mathbf{C}_2^{+}\mathbf{x}_2) \tag{14}$$

$$\mathbf{C}_{\mathbf{x}^*} = \mathbf{U}_p\mathbf{C}_{\mathbf{y}^*}\mathbf{U}_p^T \tag{15}$$

It can be shown that $\mathbf{C}_{\mathbf{x}^*}$ and $\mathbf{C}_{\mathbf{y}^*}$ are the corresponding covariance matrices for $\mathbf{x}^*$ and $\mathbf{y}^*$. Notice that this solution is not a simple generalization of Eq. (8) by substituting pseudoinverses for regular inverses, which will not constrain $\mathbf{x}^*$ to be in the subspace.

Alternatively, we can write Eq. (12) and (13) as

$$\mathbf{y}^* = (\mathbf{U}_p^T\mathbf{C}_1^{-1}\mathbf{U}_p + \mathbf{\Lambda}_p^{-1})^{-1}(\mathbf{U}_p^T\mathbf{C}_1^{-1}\mathbf{x}_1 + \mathbf{\Lambda}_p^{-1}\mathbf{y}_2) \tag{16}$$

Here $\mathbf{y}_2$ is the transformed coordinates of $\mathbf{x}_2$ in the subspace spanned by $\mathbf{U}_p$, and $\mathbf{\Lambda}_p = \mathit{diag}\{\lambda_1, \lambda_2, \ldots, \lambda_p\}$. Eq. (16) can be seen as the BLUE fusion in the subspace of two Gaussian distributions, one is $\mathcal{N}(\mathbf{y}_2, \mathbf{\Lambda}_p)$ and the other is the "intersection" (not projection!) of $\mathcal{N}(\mathbf{x}_1, \mathbf{C}_1)$ in the subspace, $\mathcal{N}((\mathbf{U}_p^T\mathbf{C}_1^{-1}\mathbf{U}_p)^{-1}\mathbf{U}_p^T\mathbf{C}_1^{-1}\mathbf{x}_1, (\mathbf{U}_p^T\mathbf{C}_1^{-1}\mathbf{U}_p)^{-1})$.

4.3. Incorporating Dynamic Prediction

The above subspace fusion provides a general formulation for (subspace) model constraining, treating the shape measurement (with heteroscedastic uncertainty) and the PCA shape model as the two information sources. In the following, we add a third source that represents the dynamic prediction from tracking.

The crucial benefits we gain from *tracking*, on top of *detection*, are the additional information from system dynamics which governs the prediction, and the fusion of information across time. Based on the analysis above, the solution to Eq. (1) has the following form:

$$\begin{aligned}\mathbf{x}_{+} =& \mathbf{C}_{x_{+}}(\mathcal{T}\{(\mathbf{C}_{\mathbf{x}_{-}} + {\mathbf{C}_{\mathbf{x}}}^{-1})^{-1} \\ & (\mathbf{C}_{\mathbf{x}_{-}}\mathbf{x}_{-} + {\mathbf{C}_{\mathbf{x}}}^{-1}\mathbf{x})\} + {\mathbf{C}_{\mathbf{m}}}^{+}\mathbf{m}),\end{aligned} \tag{17}$$

$$\begin{aligned}\mathbf{C}_{x_{+}} =& \mathbf{U}_p[\mathbf{U}_p^T(\mathcal{T}\{(\mathbf{C}_{\mathbf{x}_{-}} + {\mathbf{C}_{\mathbf{x}}}^{-1})^{-1}\} \\ & + {\mathbf{C}_{\mathbf{m}}}^{+})\mathbf{U}_p]^{-1}\mathbf{U}_p^T,\end{aligned} \tag{18}$$

This solution puts information from detection, shape model, and dynamic prediction in one unified framework.

When the predicted shape is also confined in a subspace, we can simply apply the subspace BLUE formulation (Section 4.2.2) in a nested fashion inside the transform $\mathcal{T}$.

The prediction $\mathcal{N}(\mathbf{x}_{-}, \mathbf{C}_{\mathbf{x}_{-}})$ contains information from the system dynamics. In our case, we use it to encode global motion trends such as expansion and contraction, and slow translation and rotation. $\mathcal{N}(\mathbf{x}_{-}, \mathbf{C}_{\mathbf{x}_{-}})$ can be obtained using traditional method such as the prediction filter in a Kalman setting:

$$\mathbf{C}_{\mathbf{x}_{-}} = \mathbf{S}\mathbf{C}_{\mathbf{x}_{+},\mathbf{prev}}\mathbf{S}^T + \mathbf{Q}, \tag{19}$$

where the system dynamics equation is

$$\mathbf{x}_{-} = \mathbf{S}\mathbf{x}_{+,\mathbf{prev}} + \mathbf{q}, \tag{20}$$

and $\mathbf{Q}$ is the covariance of $\mathbf{q}$, and "prev" indicate information from the previous time step.

Figure 4 shows a schematic diagram of the analysis steps where the uncertainty of detection is propagated through all the steps. In a nutshell, we evaluate at each frame multiple detection candidates by comparing their likelihood in the context of both the shape model, and the prediction from the previous frame based on the system dynamics.

5. Experiments and Evaluations

In this paper we apply and evaluate our complete framework using echocardiographic images, in particular, the 2-D B-mode sequences. This is an very important application because ultrasound imaging of the heart provides direct visualization of cardiac structure and movement, and enables quantitative evaluation of heart structure and function. Computerized automatic detection and tracking assist such tasks by extracting myocardial borders in the image sequence so that diagnostic information can be computed such as the *ventricular volumes and geometry*, *cardiac output*, and *ejection fraction*, etc. [18]

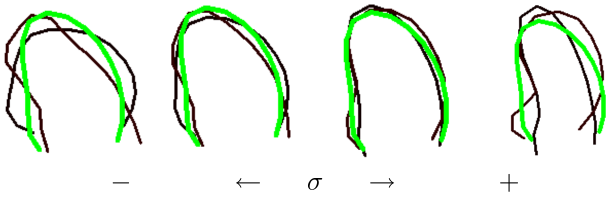


Figure 5: The PCA shape model for the apical view of left ventricle. The two dominant variations with multiples of standard deviations in both positive and negative directions. The thick green curve is the model mean.

To train models of appearance and shape, we use manually traced left ventricle endocardial borders (the inner border) as the training set. For apical views we use open contours each with 17 control points and for parasternal short axis views we use closed contours of 18 control points. The contours are drawn with a common starting point based on anatomy so that point correspondence can be maintained both across contours and with anatomical structures. Typical anatomical landmarks are the apex, the papillary muscles, and the septum. To train the boosted component detectors, image patches around the corresponding points from the manual borders are used (see Section 3 and Figure 1). Before learning shape variations, the training contours are first aligned to cancel out invariant transforms. In this paper, we consider global translation, rotation, and scaling as invariant for shapes, and model such transforms into the system dynamics instead. This ensures a small and meaningful shape subspace for the training set. The similarity alignment is achieved using an iterative Procrustes analysis approach [5]. PCA is then performed for each view. Figure 5 shows the dominant eigenshapes for the apical view along with its model mean.

5.1. Performance Measure

As performance measures, we use Mean Absolute Distance (MAD) [16] defined as follows: for the sequence S_i with m frames/contours, $\{c_1, c_2, ..., c_m\}$, where each contour c_j has n points $\{(x_{j,1}, y_{j,1}), (x_{j,2}, y_{j,2}), ..., (x_{j,n}, y_{j,n})\}$, the distance of S_i from the reference sequence S_i^0 is

$$\begin{aligned}MAD_i =& \frac{1}{m}\sum_{j=1}^{m} MAD_{i,j} \\ =& \frac{1}{m}\sum_{j=1}^{m}\frac{1}{n}\sum_{k=1}^{n}\sqrt{(x_{j,k} - x_{j,k}^0)^2 + (y_{j,k} - y_{j,k}^0)^2}\end{aligned} \tag{21}$$

The overall performance measure for a particular method is the averaged distance on the whole test set of l sequences:

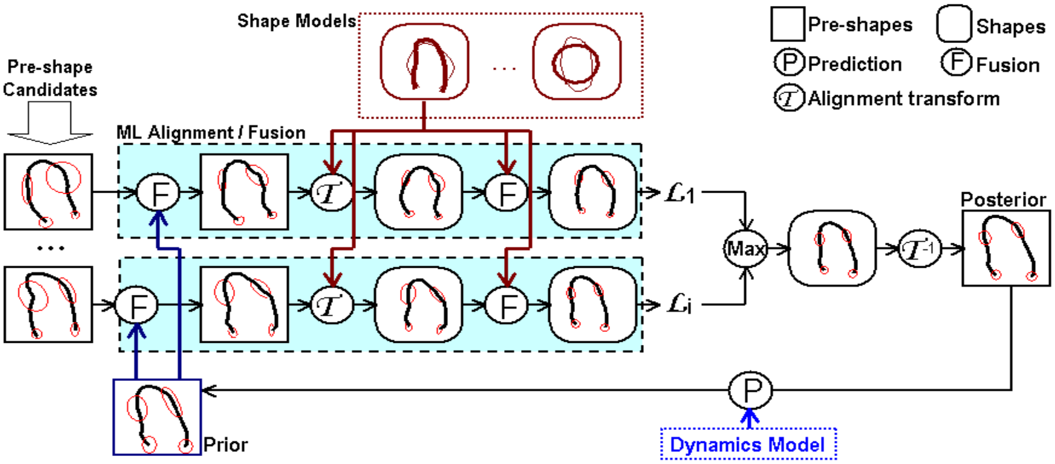


Figure 4: Uncertainty propagation during shape detection and tracking. The small red ellipses illustrate the location uncertainties. Notice that uncertainties are transformed with the shape during alignment and fused with the model and the (predicted) prior information during likelihood estimation and tracking.

$MAD = \frac{1}{l}\sum_{i=1}^{l} MAD_i$. These measures do *not* give higher weights to longer sequences which simply have more cycles. We also record the standard deviation of the distance for each frame and average the results for each sequence first, and then across sequences, again without overweighing longer sequences.

A crucial difference between our distance measures and those of [1] or [16] is that we have the point correspondence, and we want to measure the performance of these correspondences. In [1, 16], no correspondence is assumed and the nearest point from the other contour is taking as the corresponding point–As a result, motion component along the tangent of the contour *cannot* be evaluated. In reality, global or regional tangent motion are common during a cardiac cycle, and they reveal crucial information regarding cardiac function.

5.2. Localization Performance

It is observed during our interaction with cardiologists and sonographers that inter-expert variabilities on *border localization* are significant, especially for noisy cases. This is also evident in our training data. It would be insightful to put the performance of our algorithm in the context of such expert variabilities.

We tested our algorithm on data outside of the training set from 15 patients. Each patient has either one or two sequences, with about 600 frames. All sequences are traced by one expert, and some sequences or frames are traced by another expert (typically on end-diastolic or end-systolic frames). The overall detection and tracking performance is measured with reference to the first expert using MAD and compared with that of the second expert which serves as a measure of inter-expert variations. The results are shown in Table 1. We see that the performance is similar and also robust with a comparable standard deviation.

Table 1: Performance of proposed method as compared to variations between experts.

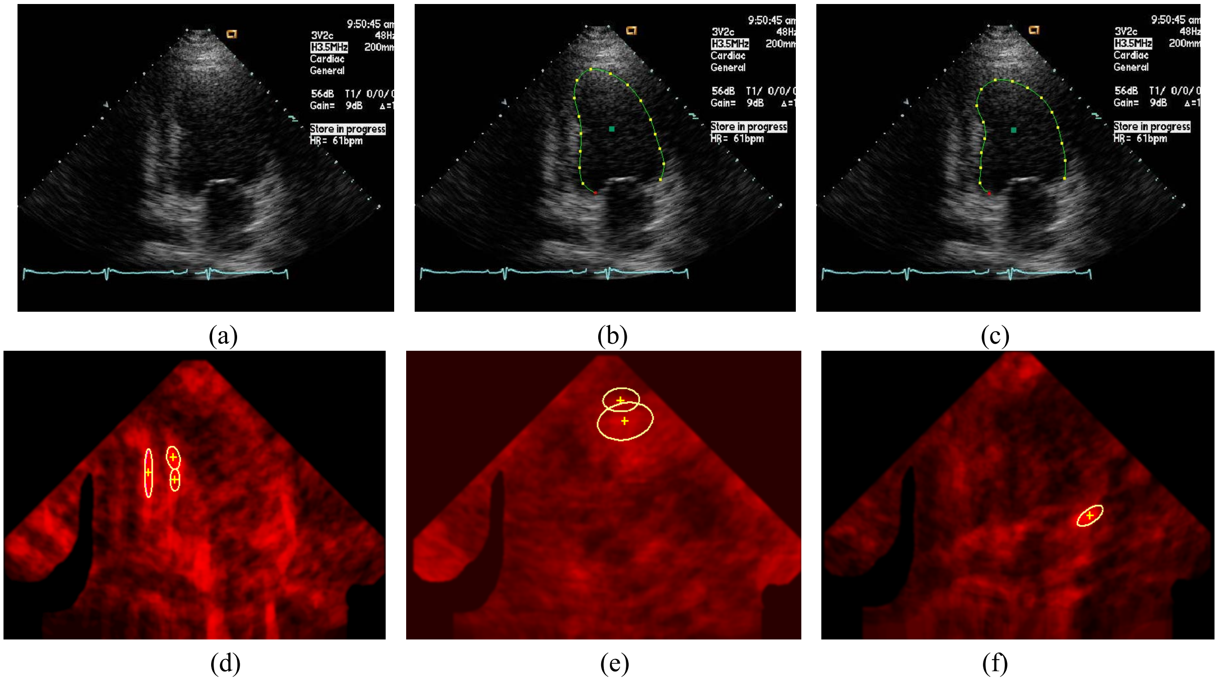
	MAD	$\bar{\sigma}_{MAD}$
ExpertVariation	7.7435	3.6800
DetectionTracking	8.8432	4.4126

5.3. Does Uncertainty Help?

A more interesting observation we have is that the inter-expert variabilities on *border localization* are much more significant than the variabilities on *motion tracking*. In other words, experts often do not agree on precise border locations, but they more or less agree on the relative motion of the border across frames[3]. Therefore, it is reasonable to evaluate algorithms against multiple experts, given that we always use for each expert his or her first contour for initialization. The resulting MAD value, unlike the previous case, should be zero ideally, if the expert is perfect in estimating motion and the algorithm agrees. Table 2 shows the results over the 15-patient test set where we use the first contour from an expert as the initialization and compare the overall performances with and without propagating uncertainties. The table clearly shows the improvement by estimating and propagating uncertainties, not only in terms of reduction of averaged error in motion capturing but also in terms of reduced variance.

Some detection and tracking examples are shown in Figure 6 and Figure 7.

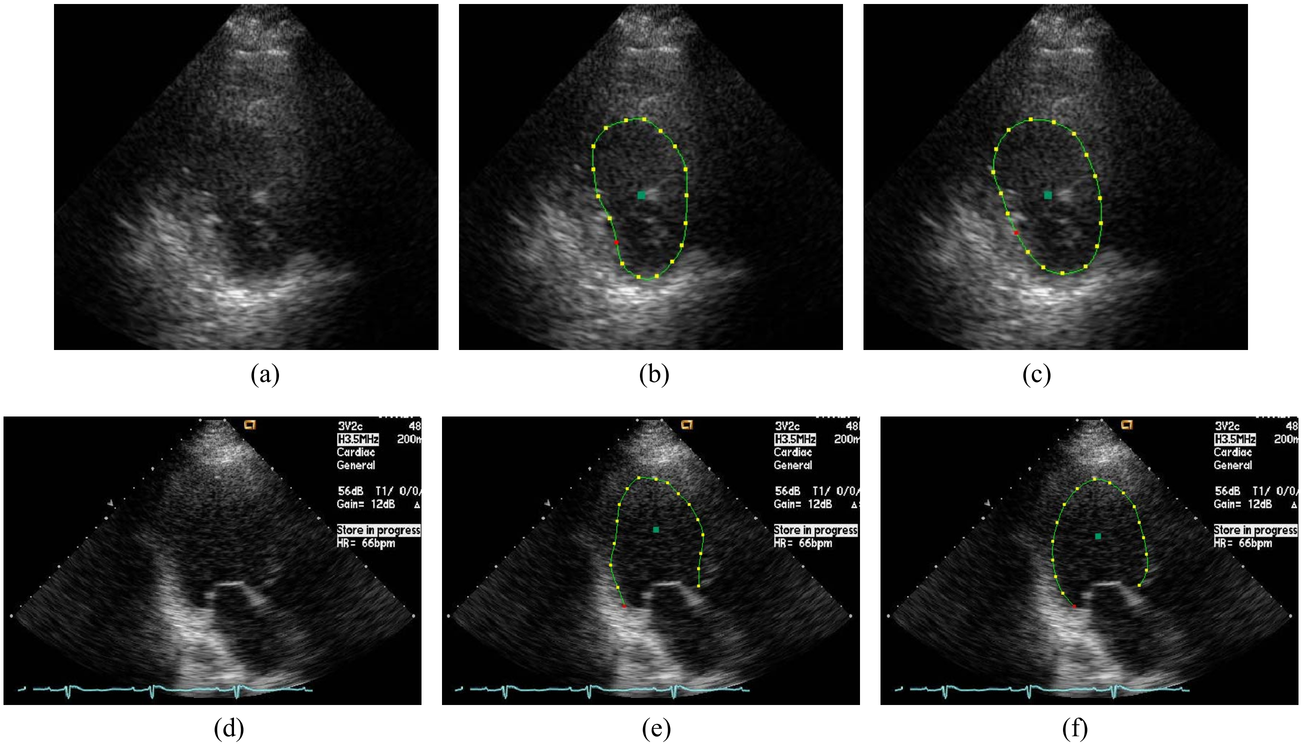
[3] This is probably explainable by the superb ability of human visual system in capturing relative motion



(a) (b) (c)

(d) (e) (f)

Figure 6: Left ventricle endocardial border detection. (a) input image; (b) detected contour; (c) contour by an expert; (d)∼(f) the local detection map for the 4th, 10th, and 17th point, respectively. Notice the heteroscedastic nature [12, 10, 14] of the local detection uncertainties (depicted by the ellipses in (d)∼(f)). Local detection ambiguities (not only between multiple modes, e.g., (d) and (e), but also within one mode) are resolved in the context of tracking and shape constraints.



(a) (b) (c)

(d) (e) (f)

Figure 7: More detection results. From left to right: the input image; the detected contour; and the contour by an expert.

Table 2: Performances with and without using the uncertainty information.

	MAD	$\bar{\sigma}_{MAD}$
WithoutUncertainty	4.0543	3.0239
WithUncertainty	3.4075	2.6488

6. Conclusions

The focus of this paper is on the uncertainty handling with a shape space constraint and a dynamic model. We presented a unified way of treating boosted detection uncertainties in an automatic shape tracking framework. From another angle, the detection benefits from the prediction information and information fusion across time. Future research includes uncertainty propagation for multiple hypothesis tracking, with the maximum likelihood formulation extended to multiple frames to achieve additional performance boost.

Acknowledgments: We would like to thank Sriram Krishnan from Siemens Medical Solutions for insightful discussions.

References

[1] Y. Akgul and C. Kambhamettu, "A coarse-to-fine deformable contour optimization framework," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 25, no. 2, pp. 174–186, 2003.

[2] V. R. M. S. T. B. B. Xie, D. Comaniciu, "Component fusion for face detection in the presence of heteroscedastic noise," in *Annual Conf. German Soc. Patt. Recog. (DAGM'03),* Magdeburg, Germany, 2003, pp. 434–441.

[3] Y. Bar-Shalom and L. Campo, "The effect of the common process noise on the two-sensor fused track covariance," *IEEE Trans. Aero. Elect. Syst.*, vol. AES-22, no. 22, pp. 803–805, 1986.

[4] A. Blake and M. Isard, *Active contours*. Springer Verlag, 1998.

[5] T. Cootes and C. Taylor, "Statistical models for appearance for computer vision," 2001. Unpublished manuscript. Available at `http://www.wiau.man.ac.uk/~bim/Models/app_model.ps.gz`.

[6] D. Cristinacce and T. Cootes, "Facial feature detection using adaboost with shape constraints," in *British Machine Vision Conference*, volume 1, 2003, pp. 231–240.

[7] Y. Freund and R. E. Schapire, "Experiments with a new boosting algorithm," in *Proc. Int'l Conf. on Machine Learning*, 1996, pp. 148–156.

[8] J. Friedman, "Regularized discriminant analysis," *Journal of the American Statistical Association*, vol. 84, no. 405, pp. 165–175, 1989.

[9] B. Heisele, T. Serre, M. Pontil, and T. Poggio, "Component-based face detection," in *CVPR01*, 2001, pp. I:657–662.

[10] M. Irani and P. Anandan, "Factorization with uncertainty," in *Proc. 6th European Conf. on Computer Vision,* Dublin, Ireland, 2000, pp. 539–553.

[11] G. Jacob, J. Noble, C. Behrenbruch, A. Kelion, and A. Banning, "A shape-space-based approach to tracking myocardial borders and quantifying regional left-ventricular function applied in echocardiography," *IEEE Trans. Medical Imaging*, vol. 21, no. 3, pp. 226–238, 2002.

[12] Y. Kanazawa and K. Kanatani, "Do we really have to consider covariance matrices for image features?," in *Proc. Intl. Conf. on Computer Vision,* Vancouver, Canada, volume II, 2001, pp. 586–591.

[13] D. G. Kendall, D. Barden, T. K. Carne, and H. Le, *Shape and Shape Theory*. Chichester: John Wiley & Sons, Ltd., 1999.

[14] Y. Leedan and P. Meer, "Heteroscedastic regression in computer vision: Problems with bilinear constraint," *Intl. J. of Computer Vision*, vol. 37, no. 2, pp. 127–150, 2000.

[15] X. Li, Y. Zhu, and C. Han, "Unified optimal linear estimation fusion - part i: Unified models and fusion rules," in *Proc. of 3rd Intl. Conf. on Information Fusion,* Paris*, France*, 2000, pp. MoC2–10–MoC2–17.

[16] I. Mikić, S. Krucinski, and J. D. Thomas, "Segmentation and tracking in echocardiographic sequences: Active contours guided by optical flow estimates," *IEEE Trans. Medical Imaging*, vol. 17, no. 2, pp. 274–284, 1998.

[17] S. Mitchell, J. G. Bosch, B. P. F. Lelieveldt, R. van der Geest, J. Reiber, and M. Sonka, "3-d active appearance models: segmentation of cardiac mr and ultrasound images," *IEEE Trans. Medical Imaging*, vol. 21, no. 9, pp. 1167–1178, 2002.

[18] C. M. Otto, *Textbook of Clinical Echocardiography*. W. B. Saunders, Philadelphia, 2 edition, 2000.

[19] C. Papageorgiou and T. Poggio, "A trainable system for object detection," *Intl. J. of Computer Vision*, vol. 38, no. 1, pp. 15–33, 2000.

[20] P. Viola, M. Jones, and D. Snow, "Detecting pedestrians using patterns of motion and appearance," in *Proc. Intl. Conf. on Computer Vision,* Nice, France, 2003.

[21] V. Vogelhuber and C. Schmid, "Face detection based on generic local descriptors and spatial constraints," in *Int'l Conf. on Patt. Recog.*, volume 1, 2000, pp. 1084–1087.

[22] X. S. Zhou, D. Comaniciu, and S. Krishnan, "An information fusion framework for robust shape tracking," in *Int'l Workshop on Statistical and Computational Theories of Vision,* Nice, France, 2003.

[23] X. S. Zhou, D. Comaniciu, and S. Krishnan, "Coupled-contour tracking through non-orthogonal projections and fusion for echocardiography," in *Proc. European Conf. on Computer Vision,* Prague, Czech Republic, 2004.