Coupled-Contour Tracking through Non-orthogonal Projections and Fusion for Echocardiography

Xiang Sean Zhou[1], Dorin Comaniciu[1], and Sriram Krishnan[2]

[1] Siemens Corporate Research, 755 College Road East, Princeton, NJ 08540, USA
[2] Siemens Medical Solutions, 51 Valley Stream Pkwy, Malvern, PA 19355, USA
xiang.zhou,dorin.comaniciu,sriram.krishnan@siemens.com

Abstract. Existing methods for incorporating subspace model constraints in contour tracking use only partial information from the measurements and model distribution. We propose a complete fusion formulation for robust contour tracking, optimally resolving uncertainties from heteroscedastic measurement noise, system dynamics, and a subspace model. The resulting *non-orthogonal subspace projection* is a natural extension of the traditional model constraint using orthogonal projection. We build models for coupled double-contours, and exploit information from the ground truth initialization through a strong model adaptation. Our framework is applied for tracking in echocardiograms where the noise is *heteroscedastic*, each heart has distinct shape, and the relative motions of epi- and endocardial borders reveal crucial diagnostic features. The proposed method significantly outperforms the traditional shape-space-constrained tracking algorithm. Due to the joint fusion of heteroscedastic uncertainties, the strong model adaptation, and the coupled tracking of double-contours, robust performance is observed even on the most challenging cases.

1 Introduction

Model constraints can significantly improve the performance of a contour tracking algorithm. In most cases, a subspace model is appropriate 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 [1]. A traditional treatment is to project into a PCA subspace [2,1]. However, this approach does not take advantage of heteroscedastic (i.e., both anisotropic and inhomogeneous) measurement uncertainties [3,4] (See Figure 1). Intuitively, a tracking algorithm should downplay measurements from uncertain regions when consulting a shape model.

A more interesting solution was to directly incorporate a PCA shape space constraint into a Kalman filter-based tracker. In [7,8], the proposal was to set the system noise covariance matrix to be the covariance of a PCA shape model. Nevertheless, this treatment has some limitations. First of all, it did not provide a systematic and complete fusion of the model information because, for example, the model mean is discarded(–as a result, the projection can be arbitrarily far from the model mean in the subspace). Secondly, it mixes the uncertainty from system dynamics with the uncertainty from the statistical shape constraint, while these two can be conceptually different. For example, we may want to use the dynamic model to capture different modes of global rigid motion, while

T. Pajdla and J. Matas (Eds.): ECCV 2004, LNCS 3021, pp. 336–349, 2004.

(a)

(b)

Fig. 1. Ellipses depicting uncertainties in feature localization and motion estimation. The heteroscedastic nature stems from either the *aperture problem* [4,5], or for echocardiograms (b), the *acoustic drop-out* [6].

applying a statistical shape model to control the modes and range of shape variations. Finally, existing solutions do not 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 reliable performance.

We decouple the uncertainty in system dynamics and the statistical shape constraint, and introduce a unified framework for fusing *a subspace shape model with the system dynamics and the measurements with heteroscedastic noise*. We build models for coupled double-contours so that more information can be integrated especially for very noisy data. The double-contour also achieves better preservation of topology[1]. To accommodate individual shape characteristics, the generic shape model is strongly adapted using information given about the current case. The subspace model can take the form of a specific subspace distribution, e.g., a Gaussian, or a simple subspace constraint, e.g., the eigenspace model [2,12]. Unlike existing *ad hoc* formulations, our framework treats the two cases in a consistent way, and combines such constraints seamlessly into the tracking framework. The new approach calls for reliable estimation of measurement uncertainties, for which we employ a recent robust solution to the motion estimation problem, which also computes the motion flow uncertainties [13].

The paper is organized as follows: The new model-constrained tracking formulation is presented in Section 2. Section 3 discusses a model adaptation scheme. Section 4 contains experimental evaluation and analysis. Related work and future directions are discussed in Sections 5 and 6, respectively.

2 Model Constraint through Projection and Fusion

Throughout this paper, we represent shapes by control or landmark points, assuming correspondence. These points are fitted by splines before shown to the user. For more implementation details, please refer to Section 4.

[1] Our coupling is probabilistic (governed by the training set) and "soft"(See Section 4.4). For deterministic coupling, "soft" or "hard", please refer to [9,10,11] and the references therein.

A typical tracking framework fuses information from the dynamic prediction and from noisy measurements. For shape tracking, additional constraints are necessary to stabilize the overall shape in a feasible space/range. In this section, we first extend the traditional subspace constraint using orthogonal projection to non-orthogonal projection. Then, we show that with a complete subspace model constraint, considering also the model mean, this can be further generalized into an information fusion formulation. Finally, these formulas are uniformly combined into a tracking framework.

2.1 Non-orthogonal Projection for Heteroscedastic Noise

Given an n-dimensional measurement point[2], $\mathbf{x}$, with uncertainty characterized by a covariance matrix $\mathbf{C}$, we want to find the "closest" point $\mathbf{y}^*$ in a p-dimensional ($p < n$) subspace, with its axes defined by the orthonormal column vectors of an $n \times p$ matrix, $\mathbf{U}_p$, $\mathbf{U}_p^T\mathbf{U}_p = \mathbf{I}$, such that the Mahalanobis distance is minimized, i.e., $\mathbf{y}^* = \textit{argmin}\ d^2$, where

$$d^2 = (\mathbf{U}_p\mathbf{y} - \mathbf{x})^T\mathbf{C}^{-1}(\mathbf{U}_p\mathbf{y} - \mathbf{x}). \tag{1}$$

This is in the form of a *weighted least square* ([14], p. 386). By taking derivative of above with respect to $\mathbf{y}$ and setting it to 0, we have

$$\mathbf{y}^* = \mathbf{C}_{\mathbf{y}^*}\mathbf{U}_p^T\mathbf{C}^{-1}\mathbf{x}, \qquad \mathbf{C}_{\mathbf{y}^*} = (\mathbf{U}_p^T\mathbf{C}^{-1}\mathbf{U}_p)^{-1}. \tag{2}$$

In general, this is a *non-orthogonal projection*. It is easy to show that the Gaussian $\mathcal{N}(\mathbf{y}^*, \mathbf{C}_{\mathbf{y}^*})$ is the conditional distribution, or *intersection*, of $\mathbf{x}$ in the subspace. Only when $\mathbf{C} = c\mathbf{I}$ with some positive scaler c, we have

$$\mathbf{y}^* = (c^{-1}\mathbf{U}_p^T\mathbf{I}\mathbf{U}_p)^{-1}\mathbf{U}_p^T(c\mathbf{I})^{-1}\mathbf{x} = \mathbf{U}_p^T\mathbf{x}, \qquad \mathbf{C}_{\mathbf{y}^*} = c\mathbf{I}_p \tag{3}$$

In our application, this means that all control points on the contour have isotropic and homogeneous uncertainties and the solution reduces to classical orthogonal projection.

2.2 Incorporating Model Distribution through Subspace Fusion

In the above we only considered the subspace constraint while the actual model distribution (assumed Gaussian with mean and covariance) represents important prior information that should not be discarded. In the sequel we show that an information fusion formulation unifies all cases within a general maximal likelihood framework.

The *information space* is the space obtained by multiplying a vector by its corresponding *information matrix*, which is, in the Gaussian case, the inverse of the error covariance matrix. Given two noisy measurements of an n-dimensional variable $\mathbf{x}$, each with a 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, $\mathbf{D}^2(\mathbf{x}, \mathbf{x}_i, \mathbf{C}_i) = (\mathbf{x} - \mathbf{x}_i)^T\mathbf{C}_i^{-1}(\mathbf{x} - \mathbf{x}_i)$, to the two centroids, i.e., $\mathbf{x}^* = \textit{argmin}\ d^2$ with

$$d^2 = (\mathbf{x} - \mathbf{x}_1)^T\mathbf{C}_1^{-1}(\mathbf{x} - \mathbf{x}_1) + (\mathbf{x} - \mathbf{x}_2)^T\mathbf{C}_2^{-1}(\mathbf{x} - \mathbf{x}_2) \tag{4}$$

[2] Care should be taken to avoid confusion over the interpretation of the term "point": the *point* here would correspond to a contour with multiple *control points*. By "inhomogeneous" noise, we refer to the inhomogeneity among different *control points* [4].

Taking derivative with respect to $\mathbf{x}$ and setting it to zero, we get:

$$\mathbf{x}^* = \mathbf{C}(\mathbf{C}_1^{-1}\mathbf{x} + \mathbf{C}_2^{-1}\mathbf{x}_2), \qquad \mathbf{C} = (\mathbf{C}_1^{-1} + \mathbf{C}_2^{-1})^{-1} \tag{5}$$

which is also known as the best linear unbiased estimate (BLUE) of $\mathbf{x}$ ([15,16]).

When one of the Gaussians is in a subspace of dimension p, e.g., $\mathbf{C}_2$ is singular, the second term of Eq. (4) can be re-written using pseudoinverse of $\mathbf{C}_2$, $\mathbf{C}_2^+$:

$$\mathbf{D}^2(\mathbf{x}, \mathbf{x}_2, \mathbf{C}_2) = \sum_{i=1}^{p} \lambda_i^{-1}[\mathbf{U}_p^T(\mathbf{x} - \mathbf{x}_2)]^2 \equiv (\mathbf{x} - \mathbf{x}_2)^T \mathbf{C}_2^+ (\mathbf{x} - \mathbf{x}_2) \tag{6}$$

with the additional constraint of $\mathbf{U}_0^T\mathbf{x} = 0$ (otherwise, d will diverge). Here $\mathbf{C}_2 = \mathbf{U}\Lambda\mathbf{U}^T$, $\mathbf{U} = [\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_n]$, $\mathbf{U}_p = [\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_p]$, $\mathbf{U}_0 = [\mathbf{u}_{p+1}, \mathbf{u}_{p+2}, \ldots, \mathbf{u}_n]$, with $\mathbf{u}_i$'s orthonormal and $\Lambda = diag\{\lambda_1, \lambda_2, \ldots, \lambda_p, 0, \ldots, 0\}$. (Here we have assumed, without loss of generality, that the subspace passes through the origin of the original space.)

With $\mathbf{U}_0^T\mathbf{x} = \mathbf{0}$, $\mathbf{x}$ resides in the subspace as $\mathbf{y} = \mathbf{U}_p^T\mathbf{x}$. Eq. (4) now takes the following general form:

$$d^2 = (\mathbf{U}_p\mathbf{y} - \mathbf{x}_1)^T\mathbf{C}_1^{-1}(\mathbf{U}_p\mathbf{y} - \mathbf{x}_1) + (\mathbf{U}_p\mathbf{y} - \mathbf{x}_2)^T\mathbf{C}_2^+(\mathbf{U}_p\mathbf{y} - \mathbf{x}_2) \tag{7}$$

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

$$\mathbf{y}^* = \mathbf{C}_{\mathbf{y}^*}\mathbf{U}_p^T(\mathbf{C}_1^{-1}\mathbf{x}_1 + \mathbf{C}_2^+\mathbf{x}_2), \qquad \mathbf{C}_{\mathbf{y}^*} = [\mathbf{U}_p^T(\mathbf{C}_1^{-1} + \mathbf{C}_2^+)\mathbf{U}_p]^{-1} \tag{8}$$

Equivalent expressions can be obtained in the original space as:

$$\mathbf{x}^* = \mathbf{U}_p\mathbf{y}^* = \mathbf{C}_{\mathbf{x}^*}(\mathbf{C}_1^{-1}\mathbf{x}_1 + \mathbf{C}_2^+\mathbf{x}_2), \qquad \mathbf{C}_{\mathbf{x}^*} = \mathbf{U}_p\mathbf{C}_{\mathbf{y}^*}\mathbf{U}_p^T \tag{9}$$

It is easy to show that $\mathbf{C}_{\mathbf{x}^*}$ and $\mathbf{C}_{\mathbf{y}^*}$ are the covariance matrices for $\mathbf{x}^*$ and $\mathbf{y}^*$.

Alternatively, we can write Eq. (8) as

$$\mathbf{y}^* = (\mathbf{U}_p^T\mathbf{C}_1^{-1}\mathbf{U}_p + \Lambda_p^{-1})^{-1}(\mathbf{U}_p^T\mathbf{C}_1^{-1}\mathbf{x}_1 + \Lambda_p^{-1}\mathbf{y}_2) \tag{10}$$

Here $\mathbf{y}_2 = \mathbf{U}_p^T\mathbf{x}_2$, and $\Lambda_p = diag\{\lambda_1, \lambda_2, \ldots, \lambda_p\}$. Interestingly, Eq. (10) is in fact the BLUE fusion of two subspace Gaussian distributions, one being $\mathcal{N}(\mathbf{y}_2, \Lambda_p)$ and the other being the *non-orthogonal projection* of $\mathcal{N}(\mathbf{x}_1, \mathbf{C}_1)$ in the subspace, $\mathcal{N}((\mathbf{U}_p^T\mathbf{C}_1^{-1}\mathbf{U}_p)^{-1}\mathbf{U}_p^T\mathbf{C}_1^{-1}\mathbf{x}_1, (\mathbf{U}_p^T\mathbf{C}_1^{-1}\mathbf{U}_p)^{-1})$ (*cf.* Eq. (2)).

2.3 Constrained Tracking through Fusion and Projection

To integrate the above projection and fusion formulas into a tracking framework, we first note that Kalman filter is essentially *fusion* in nature, which is evident in its information filter form ([17], page 138), which :

$$\mathbf{x}_{k+1|k+1} = (\mathbf{P}_{k+1|k}^{-1} + \mathbf{H}^T\mathbf{R}^{-1}\mathbf{H})^{-1}(\mathbf{P}_{k+1|k}^{-1}\mathbf{x}_{k+1|k} + \mathbf{H}^T\mathbf{R}^{-1}\mathbf{z}_{k+1}) \tag{11}$$

Here $\mathbf{x}_{i|j}$ is the state estimate at time i given the state or measurement at time j, $\mathbf{P}$ is the state covariance, and $\mathbf{H}$ is the measurement matrix. The measurement model is

$\mathbf{z}_k = \mathbf{H}\mathbf{x}_k + \mathbf{r}_k$, where $\mathbf{r}_k$ represents measurement noise with covariance $\mathbf{R}$. $\mathbf{P}$ is recursively updated as $\mathbf{P}_{k+1|k} = \mathbf{S}\mathbf{P}_{k|k}\mathbf{S}^T + \mathbf{Q}$ using information from a dynamic system model $\mathbf{x}_{k+1} = \mathbf{S}\mathbf{x}_k + \mathbf{q}_k$, where $\mathbf{q}_k$ represents system noise with covariance $\mathbf{Q}$ [17].

For the special case where $\mathbf{H}$ is a square matrix and admits an inverse, we can see Eq. (11) in a strict information fusion form, namely, the fusion of prediction and measurement in the information space (*cf.* Eq. (5)):

$$\mathbf{x}_{k+1|k+1} = (\mathbf{P}_{k+1|k}^{-1} + \mathbf{R}_{\mathbf{x}}^{-1})^{-1} \left[\mathbf{P}_{k+1|k}^{-1}\mathbf{x}_{k+1|k} + \mathbf{R}_{\mathbf{x}}^{-1}\mathbf{x}_{\mathbf{z},k+1}\right] \tag{12}$$

where $\mathbf{R}_{\mathbf{x}} = \mathbf{H}^{-1}\mathbf{R}(\mathbf{H}^{-1})^T$ and $\mathbf{x}_{\mathbf{z},k+1} = \mathbf{H}^{-1}\mathbf{z}_{k+1}$.

Because Kalman filter is a fusion filter and the information fusion operation is *associative*, we can apply the subspace fusion formula, Eq. (8), on the Kalman fusion result of Eq. (11) (In general $\mathbf{H}$ is not invertible; otherwise, Eq. (12) can be used.) and a subspace source $\mathcal{N}(\mathbf{x}_2, \mathbf{C}_2)$, to obtain a complete fusion formula:

$$\mathbf{x}_{k+1|k+1} = \mathbf{P}_{k+1|k+1}((\mathbf{S}\mathbf{P}_{k|k}\mathbf{S}^T + \mathbf{Q})^{+}\mathbf{x}_{k+1|k} + \mathbf{H}^T\mathbf{R}^{-1}\mathbf{z}_{k+1} + \mathbf{C}_2^{+}\mathbf{x}_2) \tag{13}$$

$$\mathbf{P}_{k+1|k+1} = \mathbf{U}_p[\mathbf{U}_p^T((\mathbf{S}\mathbf{P}_{k|k}\mathbf{S}^T + \mathbf{Q})^{+} + \mathbf{H}^T\mathbf{R}^{-1}\mathbf{H} + \mathbf{C}_2^{+})\mathbf{U}_p]^{-1}\mathbf{U}_p^T \tag{14}$$

Observe the symmetry of the solution which *combines all the available knowledge in the information space*. These equations provide a unified fusion of the system dynamics, a subspace model, and measurement noise information. They represent the complete representation of various uncertainties that affect the tracking system.

Compared to a PCA *shape space* representation [7,8], the above formulation uses not only the model subspace (the eigenvectors), but also the actual model distribution, in a unified fusion framework. On the other hand, if only a subspace constraint is desired, we can simply apply the special case of Eq. (2) on Eq. (11), and the resulting *non-orthogonal projection* is still within the same analytical framework.

3 Updating Shape Model: Fusion versus Model Adaptation

The use of a model learned from a pool of training samples to guide a specific case is inherently problematic, especially when novel variations commonly appear. Theoretically, what we really need is *the deformation model of the current case*. Therefore, there is a strong need to update the generic model to reflect what is already known for the current case. A natural choice is to use the initial contour (by hand or through automatic detection) to update the existing model. In the context of the preceding sections, an intriguing question would be *why don't we use* fusion *on the model and the new contour* by assigning some covariance $\mathbf{C} = \alpha\mathbf{I}$ for the new contour? The answer turns out to be negative and we will get to the reasons at the end of this section.

An alternative tool is incremental PCA (IPCA) [18], but this strategy does not adapt in a sufficiently strong manner. Therefore, we put more emphasis on the new data, and apply a *strongly-adapted-PCA* (SA-PCA) model as follows: We assume that the existing PCA model and the initial contour from the current new case *jointly* represent

the variations of the current case, but with relative energy, or representative power, being α and $(1-\alpha)$, respectively, with $0 < \alpha < 1$.

If the original covariance matrix $\mathbf{C}$ were stored (when the original dimensionality is not forbiddingly high), the adapted mean $\mathbf{x}_m^{new}$ and covariance matrix $\mathbf{C}^{new}$ would simply be the weighted sum of the two contributing sources:

$$\mathbf{x}_m^{new} = \alpha \mathbf{x}_m + (1-\alpha)\mathbf{x} \tag{15}$$

$$\begin{aligned}\mathbf{C}^{new} =& \alpha(\mathbf{C} + (\mathbf{x}_m - \mathbf{x}_m^{new})(\mathbf{x}_m - \mathbf{x}_m^{new})^T) + (1-\alpha)(\mathbf{x} - \mathbf{x}_m^{new})(\mathbf{x} - \mathbf{x}_m^{new})^T \\ =& \alpha\mathbf{C} + \alpha(1-\alpha)(\mathbf{x} - \mathbf{x}_m)(\mathbf{x} - \mathbf{x}_m)^T \end{aligned} \tag{16}$$

Eigenanalysis can be performed on $\mathbf{C}^{new}$ to obtain the new subspace model.

A more interesting and practical scenario is when $\mathbf{C}$ is not stored and $\{\mathbf{x}_m, \Lambda, \mathbf{U}\}$ resides only *in the subspace*. Denote the subspace component of $\mathbf{x}$ as $\mathbf{x}_s = \mathbf{U}^T\mathbf{x}_d$, where $\mathbf{x}_d = \mathbf{x} - \mathbf{x}_m$, and the residual vector as $\mathbf{x}_r = (\mathbf{x} - \mathbf{x}_m) - \mathbf{U}\mathbf{x}_s$. Let $\mathbf{x}_{ru}$ be the normalized unit vector of $\mathbf{x}_r$. Through straight algebraic manipulations we can arrive at the adapted eigenanalysis results $\{\mathbf{x}_m^{new}, \Lambda^{new}, \mathbf{U}^{new}\}$ with $\mathbf{U}^{new} = [\mathbf{U}, \mathbf{x}_{ru}]\mathbf{R}$, where $\mathbf{R}$ and Λ^{new} are solutions to the following eigenanalysis problem:

$$\left(\alpha \begin{bmatrix} \Lambda & \mathbf{0} \\ \mathbf{0}^T & 0 \end{bmatrix} + \alpha(1-\alpha) \begin{bmatrix} \mathbf{x}_s\mathbf{x}_s^T & e_r\mathbf{x}_s \\ e_r\mathbf{x}_s^T & e_r^2 \end{bmatrix}\right)\mathbf{R} = \mathbf{R}\Lambda^{new} \tag{17}$$

where $e_r = \mathbf{x}_{ru}^T(\mathbf{x} - \mathbf{x}_m)$ is the residual energy.

The above formulas are extensions of IPCA or eigenspace merging formula of [18], with tunable energy ratios between the new data and the old data. With α set at a smaller value (we use 0.5), the PCA model is strongly adapted toward the current case, hence the name. Now we are ready to point out the differences between fusion and IPCA or SA-PCA. First of all, a fused model cannot break out of the subspace, while IPCA or SA-PCA can. More fundamentally, fusion provides the "intersection" of the information sources [19], while IPCA or SA-PCA yield some "union" of the sources. We need to *augment* instead of *constrain* the generic model, so fusion is not the proper choice.

With SA-PCA, our framework now incorporates four information sources: the system dynamic, measurement, subspace model, and the initial contour. This last addition is especially useful for periodic shape deformations such as cardiac motion.

4 Implementation, Evaluation, and Analysis

In this paper we test the proposed framework using ultrasound heart sequences. Ultrasound is the noisiest among common medical imaging modalities such as MRI or CT. Echocardiogram (ultrasound heart images) is even worse due to the fast motion of the heart muscle and respiratory interferences [6]. With spatially varying noise characteristics, echocardiograms are ideal for testing our heteroscedastic fusion framework.

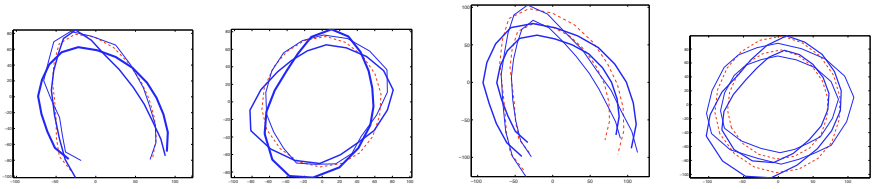


Fig. 2. The dominant eigenshapes for: (a,b) single contour; (c,d) coupled contours; (a,c) Apical views; (b,d) Short axis views. The dashed curves are the model mean.

4.1 Tracking in Echocardiography

We use manually traced left ventricle borders in echocardiography images as the training set. Both apical two- or four-chamber views and parasternal long and short axis views are trained and tested. Landmark points are assigned based on anatomic features (apex, papillary muscles, and septum, etc.). The algorithm can tolerate some variability on the location of the landmark points, partly due to the application of SA-PCA.

The training contours are aligned using the iterative Procrustes analysis approach described by Cootes and Taylor [20] to cancel out global translation, rotation and scaling. PCA is then performed and the original dimensionality is reduced to retain 80-97% of energy, separately tuned for each model. Figure 2 shows the dominant eigenshapes (without splining) for two views along with their model means trained on about 200 contours each for both single and double-contours. A double-contour is treated as a single point in a high-dimensional space.

During testing we assume manual initialization on the first frame, and use a simple dynamic model to impose a temporal smoothness constraint. Without prior knowledge, we employ a diagonal matrix to model the uncertainty in system dynamics, and set the relative confidence of this model empirically. Since we perform alignment on the training shapes before the PCA, at each tracking step the model is aligned to the fusion result $\{\tilde{\mathbf{x}}, \tilde{\mathbf{C}}\}$ using the measurement and the dynamic model. We adopt the optimal transformation $\mathcal{T}_o$ which minimizes a weighted sum-of-squares measure of point difference subject to translation, rotation and scaling ([20], p. 102), with the weighting matrix being $\tilde{\mathbf{C}}^{-1}$. The system transforms the model mean as well as *the model covariance* using $\mathcal{T}_o$ before the final fusion.

4.2 Motion Estimation with Uncertainty

To measure the motion of each of the control points we use an adaptation of the frame-to-frame motion estimation algorithm described in [13], which has been shown to be very competitive in terms of performance evaluation using standard sequences. We present in the sequel a summary of the algorithm. For more details, see [13].

The main idea is that the *motion in a certain neighborhood can be robustly estimated as the most significant mode of some initial motion estimates* (expressed by mean vectors and associated covariance matrices). The most significant mode is defined by mode

Fig. 3. The 95% confidence ellipses corresponding to the local measurement uncertainty on each control point. The first image shows a single contour for endocardium. The other two show coupled double contours for both endocardium and epicardium.

tracking across scales, while the underlying mechanism for mode detection relies on the variable-bandwidth mean shift [21].

In the current work, for each control point we compute initial estimates using 17×17 windows and fuse the results on $n = 5 \times 5$ neighborhoods. A pyramid of three levels is employed with covariance propagation across levels. Figure 3 depicts the uncertainty calculated at the bottom of the pyramid for the contour points.

To avoid error accumulation from frame to frame, the motion is always computed with reference to the neighborhood of the control point in the first frame of the sequence (i.e., the current frame is always compared to a model extracted from the first frame). Since we update the location of the model at each frame, the motion estimation process always starts with a good initialization, hence the error accumulation is canceled. The overall procedure is suitable for the tracking of periodic sequences such as the heart ultrasound data. It resembles to a template-based tracker, which benefits from the fast computation of frame-to-frame motion.

4.3 Performance Evaluation and Analysis

For systematic evaluation, a set of 30 echocardiogram sequences are used for testing, including parasternal long- and short-axis views and apical two- or four-chamber (AC) views, all with expert-annotated ground-truth contours.

We use two distance measures: the Mean Sum of Squared Distance (MSSD) [22] and the Mean Absolute Distance (MAD) [23]. With consistent results, we report MSSD only in this paper. For the sequence S_i with m frames, $\{c_1, ..., c_m\}$, where each contour c_j has n points $\{(x_{j,1}, y_{j,1}),\ldots,(x_{j,n}, y_{j,n})\}$, the distances to the ground truth S_i^0 are

$$MSSD_i = \frac{1}{m}\sum_{j=1}^{m} MSSD_{i,j} = \frac{1}{m}\sum_{j=1}^{m}\frac{1}{n}\sum_{k=1}^{n}((x_{j,k} - x_{j,k}^0)^2 + (y_{j,k} - y_{j,k}^0)^2) \quad (18)$$

The overall performance measure for a particular method is the averaged distance on the whole test set of l sequences. A critical difference between our distance measures and those of [22] or [23] is that we have the point correspondence through tracking. As a

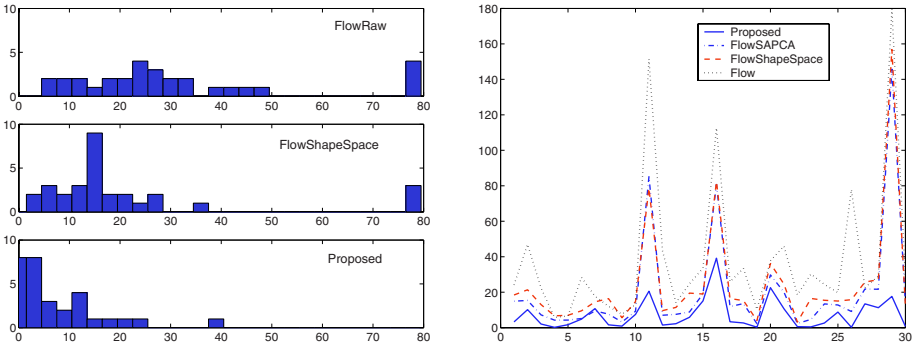


Fig. 4. (a) MSSD histograms over the test set; (b) MSSD curves for the 30 test sequences.

result, we could capture tangent motion components along the contour which can reveal crucial information about cardiac function.

Our proposed framework is compared to three alternatives. The first is a tracker based on the optical flow algorithm without shape constraint ("FlowRaw") [13]. The second approach is the same tracker but adding orthogonal PCA shape space constraints [8,20, 7]("FlowShapeSpace"). The third is "FlowShapeSpace" but using our SA-PCA model ("FlowSAPCA") Figure 4 shows the comparison of these methods. Our proposed method ("Proposed") significantly outperforms others, with an average MSSD of 7.4 ($\sigma = 12.3$) as opposed to 24.3 ($\sigma = 35.4$) by the current approach ("FlowShapeSpace"). Our SA-PCA model alone ("FlowSAPCA") already brought significant improvement, achieving an average MSSD of 20.4 ($\sigma = 32.8$). The fusion alone (without SA-PCA) had an average MSSD of 18.4 ($\sigma = 22.7$). The MSSD of "FlowRaw" is 38.2 ($\sigma = 83.7$). The combined use of fusion and SA-PCA (i.e., "Proposed") has apparently brought out a significant performance boost over each alone. Figure 5 shows some tracked sequences. Please also refer to the supplementary videos.

When the measurement process makes a large error in a drop-out or high-noise region, the corresponding localization uncertainty is usually high as well, due to the lack of trackable patterns. Our fusion can correct such errors to a larger extent than what an orthogonal projection can do. This is illustrated by an example in Figure 6.

Our SA-PCA model is especially helpful for shapes that differ significantly from the training set. Figure 7 shows a comparison of IPCA and SA-PCA. In this example, we deliberately used a "wrong" model, i.e., we use the model for apical four chamber (A4C) views (see Figure 2a) to constrain the tracking of this parasternal long axis (PLA) view. PLA views have distinctive patterns that are not seen in apical views (e.g., the upper concave portion). The incremental PCA model, taking in the initial contour (Figure 7a) but with a very small weight ($< 0.01\%$), fails to follow such distinctive patterns; and has constrained the contours to a typical A4C shape (Figure 7b). SA-PCA yields a contour that fits much better to the true border (Figure 7c).

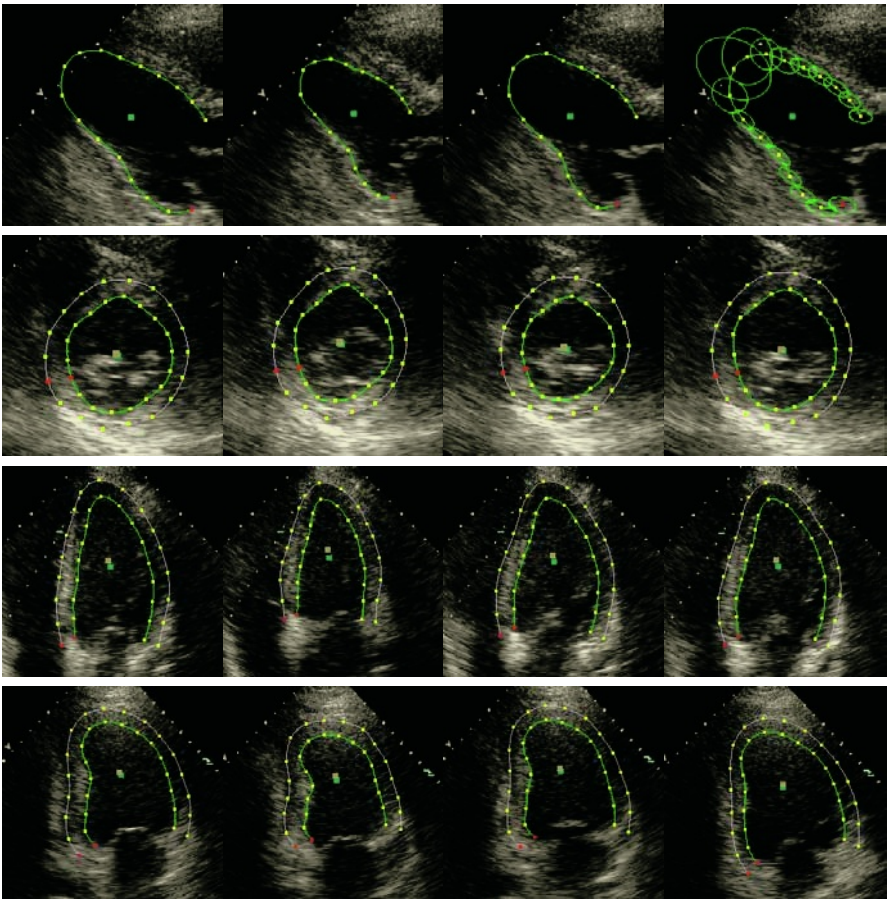


Fig. 5. Four tracking examples in rows, with 4 snapshots per sequence. The frame numbers are 1, 16, 24, and 32; 1, 12, 67, and 81; 1, 12, 18 and 23; 1, 15, 23, and 30; respectively.

4.4 Double-Contour versus Single Contour

Although harder to train a model (requiring more data), coupled double-contours have some advantages: A double-contour approach integrates more spatial information, thus can provide more robust tracking of the two borders. In many cases epicardium is less visible than endocardium (except for the case of pericardial effusion for which the opposite is true!), a double-contour can propagate information from the endocardium to guide the localization of the epicardium (or vice versa). Furthermore, a double-contour can better preserve topology and reduce the chance of crossing (assuming no crossing in the training set). With our explicit constraint from the model distribution using Eq. (9), we limit not only the mode but also the range of shape deformations. Figure 8 shows an example where the double-contour approach clearly improves the performances by single contours alone. Notice the complete appearance change on the right, along with

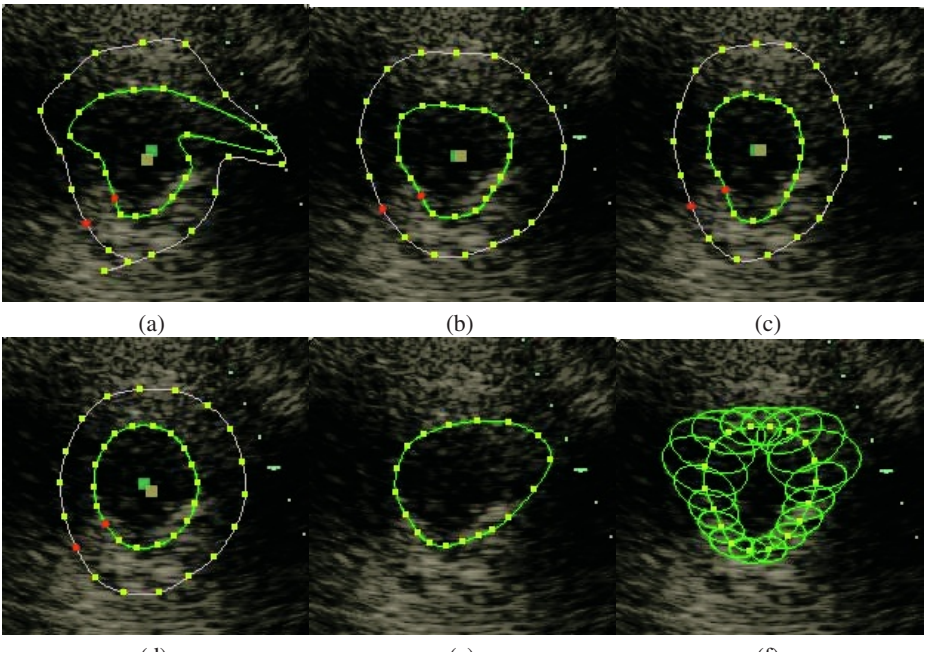


Fig. 6. Comparison of orthogonal projection with our fusion approach in handling large measurement errors. On the same frame we show: (a) the un-constrained flow results; (b) orthogonal projection into the SA-PCA space; (c) result from our fusion framework; (d) by an expert; (e) same as (b) but only for endocardial border; (f) same as (c) for endocardial border (also shown are the uncertainty ellipses). Notice the stronger correction the fusion method brought over the local measurement errors for both single and double contours.

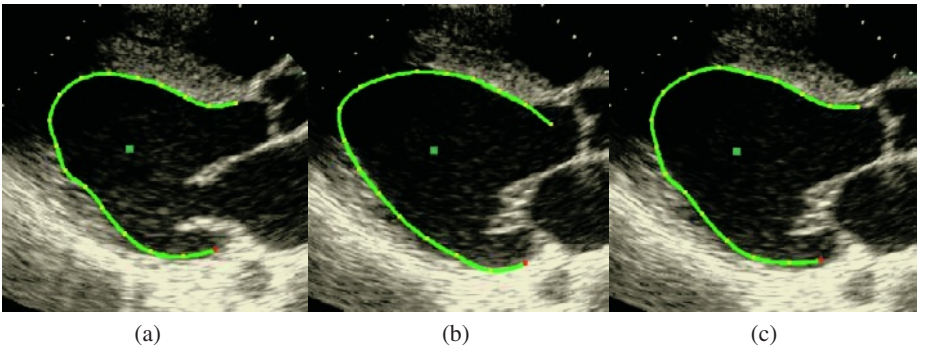
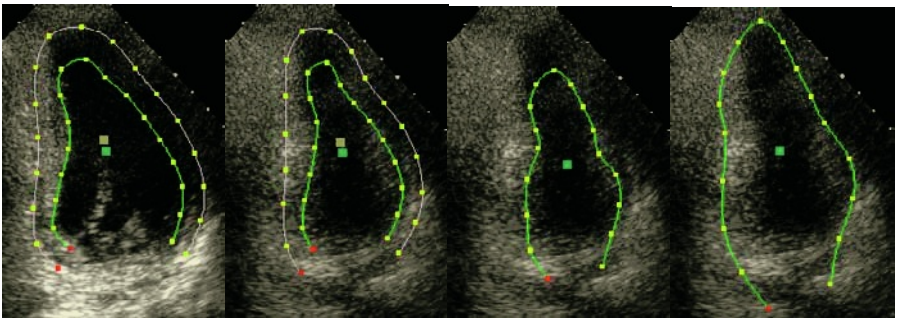


Fig. 7. SA-PCA versus incremental PCA. (a) the initial contour; (b) the 14th frame using an incremental PCA model [18]; (c) the same frame using an SA-PCA model ($\alpha = 0.5$).



(a) (b) (c) (d)

Fig. 8. Double versus single contours. (a) the initial contours; (b) 7th frame tracked by the double-contour; (c) 7th frame with the inner contour; (d) 7th frame with the outer contour.

the large intensity shift at the base, which is in large part to blame for the errors made by the single contours. The double-contour combines information from all locations plus a double-contour model to make the decisions for both contours jointly, thus achieving more robust performance. Our tracker runs at about 30 fps on a 3GHz PC.

5 Related Work

With heteroscedastic measurement noise, an orthogonal projection into the model subspace is not only unjustified, but also damaging in terms of information loss [24]. It can only be justified when the noise is isotropic and homogeneous (*cf.* [20,7,8,2]).

Measurement uncertainty has been exploited for tracking and motion estimation in different contexts. However, none has put all relevant information sources into a unified fusion formulation. Both Brand [24] and Irani [25] use measurement uncertainties, but they did not provide a complete fusion-based tracking framework that combines all the information sources. A rank-constrained flow estimation formulation was proposed by Bregler et al. [26]. They use constraints from both rigid and non-rigid motion represented by basis-shapes. Although occlusion is addressed, measurement uncertainty in general is not optimally exploited. Leedan, Matei, and Meer (e.g. [3]) applied heteroscedastic regression for fitting ellipses and fundamental matrices. The fitting is achieved in the original space with parameterized models. In our formulation, we avoid the parameterization of shape variations. Instead, we build subspace probabilistic models through PCA and obtain closed-form solutions on both the mean and covariance of the fitted data. Although simple, this model proves to be flexible and powerful for the current application, especially with the use of an SA-PCA model. Nevertheless, for applications where this model is too restrictive, a future research is to integrate more sophisticated nonlinear models. (e.g., [27]). Robust model matching [28] relying on M-estimators or RANSAC has been applied to limit or eliminate the influence of data components that are outliers with respect to the model. Again, the locally (in space or time) varying uncertainties are not exploited in these frameworks.

There is much research work done in medical domain that tracks heart motion using various techniques (e.g., [8,23,22], etc.). However, none of these addresses the issue of heteroscedastic noise and its fusion with other information sources.

6 Conclusions and Future Work

This paper presented a joint information fusion framework to track shapes under heteroscedastic noise with a strongly adapted subspace model constraint.

Extensions to 3D or 2D+T(time) are natural. Extension to triple(or more)-contours are feasible if sufficient training data are available. Our framework is general and can be applied to other applications. Considering the heavy noise situation in the ultrasound data, success of this approach on other data is expected in our future efforts.

Acknowledgment. We would like to thank Visvanathan Ramesh from Siemens Corporate Research for insightful discussions on the subject. We are grateful to Alok Gupta from the CAD Group, Siemens Medical Solutions for his support and encouragement.

References

1. Cootes, T., Taylor, C.: Active shape models-'smart snakes'. In: Proc. British Machine Vision Conference. (1992) 266–275
2. Turk, M.A., Pentland, A.P.: Face recognition using eigen-face. In: Proc. IEEE Conf. on Computer Vision and Pattern Recognition, Hawaii. (1991) 586–591
3. Leedan, Y., Meer, P.: Heteroscedastic regression in computer vision: Problems with bilinear constraint. Intl. J. of Computer Vision **37** (2000) 127–150
4. Kanazawa, Y., Kanatani, K.: Do we really have to consider covariance matrices for image features? In: Proc. Intl. Conf. on Computer Vision, Vancouver, Canada. Volume II. (2001) 586–591
5. Irani, M., Anandan, P.: Factorization with uncertainty. In: Proc. 6th European Conf. on Computer Vision, Dublin, Ireland. (2000) 539–553
6. Oh, J.K., Seward, J.B., Tajik, A.J.: The Echo Manual. Lippincott Williams & Wilkins, Philadelphia, (1999)
7. Blake, A., Isard, M.: Active contours. Springer Verlag (1998)
8. Jacob, G., Noble, A., Blake, A.: Robust contour tracking in echocardiographic sequence. In: Proc. Intl. Conf. on Computer Vision, Bombay, India. (1998) 408–413
9. Paragios, N.: A variational approach for the segmentation of the left ventricle in cardiac images. In: Proc. IEEE Workshop on Variational and Level Set Methods in Computer Vision, Vancouver, Canada. (2001)
10. Goldenberg, R., Kimmel, R., Rivlin, E., Rudzsky, M.: Cortex segmentation: A fast variational geometric approach. IEEE Trans. Medical Imaging **21** (2002) 1544–51
11. Wang, S., Ji, X., Liang, Z.P.: Landmark-based shape deformation with topology-preserving constraints. In: Proc. Intl. Conf. on Computer Vision, Nice, France. (2003)
12. Black, M., Jepson, A.: Eigentracking: Robust matching and tracking of articulated objects using a view-based representation. In: Proc. European Conf. on Computer Vision, Cambridge, UK. (1996) 610–619

13. Comaniciu, D.: Nonparametric information fusion for motion estimation. In: Proc. IEEE Conf. on Computer Vision and Pattern Recognition, Madison, Wisconsin. Volume I. (2003) 59–66
14. Scharf, L.L.: Statistical Signal Processing. Addison Wesley, Reading, MA (1991)
15. Bar-Shalom, Y., Campo, L.: The effect of the common process noise on the two-sensor fused track covariance. IEEE Trans. Aero. Elect. Syst. **AES-22** (1986) 803–805
16. Li, X., Zhu, Y., Han, C.: Unified optimal linear estimation fusion - part i: Unified models and fusion rules. In: Proc. of 3rd Intl. Conf. on Information Fusion, Paris, France. (2000) MoC2–10–MoC2–17
17. Anderson, B., Moore, J.: Optimal filtering. Prentice-Hall (1979)
18. Hall, P., Marshall, D., Martin, R.: Merging and splitting eigenspace models. IEEE Trans. Pattern Anal. Machine Intell. **22** (2000) 1042–1048
19. Julier, S., Uhlmannn, J.: A non-divergent estimation algorithm in the presence of unknown correlations. In: Proc. American Control Conf., Alberqueque, NM. (1997)
20. Cootes, T., Taylor, C.: Statistical models for appearance for computer vision. (2001) Unpublished manuscript. Available at `http://www.wiau.man.ac.uk/~bim/Models/app_model.ps.gz`.
21. Comaniciu, D.: An algorithm for data-driven bandwidth selection. IEEE Trans. Pattern Anal. Machine Intell. **25** (2003) 281–288
22. Akgul, Y., Kambhamettu, C.: A coarse-to-fine deformable contour optimization framework. IEEE Trans. Pattern Anal. Machine Intell. **25** (2003) 174–186
23. 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
24. Brand, M., Bhotika, R.: Flexible flow for 3D nonrigid object tracking and shape recovery. In: Proc. IEEE Conf. on Computer Vision and Pattern Recognition, Hawaii. Volume I. (2001) 315–322
25. Irani, M.: Multi-frame optical flow estimation using subspace constraints. In: Proc. Intl. Conf. on Computer Vision, Kerkyra, Greece. (1999) 626–633
26. Bregler, C., Hertzmann, A., Biermann, H.: Recovering non-rigid 3d shape from image streams. In: Proc. IEEE Conf. on Computer Vision and Pattern Recognition, Hilton Head, SC. Volume II. (2000) 690–696
27. Cremers, D., Kohlberger, T., Schnorr, C.: Nonlinear shape statistics in mumford-shah based segmentation. In: Proc. European Conf. on Computer Vision, Copenhagen, Denmark. (2002) 93–108
28. Black, M., Ananadan, P.: The robust estimation of multiple motions: Parametric and piecewise-smooth flow fields. Computer Vision and Image Understanding **63** (1996) 75–104