An Information Fusion Framework for Robust Shape Tracking

Xiang Sean Zhou, *Member*, *IEEE*, Dorin Comaniciu, *Senior Member*, *IEEE*, and Alok Gupta

Abstract—Existing methods for incorporating subspace model constraints in shape tracking use only partial information from the measurements and model distribution. We propose a unified framework for robust shape tracking, optimally fusing heteroscedastic uncertainties or noise from measurement, system dynamics, and a subspace model. The resulting *nonorthogonal subspace projection* and *fusion* are natural extensions of the traditional model constraint using orthogonal projection. We present two motion measurement algorithms and introduce alternative solutions for measurement uncertainty estimation. We build shape models offline from training data and exploit information from the ground truth initialization online through a strong model adaptation. Our framework is applied for tracking in echocardiograms where the motion estimation errors are *heteroscedastic* in nature, each heart has a distinct shape, and the relative motions of epicardial and endocardial borders reveal crucial diagnostic features. The proposed method significantly outperforms the existing shape-space-constrained tracking algorithm. Due to the complete treatment of heteroscedastic uncertainties, the strong model adaptation, and the coupled tracking of double-contours, robust performance is observed even on the most challenging cases.

Index Terms—Shape tracking, subspace constraint, motion estimation with uncertainty, heteroscedastic noise, active shape model, model adaptation.

1 Introduction

Model constraints can significantly improve the performance of a shape 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 [18]. A traditional treatment is to project into a PCA (Principal Component Analysis) subspace [54], [18]. However, this approach does not take advantage of heteroscedastic (i.e., both anisotropic and inhomogeneous) measurement uncertainties [33], [31] (see Fig. 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 model constraint into a Kalman filter-based tracker. In [9], [29], the proposal was to set the system noise covariance matrix to be the covariance of a PCA shape model. However, it did not provide a systematic and complete fusion of the model information because the model mean is discarded (as a result, the projection can be arbitrarily far from the model mean in the subspace). More importantly, 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 applying a statistical shape model to control the modes and range of shape variations. In addition, existing solutions do not specifically address the issue of heteroscedastic measurement uncertainties and its influence during the fusion with other information sources. When measurement errors are inhomogeneous and anisotropic, 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 uncertainties*. We build models for coupled double-contours so that more information can be integrated, especially for very noisy data [58]. The double-contour also achieves better preservation of topology. To accommodate individual shape characteristics, the shape model learned from a collection of training cases 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 [54], [8]. 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 [16].

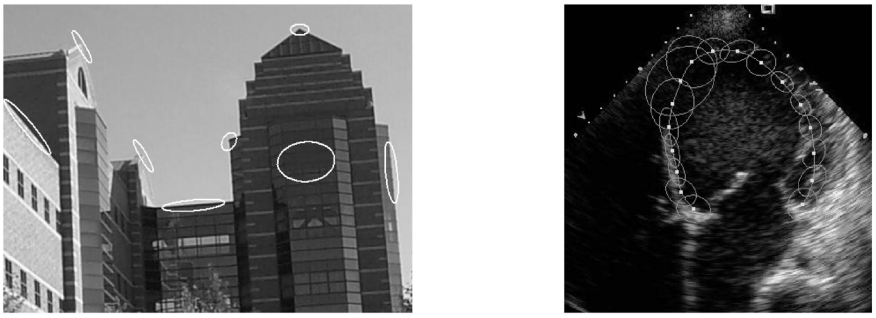
By *shape tracking* instead of *contour tracking*, we stress that we detect actual control point movement (instead of movement on a perpendicular line) and maintain point labels throughout tracking. This facilitates alignment and structural analysis and tangential motion estimation.

1.1 Related Work

When the measurement errors are heteroscedastic, an orthogonal projection into the model subspace is not only unjustified, but also damaging in terms of information loss

- *X.S. Zhou and D. Comaniciu are with the Integrated Data Systems Department, Siemens Corporate Research, 755 College Road East, Princeton, NJ 08540.*
  E-Mail: {Xiang.Zhou, dorin.comaniciu}@siemens.com.
- *A. Gupta is with the Computer Aided Diagnosis Group, Siemens Medical Solutions, 51 Valley Stream Parkway, Malvern, PA 19355.*
  E-Mail: alok.gupta@siemens.com.

Manuscript received 25 Jan. 2004; revised 16 May 2004; accepted 2 June 2004. Recommended for acceptance by S. Sclaroff.
For information on obtaining reprints of this article, please send e-mail to: tpami@computer.org, and reference IEEECS Log Number TPAMI-0049-0104.



(a) (b)

Fig. 1. Ellipses depicting uncertainties or confidence regions in feature localization and motion estimation. The heteroscedastic nature stems from either the *aperture problem* [31], [28], or for echocardiograms (b), the *acoustic drop-out* [41].

[10]. It can only be justified when the error is isotropic and homogeneous (cf. [19], [9], [29], [54]).

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. Brand and Bhotika [10] and Irani [27] 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. [11]. They used constraints from both rigid and nonrigid motion represented by basis-shapes. Although occlusion is addressed, measurement uncertainty in general is not optimally exploited. Leedan and Meer (e.g., [33]) 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 effective for the current application, especially with the use of our proposed SA-PCA model (see Section 3). Nevertheless, for applications where this model is too restrictive, a future research is to integrate more sophisticated nonlinear kinematic models (e.g., [43]). Robust model matching [7], [44] 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 the medical domain that tracks heart motion using various techniques (e.g., [29], [38], [1], [13], [36], [52], [37], etc.). These approaches explored various techniques such as optical flow, active contours, PCA shape constraints, joint appearance and shape models, Kalman filters, etc. However, none of these addressed the issue of heteroscedastic noise and its fusion with other information sources.

The paper is organized as follows: The new tracking formulation using fusion is presented in Section 2. Section 3 introduces our proposed model adaptation scheme. Section 4 discusses the issues related to local motion measurement and uncertainty estimation. Section 5 addresses the problem of uncertainty handling during shape alignment. Experimental evaluation and analysis are presented in Section 6. Conclusions are drawn in Section 7, with a discussion of future research directions.

2 Fusion-Based Subspace Model Constraining during Tracking

Throughout this paper, we represent shapes by configurations of labeled control points or landmark points, assuming correspondence. For displaying on the user interface, these points are fitted by splines into one or more contours (see Fig. 12). During analysis, however, point coordinates for each shape (with one or more contours) are concatenated into one vector [18]. For more implementation details, please refer to Section 6.

A typical tracking framework fuses information from 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 nonorthogonal projection. Then, we show that, with a complete subspace model constraint, this can be further generalized into an information fusion formulation. Finally, these formulas are uniformly combined into the tracking framework.

2.1 Nonorthogonal Projection for Heteroscedastic Uncertainties

Unless otherwise noted, bold lower-case letters are used to represent vectors, bold upper-case letters are for matrices, and italic letters represent scalars.

Given an n-dimensional measurement point[1] represented by a multivariate normal distribution with a mean vector $\mathbf{x}$ and 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., $\mathbf{y}^* = argmin\ d^2$, where

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

This is in the form of a *weighted least square* ([45, p. 386]). By taking a derivative of the above with respect to $\mathbf{y}$ and setting it to 0, we have

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

1. 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," we refer to the inhomogeneity among different *control points* [31].

In general, this is a *nonorthogonal 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 scalar c, we have

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

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 (or the mean and covariance, in case of Gaussian) 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 basic idea is to treat both the *measurement* and the *shape model* (and, later, the *prediction* as well) as noisy measurements with covariance matrices and fuse all information in an optimal way.

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, ${D_i}^2(\mathbf{x}) = (\mathbf{x} - \mathbf{x}_i)^T\mathbf{C}_i^{-1}(\mathbf{x} - \mathbf{x}_i)$, to the two centroids, i.e., $\mathbf{x}^* = argmin\ d^2$ with

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

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

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

which is also known as the best linear unbiased estimate (BLUE) of $\mathbf{x}$, assuming that the two sources are independent ([4], [34]). When dependency or correlation exists, one possible alternative is to use *covariance intersection* which can avoid overly confident estimate [30].

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

$${D_2}^2(\mathbf{x}) = \sum_{i=1}^{p} \lambda_i^{-1}[\mathbf{U}_p^T(\mathbf{x} - \mathbf{x}_2)]^2 \equiv (\mathbf{x} - \mathbf{x}_2)^T\mathbf{C}_2^+(\mathbf{x} - \mathbf{x}_2) \quad (6)$$

with the additional constraint of $\mathbf{U}_0^T\mathbf{x} = 0$ (otherwise, d will diverge). Here, $\mathbf{C}_2 = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^T$, $\mathbf{U} = [\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_n]$, $\mathbf{U}_p = [\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_p]$, $\mathbf{U}_0 = [\mathbf{u}_{p+1}, \mathbf{u}_{p+2}, \ldots, \mathbf{u}_n]$, with $\mathbf{u}_i$'s orthonormal and $\mathbf{\Lambda} = diag\{\lambda_1, \lambda_2, \ldots, \lambda_p, 0, \ldots, 0\}$. (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 $\mathbf{y} = \mathbf{U}_p^T\mathbf{x}$. Equation (4) now takes the following general form:

$$\begin{aligned} d^2 = &(\mathbf{U}_p\mathbf{y} - \mathbf{x}_1)^T\mathbf{C}_1^{-1}(\mathbf{U}_p\mathbf{y} - \mathbf{x}_1) \\ &+ (\mathbf{U}_p\mathbf{y} - \mathbf{x}_2)^T\mathbf{C}_2^+(\mathbf{U}_p\mathbf{y} - \mathbf{x}_2). \end{aligned} \quad (7)$$

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

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

Equivalent expressions can be obtained in the original space as:

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

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 (8) as

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

Here, $\mathbf{y}_2 = \mathbf{U}_p^T\mathbf{x}_2$, and $\mathbf{\Lambda}_p = diag\{\lambda_1, \lambda_2, \ldots, \lambda_p\}$. Interestingly, (10) is, in fact, the BLUE fusion of two subspace Gaussian distributions, one being $\mathcal{N}(\mathbf{y}_2, \mathbf{\Lambda}_p)$ and the other being the *nonorthogonal 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. (2)).

The fusion formulation ((8) through (10)) is general and it takes as special cases the nonorthogonal projection of (2) (when the distribution in the subspace is ignored, leaving just a subspace as the constraint) and the traditional approach of orthogonal projection of (3) (when considering only the subspace and an isotropic $\mathbf{C}_1$). In this general formulation, the model constraint can be interpreted as an independent "virtual measurement" and can be treated interchangeably with the true measurement.

2.3 Constrained Tracking through Fusion and Projection

To integrate the above projection and fusion formulas into a tracking framework, we first note that the Kalman filter is essentially *fusion* in nature, which is evident in its information filter form ([3, p. 138]):

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

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 uncertainty 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 $\mathbf{Q}$ ([5, p. 56]).

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

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

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 the Kalman filter is a fusion filter and the information fusion operation is *associative*, we can apply the subspace fusion formulas, (8) and (9), on the Kalman fusion result of (11) (in general, $\mathbf{H}$ is not square nor invertible; otherwise, (12) can be used) and a subspace source $\mathcal{N}(\mathbf{x}_2, \mathbf{C}_2)$, to obtain a complete fusion formula:

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

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

It is worth pointing out that this solution *combines all the available knowledge in the information space*. It provides a unified fusion of the system dynamics, a subspace model, and measurement uncertainty. It addresses, in a principled way, uncertainties from various sources that affect the tracking system.

Compared to a PCA *shape space* representation [9], [29], the fusion formulation takes into account not only the model subspace (the eigenvectors), but also the actual model distribution, in a unified framework. On the other hand, if only a subspace constraint is desired, we can simply apply the special case of (2) on (11) and the resulting *nonorthogonal 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 shapes 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 discussions on *fusion*, a tempting idea would be *to fuse the model and the new contour* by assigning some covariance $\mathbf{C} = \alpha\mathbf{I}$ for the new contour. This turns out to be inappropriate and we will discuss this at the end of this section.

An alternative tool is incremental PCA (IPCA) [22], where a PCA model can be updated without using previous training data. However, to effectively exploit the prior, one needs to put more emphasis on the initial contour. Therefore, we apply a *strongly adapted-PCA* (SA-PCA) model as follows: We assume that the existing PCA model and the initial contour of the current case *jointly* represent the variations of the current case, but with relative energy being α and $(1-\alpha)$, $0 < \alpha < 1$. In other words, a portion of the shape variations of the current case is captured by the generic model, while the rest is captured *in the direction of* the initial contour in the model space.

The PCA model has its mean, eigenvalue matrix, and eigenvector matrix, denoted by $\mathbf{x}_m$, $\mathbf{\Lambda}$, and $\mathbf{U}$, respectively. If the original covariance matrix $\mathbf{C}$ were stored, the adapted mean $\mathbf{x}_m^{new}$ and covariance matrix $\mathbf{C}^{new}$ would simply be the weighted sum of the two contributing sources:

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

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

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

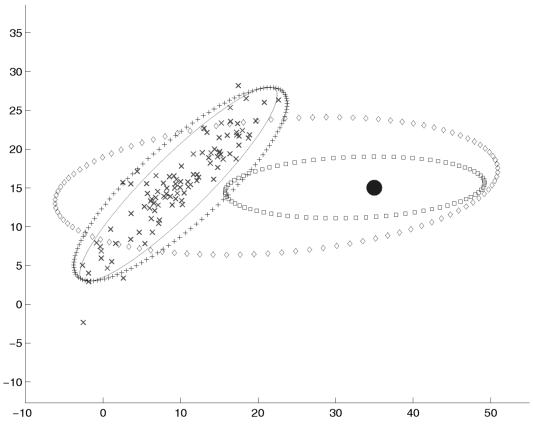


Fig. 2. SA-PCA versus IPCA. The dark crosses ("x") are 99 training points, the black dot is the new point. The solid ellipse shows the original model distribution, the black "+"-ellipse is IPCA (equivalent to SA-PCA with $\alpha = 0.99$), the diamond-ellipse is SA-PCA with $\alpha = 0.5$, and the square-ellipse corresponds to $\alpha = 0.1$. Each ellipse depicts the 90 percent equal-probable contour of the corresponding distribution.

A more interesting and practical scenario is when $\mathbf{C}$ is not stored and $\{\mathbf{x}_m, \mathbf{\Lambda}, \mathbf{U}\}$ 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 $\{\mathbf{x}_m^{new}, \mathbf{\Lambda}^{new}, \mathbf{U}^{new}\}$ with $\mathbf{U}^{new} = [\mathbf{U}, \mathbf{x}_{ru}]\mathbf{R}$, where $\mathbf{R}$ and $\mathbf{\Lambda}^{new}$ are solutions to the following eigenanalysis problem:

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

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

The above formulas are extensions of IPCA or the eigenspace merging formula of [22], with tunable energy ratios between the new data and the old data. With α set at a smaller value, the PCA model is strongly adapted toward the current case, hence the name. Fig. 2 shows a simple two-dimensional illustration of IPCA and SA-PCA with different α values. One open issue for SA-PCA is how to set the value for α. Based on our experiments, we observe that optimal performance is reached when α is between 0.3 and 0.6 [57]. In this work, we set it at 0.5.

We are now in a position 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 [30], 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 Estimating Local Motion and Uncertainty

In this section, we first review existing research on the estimation of measurement or feature uncertainties,

particularly *heteroscedastic* (i.e., location-dependent and anisotropic) analysis. We then discuss our selection of matching algorithms and the corresponding formulations for uncertainty estimation.

4.1 Measurement Uncertainty Estimation

Motion estimation uncertainty is jointly influenced by three factors: 1) *the noise characteristics and levels*, 2) *the intrinsic* tractability *of the target pattern*. A typical example is the so-called "aperture problem": Points on an edge are intrinsically harder to track than a point sitting on a corner. This factor is intrinsic because it exists regardless of what algorithms to use, and 3) *the complexity of the appearance change with respect to the capability of the estimator*. For example, affine deformations of the target would result in low confidence estimates if a tracker assumed only translational motion [48]. Even for a good feature such as a corner or a patch with "high texture content" [48], its motion uncertainty can be high if the feature is significantly altered or lost in the second image.

Under some limiting assumptions, the covariances of the motion estimates can be obtained analytically. For example, with the assumptions of "conservation of image intensity" and Gaussian noise, the gradient-based optical flow calculated by least squares has a covariance matrix

$$\mathbf{C} = \sigma^2 \mathbf{G}^{-1}, \tag{18}$$

i.e., with its spread proportional to the noise variance σ^2 and its *anisotropy* or directionality governed by the inverse of the gradient Grammian $\mathbf{G}$ ([25], [45], [53, p. 378]), where

$$\mathbf{G} = \sum_{(x_i,y_j)\in W} \begin{bmatrix} I_x(x_i,y_j)^2 & I_x(x_i,y_j)I_y(x_i,y_j) \\ I_x(x_i,y_j)I_y(x_i,y_j) & I_y(x_i,y_j)^2 \end{bmatrix}. \tag{19}$$

Here, I_x and I_y are the spatial gradients of the image function I, and (x_i, y_j) are the points used to estimate the flow. Not surprisingly, the same matrix $\mathbf{G}$ is used in other contexts such as *interest point detection*, where an interest point is declared if this matrix has two significant eigenvalues [46], point feature detection and monitoring during tracking [48], and other computer vision problems such as homography or fundamental matrix computation [31].

Note here that the uncertainty depends only on the intrinsic property of the target (by the Grammian) and the noise and not on the actual matching candidate image, which has been assumed to be simply a shifted version of the target in noise. The noise, however, can be divided into two types and modeled explicitly, one for errors in temporal derivative measurements and the other for variations in spatial gradients. If both are Gaussian, a closed-form solution can be obtained for the distribution of the flow [49].

However, motion uncertainty is not easy to formulate when complex, non-Gaussian appearance variations occur in the second image, such as affine or nonlinear distortions, partial occlusion/disocclusion, and signal drop-outs, as often occur in echocardiography [41], etc. In an early proposal, Anandan [2] used multiple sum-of-squared-differences (SSD) matching scores $SSD(x, y)$ around the best match to estimate matching confidence. The two principal curvatures of the SSD surface were estimated to represent the anisotropic matching confidence. This scheme was revised subsequently by, e.g., Singh [50] and Nickels and Hutchinson [40], where a *response distribution*, $RD(x, y)$, is used instead of the original SSD map:

$$RD(x,y) = \exp(k\, SSD(x,y)), \tag{20}$$

where the factor k is used to normalize RD, e.g., to sum to 1 [40].

The use of SSD and RD are both justifiable under certain assumptions, but one has to use the right formulation for the covariance estimate. The guideline for choosing an uncertainty measure is that it should conform to (18) under the same assumptions. Indeed, under the assumptions that led to (18), the SSD surface $SSD(x, y)$ can be approximated using a Taylor expansion around the best-match location (which is set at $(0, 0)$):

$$\begin{aligned} SSD(x,y) &= \sum_{i,j} (I(i+x, j+y) - I(i,j))^2 \\ &\approx \sum_{i,j} (I(i,j) + I_x x + I_y y - I(i,j))^2 \\ &= (\,x \quad y\,)\mathbf{G}\begin{pmatrix} x \\ y \end{pmatrix}, \end{aligned} \tag{21}$$

where $\mathbf{G}$ has the same form as (19).

This indicates that the motion covariance can be estimated, up to scale, through a *quadratic fitting* of $SSD(x, y)$ (cf. [31]).

Combining (20) and (21), it becomes apparent that the $RD(x, y)$, when approximated by a Gaussian distribution, has a covariance that is proportional to that of (18). In other words, motion covariance can be directly estimated, up to scale, by the covariance of $RD(x, y)$ around the best-match location.

Given $RD(x, y)$, assuming additive zero-mean independent errors, the covariance matrix is [50], [40]:

$$\mathbf{C} = \frac{1}{\sum_{i,j} RD(x_i,y_j)} \sum_{i,j} \begin{bmatrix} RD(x_i,y_j)(x_i-x)^2 & RD(x_i,y_j)(x_i-x)(y_j-y) \\ RD(x_i,y_j)(x_i-x)(y_j-y) & RD(x_i,y_j)(y_j-y)^2 \end{bmatrix}, \tag{22}$$

where x_i and y_j are sampled in the neighborhood of (x, y).

Alternatively, one can obtain a covariance estimate based on the estimated Hessian matrix, assuming Gaussianity. For a Gaussian distribution with a covariance matrix $\mathbf{C}$, the Hessian matrix at the peak is:

$$\mathbf{G} = -(2\pi)^{-d/2} |\mathbf{C}|^{-1/2} \mathbf{C}^{-1}. \tag{23}$$

By taking determinants on both sides, $|\mathbf{C}|$ can be expressed as a function of $\mathbf{G}$. Combining this with (23), we get:

$$\mathbf{C} = \left|2\pi(-\mathbf{G}^{-1})\right|^{-\frac{1}{d+2}} (-\mathbf{G}^{-1}). \tag{24}$$

This estimate is invariant to vertical shift of the RD-surface because the estimate depends only on the derivatives (see also [23]).

4.2 Local Matching of Control Points

To measure the local motion at each control point of a contour, different algorithms can be applied. Here, we discuss two approaches: The first is the Bayesian kernel matching algorithm [14]. The second is an optical flow-based matching algorithm [16].

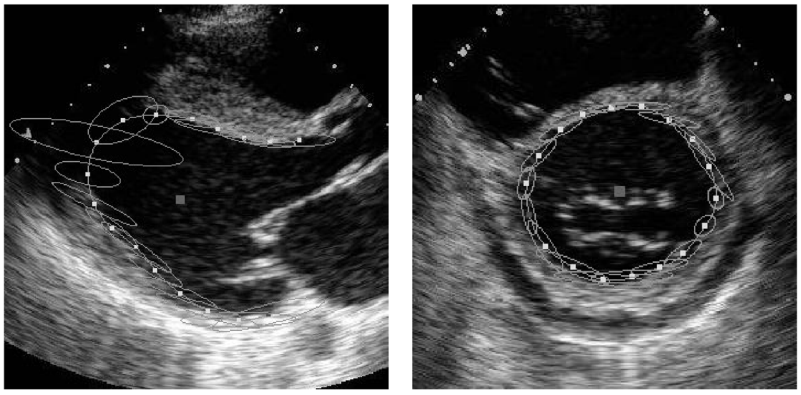


Fig. 3. The confidence ellipses corresponding to the local measurement uncertainty on each control point, estimated by the Bayesian matching algorithm of Section 4.2.1 and (24).

4.2.1 Bayesian Kernel Matching with Uncertainty

We adopt the algorithm proposed in [14], based on nonparametric density estimation.

Both the target model and the candidate image patch are represented by probability densities in the *joint spatial-intensity domain*. The candidate is assumed to be a transformed version of the target in this augmented space. The parametric transformation vector can be estimated using a maximum-likelihood formulation. The optimal transformation vector reveals the 2D motion and appearance change of the target. In contrast to classical correlation-based methods or SSD-based methods [2], [51], this approach captures appearance changes in a natural and principled way, and is robust to outliers caused by, for example, partial occlusion or impulsive noise.

Assuming that the transformation of the candidate density is only a translation in the joint spatial-intensity space, the log-likelihood of matching has the following form [14]:

$$L_{\mathbf{y}} = \sum_{r=1}^{N} \log p(\mathbf{x}_r + \mathbf{y}) = \sum_{r=1}^{N} \log \frac{1}{nh^d} \sum_{i=1}^{n} k\left(\left\| \frac{\mathbf{x}_r + \mathbf{y} - \mathbf{x}_i}{h} \right\|^2 \right), \tag{25}$$

where $Q = \{\mathbf{x}_r, r = 1N\}$ is a sample drawn i.i.d. from the target probability density q in the d-dimensional joint space, $\{\mathbf{x}_i, i = 1n\}$ is the sample drawn from the candidate density, $\mathbf{y}$ is the translation vector, and $k(\bullet)$ is the kernel for density estimation from samples, e.g., Gaussian kernel, with h its bandwidth parameter.

Derived from a mean shift procedure [17], we can maximize this term using the following iterative formula:

$$\mathbf{y}_1 = \frac{1}{N} \sum_{r=1}^{N} \frac{\sum_{i=1}^{n} (\mathbf{x}_i - \mathbf{x}_r) k\left(\left\| \frac{\mathbf{x}_r + \mathbf{y}_0 - \mathbf{x}_i}{h} \right\|^2 \right)}{\sum_{i=1}^{n} k\left(\left\| \frac{\mathbf{x}_r + \mathbf{y}_0 - \mathbf{x}_i}{h} \right\|^2 \right)}, \tag{26}$$

where $\mathbf{y}_0$ is the previous translation vector, and $\mathbf{y}_1$ is the updated vector. To start with, one can simply initialize $\mathbf{y}_0$ to be 0. The iteration is terminated when the increment is smaller than a threshold. A multiscale, coarse-to-fine image pyramid is used in our implementation, with three levels for the pyramid. The window size for the target model and the candidate patch is 17×17 for all levels.

To estimate the localization uncertainty, we perturb and estimate the log-likelihoods of (25) in the vicinity of the optimally matched location. Specifically, for each spatial location $(i, j) \in N_o$, where N_o is the neighborhood around the optimal spatial location, we find the optimal shift in intensity and record the corresponding log-likelihood $L(i, j)$. This can be done using (26), but *without updating the spatial translation during the iteration*.

According to the discussions in Section 4.1, we recover the response distribution as $RD(i, j) = \exp(k\, L(i, j))$, where k is used to normalize RD to sum to 1.

Equation (22) or (24) can then be applied for the estimation of the covariance. Fig. 3 depicts the uncertainty calculated using (24).

4.2.2 Flow-Based Matching Using Nonparametric Information Fusion

Our second matching algorithm is an adaptation of the frame-to-frame motion estimation algorithm described in [16], which is very competitive in terms of performance evaluation using standard sequences. We present in the sequel a brief summary of the algorithm.

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 tracking across scales, while the underlying mechanism for mode detection relies on the variable-bandwidth mean shift [15].

Assume a constant velocity model, the initial motion estimates for each location of a window of dimension $M \times M$ are computed through the traditional Least Squares [35] with some lower bound on the covariance eigenvalues to avoid instabilities, or through Biased Least Squares [24]. Then, using a robust fusion framework, we determine the location of the most significant mode in the two-dimensional velocity space by performing mode tracking across scales.

In the current work, for each control point, we compute initial estimates using 17×17 windows and fuse the results on 5×5 neighborhoods. A pyramid of three levels is employed with covariance propagation across levels. Fig. 4 depicts the uncertainty calculated at the bottom of the pyramid for the contour points from two sequences, using (18).

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). Thus, error accumulation from frame to frame is avoided. Since we update the

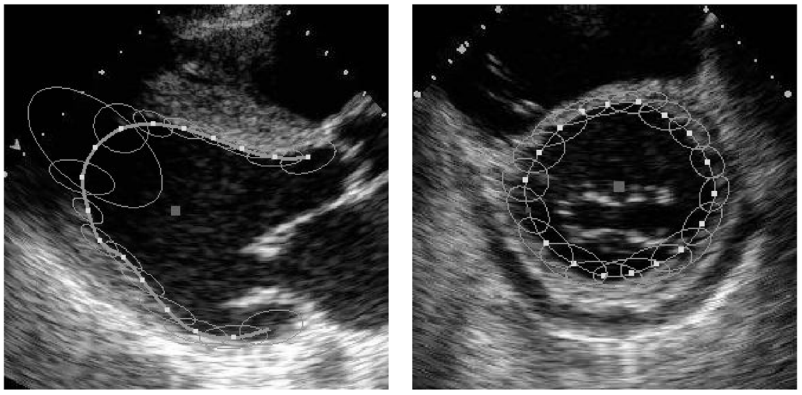


Fig. 4. The confidence ellipses corresponding to the local measurement uncertainty on each control point, estimated by the Flow algorithm of Section 4.2.2 and (18).

location of the model at each frame, the motion estimation process always starts with a good initialization. The overall procedure is suitable for the tracking of periodic sequences such as the heart ultrasound data. It resembles a template-based matching algorithm, which benefits from the fast computation of frame-to-frame motion.

5 Exploiting Uncertainty for Shape Alignment

To build an accurate PCA shape model, one needs an *aligned* training set. Without proper alignment, *trivial* global transformations (such as translation) may dominate the model. Of course, what types of variations are trivial depends on the application.

In this work, the training set is aligned using the iterative Procrustes analysis approach described by Cootes and Taylor [19] to cancel out global translation, rotation, and scaling. PCA is then performed to obtain the initial shape model. Therefore, before fusing the measured set of control points (Section 4.2) with the updated shape model (Section 3), we need to first align the measured point set with the model. In this section, we discuss the exploitation of heteroscedastic uncertainties during this alignment process.

5.1 Maximum-Likelihood Fusion with Alignment

We will borrow terms from Kendall et al. [32] and call an un-aligned control point set a *preshape* while referring to an aligned set as a *shape*. With a common reference, the assemblage of all possible shapes forms the *shape space* [32].[2]

Ideally, we should jointly optimize the alignment transform inside the fusion formulation of Section 2. Fig. 5 illustrates this idea: Given a measurement point $\mathbf{X}$ with covariance $\mathbf{C_X}$, and the model $\mathcal{N}(\mathbf{M}, \mathbf{C_M})$ in a subspace, we would like to find the optimal transform $\mathcal{T}^*$ and the point $\mathbf{X^M}$ (and its covariance $\mathbf{C_{X^M}}$) which has maximum-likelihood of generating $\mathcal{N}(\mathbf{M}, \mathbf{C_M})$ and $\mathcal{N}(\mathbf{X}^*, \mathbf{C_X}^*)$, where $\mathbf{X}^* = \mathcal{T}^*(\mathbf{X})$.

Unlike traditional Procrustes analysis formulations [47], [20], the problem here involves covariance matrices on both sides. To our best knowledge, it does not have a close-form solution over the similarity transform set. The global optimal can be sought numerically, but the computation will be expensive.

2. This definition of shape space is in agreement with that of Cootes and Taylor [19], but somewhat different from that of Blake and Isard ([9, p. 74]) which is defined as a linear subspace of the preshape space.

We resort to a two-step optimization approach, with close-form solutions for both steps. This scheme can be explained on Fig. 5: The first step is to go from $\mathbf{X}$ to $\mathbf{X}^*$, i.e., 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*.

5.2 Shape Alignment with Uncertainty

The problem of matching two (or more) point patterns is a classical one and was addressed as early as in the 1960s (for a list of papers see [20, p. 83]). Also known as the orthogonal Procrustes problem, most of its various solutions had implied either isotropic uncertainties on the point patterns or anisotropic weighting of the axes of the coordinate system [6]. Our goal, however, is to consider the component uncertainties during the transform of the preshape and its covariance matrix toward the model mean. We minimize

$$d^2 = (\mathbf{m} - \mathbf{x}')^T \mathbf{C}_\mathbf{x}'^{-1} (\mathbf{m} - \mathbf{x}'), \tag{27}$$

where $\mathbf{x}' = \mathcal{T}(\mathbf{x})$ and $\mathbf{C}'_\mathbf{x} = \mathcal{T}(\mathbf{C_x})$.

When $\mathcal{T}$ is the similarity transform, we have $\mathbf{x}' = \mathbf{R}\mathbf{x} + \mathbf{t}$, where $\mathbf{t}$ is the translation vector with two free parameters and $\mathbf{R}$ is a block diagonal matrix, with each block being

$$R = \begin{pmatrix} a & -b \\ b & a \end{pmatrix} \tag{28}$$

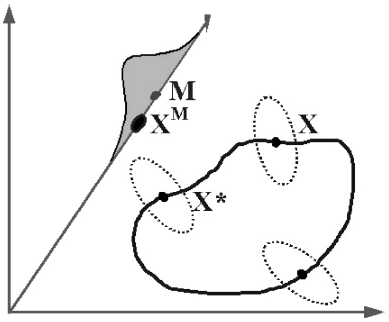


Fig. 5. Invariance manifold for shape alignment. Under invariant transforms, the pre-shape $\mathbf{X}$ traverses a manifold, $\mathcal{C}$, illustrated by the thick curve. In general, $\mathcal{C}$ will not intersect the model subspace $\mathcal{F}$ (the slanted axis containing the model centroid $\mathbf{M}$).

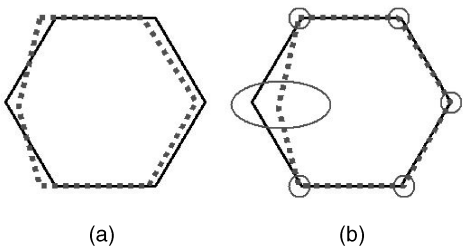


Fig. 6. Shape alignment. (a) Without considering uncertainties in localization and (b) with heteroscedastic [31], [28], [33] uncertainties. The ellipses depicts the covariance on point locations, representing information in a block-diagonal $\mathbf{C_x}$.

with the vector of free parameters $\mathbf{r} = (a, b)^T$, encapsulating rotation and scaling factors.

A subtly different formulation was solved in [19], [56], where the weight matrix is not a function of the transformation. We can rewrite (27) in the form of [19], [56] so that their results can be borrowed directly:

$$\begin{aligned} d^2 =& (\mathbf{R}^{-1}(\mathbf{m}-\mathbf{t})-\mathbf{x})^T\mathbf{C_x^{-1}}(\mathbf{R}^{-1}(\mathbf{m}-\mathbf{t})-\mathbf{x}) \\ =& (\mathcal{T}^{-1}(\mathbf{m})-\mathbf{x})^T\mathbf{C_x^{-1}}(\mathcal{T}^{-1}(\mathbf{m})-\mathbf{x}). \end{aligned} \tag{29}$$

When measurement uncertainties are estimated for different control points $\mathbf{x}_i$ independently, $\mathbf{C_x}$ becomes block-diagonal with 2×2 C_i on the diagonal and a close-form solution can be concisely written as follows:

$$\begin{pmatrix} \bar{\mathbf{r}} \\ \bar{\mathbf{t}} \end{pmatrix} = \begin{pmatrix} \sum C_i^{-1} Y_i, & \sum C_i^{-1} \\ \sum Y_i^T C_i^{-1} Y_i, & \sum Y_i^T C_i^{-1} \end{pmatrix}^{-1} \begin{pmatrix} \sum C_i^{-1}\mathbf{x}_i \\ \sum Y_i^T C_i^{-1}\mathbf{x}_i \end{pmatrix}, \tag{30}$$

where $\bar{\mathbf{r}} = (\bar{a}, \bar{b})^T$, $Y_i = (\mathbf{m}_i, J\mathbf{m}_i)$, and

$$J = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}.$$

Finally, we recover the rotation and scaling matrix R, and the translation vector $\mathbf{t}$, as follows:

$$R = \begin{pmatrix} \bar{a} & -\bar{b} \\ \bar{b} & \bar{a} \end{pmatrix}^{-1}, \quad \mathbf{t} = -\mathbf{R}\bar{\mathbf{t}}. \tag{31}$$

Fig. 6 illustrates shape alignment without and with considering uncertainties in point locations. Intuitively, the latter trusts the points with higher confidence during the alignment more.

5.3 Model Constraining with Uncertainty

With the preshape aligned with the model, the shape with maximum likelihood of being generated by the two competing information sources, namely, the aligned shape $\mathbf{X}^*$ versus the (subspace) model $\mathbf{M}$, can be obtained using (14) of Section 2.

Because the complete fusion is performed on the aligned shapes with a transformation $\mathcal{T}^*$, we need to transform the fusion result back to the pre-shape space by $\mathcal{T}^{*-1}$ before the prediction takes place for the next time step.

Fig. 7 shows a schematic diagram of the analysis steps where the uncertainty of measurement and prediction is propagated through all the steps. To take into account the shape invariances, the complete fusion formulation is implemented in two steps, separated by shape alignment.

6 Implementation, Evaluation, and Analysis

In this section, we test the proposed framework using synthetic sequences and real-world ultrasound heart sequences.

Echocardiography data collected from 60 patients are used to form the training and the test sets. Both sets were manually traced by experts, and confirmed by a cardiologist, to provide ground truth contours. Both apical two/four-chamber views and parasternal long/short axis views are trained and tested. Landmark points are assigned based on anatomic features (e.g., the apex, the papillary muscles, etc.). We order the control points in clockwise order, with the first point at the mid inferior region for the short axis view and at the basal septal/inferior region for apical views. Furthermore, the mid-point for the apical view is assigned at the apex. We use 17 control points for the apical views and the long-axis view and 18 for the short axis views. The algorithm can tolerate some variability on the location of the landmark points, partly due to the application of SA-PCA.

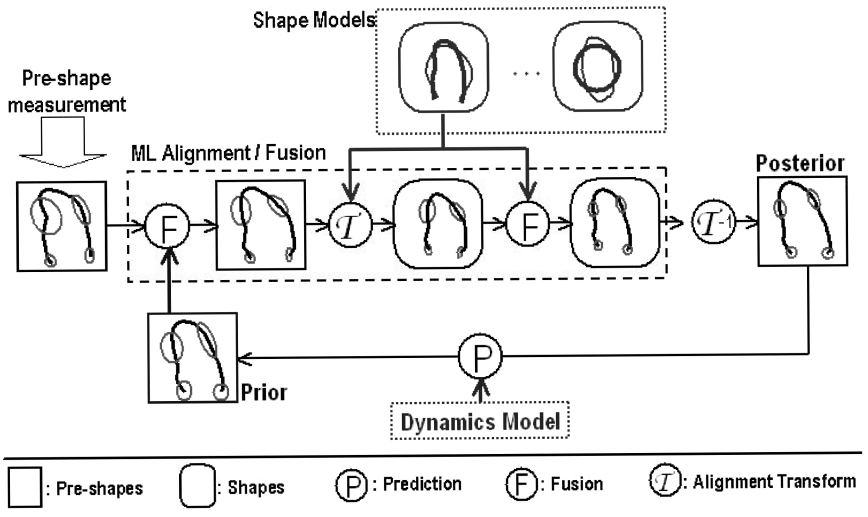


Fig. 7. Uncertainty propagation during shape tracking. The small gray ellipses illustrate the location uncertainties. Notice that uncertainties are transformed with the shape during alignment and fused with the model and the (predicted) prior information during likelihood estimation and tracking.

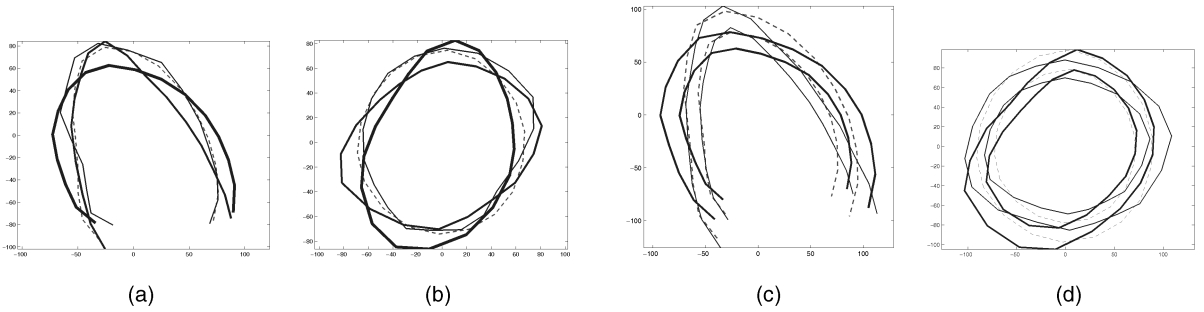


Fig. 8. The dominant eigenshapes for: (a) and (b) single contour, (c) and (d) coupled contours, (a) and (c) apical views, (b) and (d) short axis views. The dashed curves are the model mean.

The training contours are aligned using the iterative Procrustes analysis approach described by Cootes and Taylor [19] to cancel out global translation, rotation, and scaling. PCA is then performed and the dimensionality is reduced to retain 80-97 percent of energy, separately tuned for each model. Fig. 8 shows the most dominant eigenshapes for two views 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 and its model is trained in the same way as a single contour model.

During testing, we assume manual initialization and use the dynamic model to impose a temporal smoothness constraint. Without prior knowledge regarding the possible motion abnormalities for the current patient under study, we employ a 0th-order dynamic model to represent the uncertainty in system dynamics and set the relative confidence of this model empirically. Although a more sophisticated dynamic model encoding the periodicity in heart motion may provide stronger predictive power, our goal is to detect *abnormal* and *unpredictable* local motions (e.g., hypokinetic, akinetic, or paradoxical motion of regional heart muscles)—A strong predictive motion model can also drown out subtle manifestations of regional wall motion abnormalities. Since we take local motion measurements in a coarse-to-fine fashion on a three-level pyramid, our algorithm can track large motion without a strong prior motion model. Our simple dynamic model suffices and serves effectively as a temporal smoothness constraint.

At each tracking step, before the fusion step, the measurement is aligned to the model using the transformation of (31). The covariance matrix is also transformed accordingly.

6.1 Performance Measures

We tested two distance measures: the Mean Sum of Squared Distance (MSSD) [1] and the Mean Absolute Distance (MAD) [38]. With consistent results, we report only the MSSD in this paper. For the sequence S_i with m contours, $\{c_1, \ldots, c_m\}$, where each contour c_j has n points $\{(x_{j,1}, y_{j,1}), \ldots, (x_{j,n}, y_{j,n})\}$, the distance to the ground truth S_i^0 is defined as $MSSD_i = \frac{1}{m}\sum_{j=1}^{m}\frac{1}{n}\sum_{k=1}^{n} SD_{i,j,k}$ and the mean spatial error deviation is defined as $\sigma_i = \frac{1}{m}\sum_{j=1}^{m}\sigma_{i,j}$, where

$$SD_{i,j,k} = \left((x_{j,k} - x_{j,k}^0)^2 + (y_{j,k} - y_{j,k}^0)^2\right) \tag{32}$$

and $\sigma_{i,j}$ is the sample standard deviation of $SD_{i,j,k}$ across k.

The overall performance measures for a particular method are the averages of $MSSD_i$ and σ_i across the whole test set of l sequences, $\overline{MSSD}$ and $\bar{\sigma}$. As an approximate measure of pixel errors, we use the square root of $\overline{MSSD}$. A critical difference between our distance measures and those of [1] or [38] is that we have the point correspondence through tracking. As a result, we could capture tangent motion components along the contour which can reveal crucial information about cardiac function.

Five methods are compared. The first is a tracker without shape constraint ("Unconstrained"), using either the Bayesian kernel tracker or the flow-based tracker (Section 4.2) to track the control points independently. The second approach is the same tracker but adding orthogonal PCA shape model constraints, which represents the current approaches [29], [19], [9] ("ShapeSpace"). The third is "ShapeSpace" but using our SA-PCA model ("SAPCA"). The fourth approach is our fusion framework using traditional PCA models ("F-PCA"). The last is the proposed framework ("Proposed"), which uses both fusion and the SA-PCA models.

6.2 Tracking Synthetic Sequence

To ensure unbiased evaluation, we first use synthetic sequences with known ground truth for performance comparison. The sequences consist of a shrinking circular ring moving right-downward for 20 frames (800×720), with portions of the ring disappearing in three frames out of every six frames (Fig. 9). Three kinds of noise are added to each frame: zero-mean Gaussian white noise (with variance σ^2), multiplicative speckle noise (which changes a pixel value I to $I + \lambda I$, where λ is uniformly distributed random noise with mean 0 and variance σ_λ^2), and salt-and-pepper noise (Table 1).[3] The selection of noise is partially motivated by the noise occurring in ultrasound images [12], [41], but we are interested in not only ultrasound sequences but also signals with general noise settings since the proposed framework is general. Fig. 9 shows some screen shots of the tracking results. Fig. 10 presents comparisons of the four model-constraining approaches based on either the Bayesian or the flow-based trackers. Observe the significant and consistent improvements of the proposed approaches over the existing "ShapeSpace" approach. As the shape model, we have borrowed the model for the parasternal short axis view (Fig. 8b).

It should be noted that the algorithms do not assume a circular shape other than statistically exploiting the initial contour when SAPCA is employed. SAPCA in this case did

3. Since the multiplicative speckle noise is added on top of the Gaussian noise, the increase of noise variance of the Gaussian also means increased influence from the speckle noise.

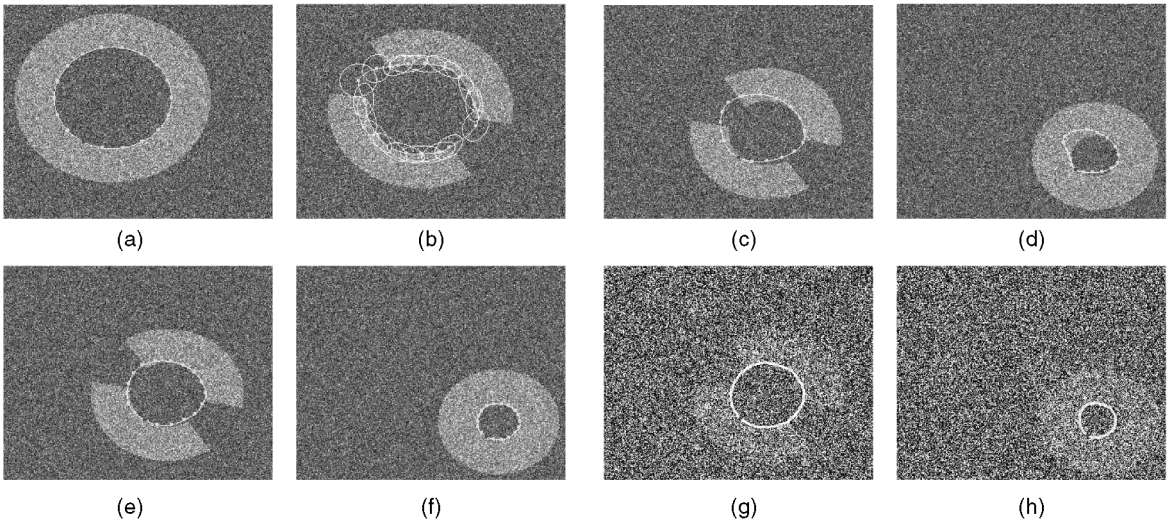


Fig. 9. Comparison of tracking results on synthetic sequences. (a) The initial contour. (b) The unconstrained tracking on the fourth frame with ellipses depicting motion uncertainties. (c), (d), (e), and (f) The 12th and 20th frames tracked by different methods: (c) and (d) "ShapeSpace" method using flow, (e) and (f) "Proposed" method using flow (noise level 3). (g) and (h) "Proposed" method using Bayesian tracker (noise level 8).

TABLE 1
Noise Compositions and Levels for the Synthetic Sequences

	Noise levels (image intensities in [0, 1])							
Noise / Drop-out	1	2	3	4	5	6	7	8
Gaussian (σ^2)	0	0.02	0.02	0.02	0.05	0.08	0.1	0.3
Speckle (σ_λ^2)	0	0	0.04	0.04	0.04	0.04	0.04	0.1
Salt&Pepper (%)	0%	0%	0%	5%	5%	5%	5%	10%
Signal Drop-outs	yes	yes	yes	yes	yes	yes	yes	yes

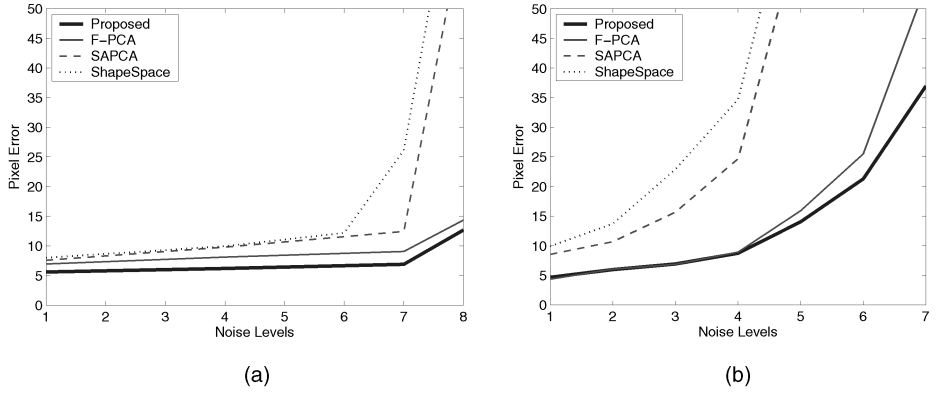


Fig. 10. Performance comparison over noisy synthetic sequences. (a) Bayesian kernel tracking. (b) Flow-based tracking.

not play a significant role since the model itself can represent a circle very well, if not perfectly, and "F-PCA" is quite close to "Proposed." However, in general, this is not the case (see the next experiment).

The Bayesian tracker has better performance than the flow-based tracker. Nevertheless, the trade off is speed—it is more than 100 times slower than the Flow tracker. In this experiment, the number of iterations for (26) varies and can reach 60, with an average around 20.

6.3 Tracking Echocardiography Sequences

Due to the portability of ultrasound machines and the relatively low cost and wide availability of echocardiogram exams, automatic processing and analysis of echocardiography sequences is extremely valuable in aiding sensitive and robust early detection of common heart diseases. Automatic tracking of heart borders can reveal early signs of vascular and myocardium dysfunctions. However, ultrasound is noisier than other common medical imaging modalities such as MRI or CT. Echocardiogram is even worse due to the fast motion of the heart muscle and respiratory interferences. It usually exhibits spatially varying structural noise characteristics such as signal drop-outs [41]. Our proposed framework exploiting heteroscedastic measurement uncertainties is ideal for tracking in echocardiography sequences.

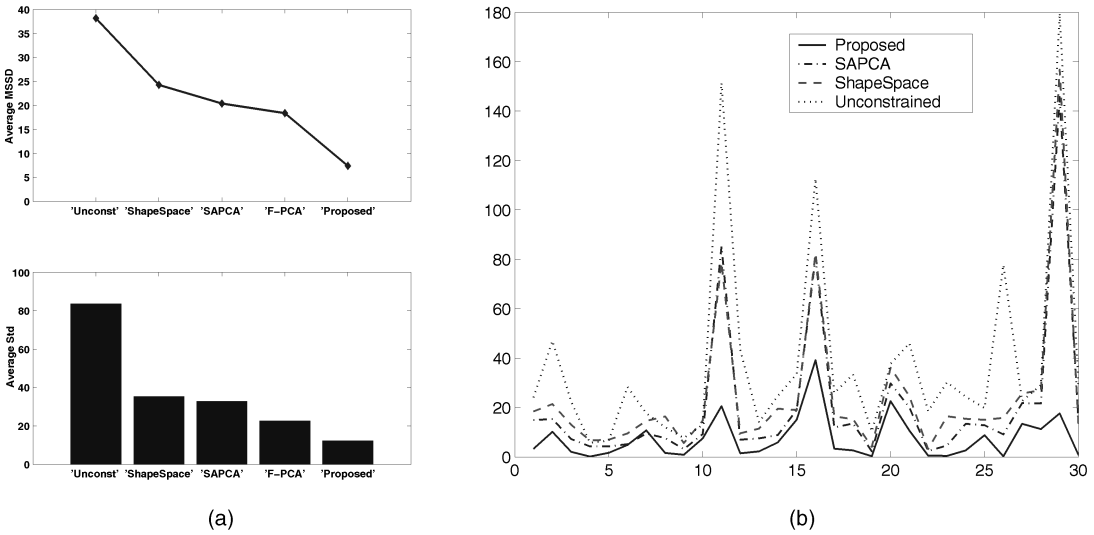


Fig. 11. Comparison of tracking results on real-world sequences. (a) The averaged MSSDs and standard deviations for different methods. (b) MSSD curves for the 30 test sequences.

For systematic evaluation on real-world data, a set of 30 echocardiography sequences are used for testing, with lengths ranging from 18 to 90 frames, the frame sizes are from 480×430 to 640×480. Due to the practical requirement of real-time performance, we use the flow-based matching algorithm in our current system, which runs at 30 frames per second on a 3GHz PC.

Fig. 11 shows the comparison of the four subspace shape model constraining methods as well as the unconstrained flow measurements. Our proposed method ("Proposed") significantly outperforms others, with an average MSSD of 7.4 ($\bar{\sigma} = 12.3$) as opposed to 24.3 ($\bar{\sigma} = 35.4$) by the current model constraining approach ("ShapeSpace"). Our SA-PCA model alone ("SAPCA") already brought significant improvement, achieving an average MSSD of 20.4 ($\bar{\sigma} = 32.8$). The fusion alone ("F-PCA") had an average MSSD of 18.4 ($\bar{\sigma} = 22.7$). The combined use of fusion and SA-PCA (i.e., "Proposed") has apparently brought out a significant performance boost over each alone (Fig. 11a). Fig. 12 shows some tracked sequences.

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 Fig. 13.

Our SA-PCA model is especially helpful for shapes that differ significantly from the training set. Fig. 14 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 Fig. 8a) 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 (Fig. 14a), but with a very small weight ($< 0.01\%$), fails to follow such distinctive patterns and has constrained the contours to a typical A4C shape (Fig. 14b). SA-PCA yields a contour that fits much better to the true border (Fig. 14c). Regarding the sensitivity with respect to α, the interested reader is referred to [57].

6.4 Coupled Double-Contour versus Single Contour

Cardiologists often use *wall thickening*, if visible, in addition to *wall motion* as a more reliable indicator for diagnosing myocardium functions [41] because *wall motion* is often susceptible to corruptions from patient or probe movement or respiratory motion. Our framework can be directly applied for coupled double-contour tracking by simply treating the inner and the outer contours together as a single shape.

There have been various research efforts on multiple contour segmentation and tracking incorporating topological constraints. Recent examples include Wang et al.'s work [55], where topology-preserving constraints are incorporated in a support vector machine (SVM) regression setting, and Goldenberg et al.'s [21] and Paragios' [42] work based on geometric variational approaches using geodesic active surface or active region models. While all the above approaches use explicit, "hard" constraints, our coupling is implicit and probabilistic (governed by the training set) and, thus, "soft." To minimize the chance of inner/outer contour crossing, we shall first make sure no crossing happens in the training set. Additionally, with our complete constraint using the model distribution using (9), we constrain not only *the mode* but also *the range of shape deformations*; this, in turn, can further reduce the probability of crossing.

For echocardiography, in many cases, the epicardium is less visible than the endocardium (except for the case of pericardial effusion for which the opposite is true!). Intuitively, a double-contour approach can propagate information from the endocardium to guide the localization of the epicardium (or vice versa), thus achieving more robust tracking of the two borders than tracking them independently. Fig. 15 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 the large intensity shift at the base, which is in large part to blame for the errors made by the single contours. However, a statistically significant evaluation of epicardial border localization is difficult due to the lack of sufficient and reliable ground truth data: The localization ambiguity and interexpert variability are much higher for the

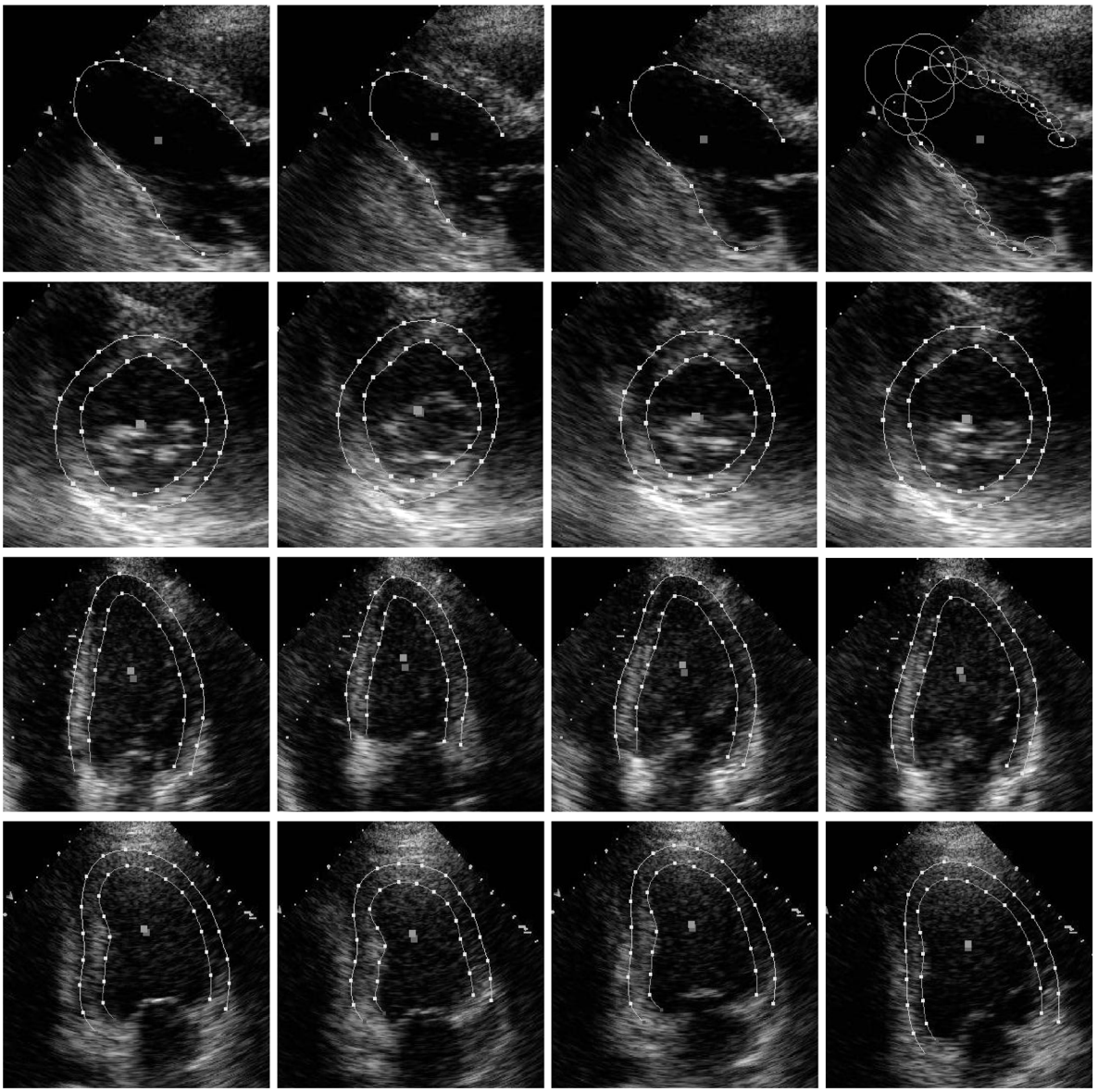


Fig. 12. Four tracking examples in rows, corresponding to the four common views of echocardiogram: parasternal long axis, parasternal short axis, apical four chamber, and apical two chamber views, respectively. The frame numbers are 1, 16, 24, and 32; 1, 12, 67, and 81; 1, 12, 18, and 23; 1, 15, 23, and 30; respectively.

epicardial border than the endocardial border. A recent study turned in only inconclusive results and recommended the use of ground truth from registered magnetic resonance (MR) images [29]. We suggest that another meaningful way to evaluate is to compare final classification results (instead of border locations) from the machine versus cardiologists. This is among our future efforts.

7 Conclusions, Discussions, and Future Work

This paper presented a joint information fusion framework to track shapes under heteroscedastic uncertainties with a strongly adapted subspace model constraint. Heteroscedastic measurement uncertainties are considered during the subspace shape model constraining process and it has been shown to be a natural extension of the classical orthogonal projection-based model constraining approach. Furthermore, we integrated this fusion formulation into the Kalman tracking framework so that subspace models can be incorporated in a unified and flexible way, allowing either a subspace itself or a subspace model (i.e., a Gaussian distribution) as the constraint. We discussed the issues of measurement uncertainty estimation and presented two methods for local motion and uncertainty estimation. When invariant shape transformations are adopted during the model construction step, measurement uncertainty is also exploited during the shape alignment process. The proposed framework was evaluated using both synthetic data and real-world echocardiography sequences. The results showed that, with heteroscedastic uncertainties, the proposed fusion framework with a strong model adaptation yields significant performance improvement over existing approaches.

Although our motivating application is the tracking of echocardiography, the framework is general and can be easily extended for other applications. For example, we have tested

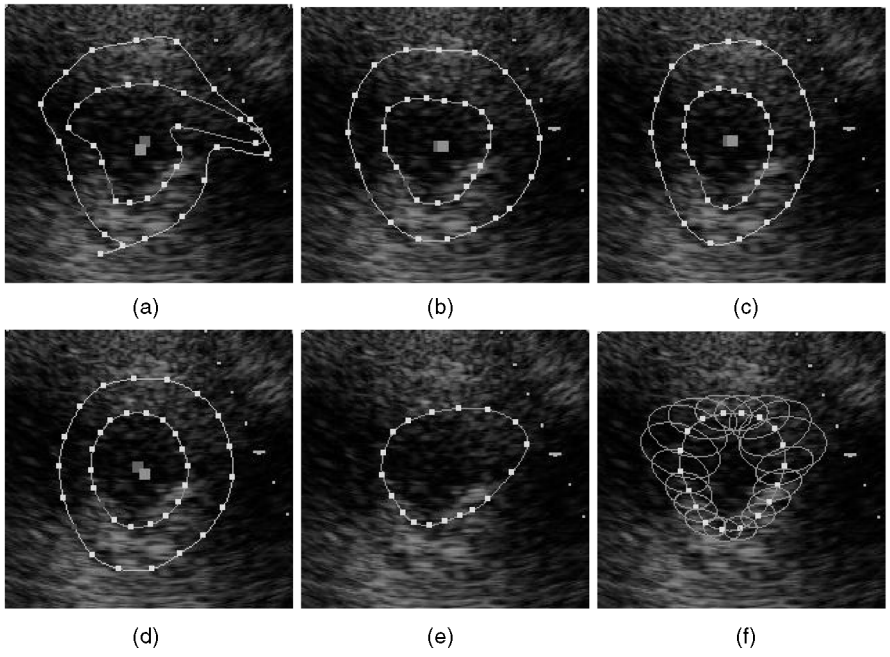


Fig. 13. Performance comparison in handling large measurement errors. For the same frame we show: (a) "Unconstrained" flow, (b) "SAPCA," (c) "Proposed," (d) by an expert, (e) "SAPCA" but only for endocardial border, and (f) "Proposed" but for endocardial border (also shown are the uncertainty ellipses). Notice the stronger correction the "Proposed" brought over the local measurement errors for both single and double contours.

Fig. 14. SA-PCA versus incremental PCA. (a) The initial contour, (b) the 14th frame using an incremental PCA model [22], and (c) the same frame using an SA-PCA model ($\alpha = 0.5$).

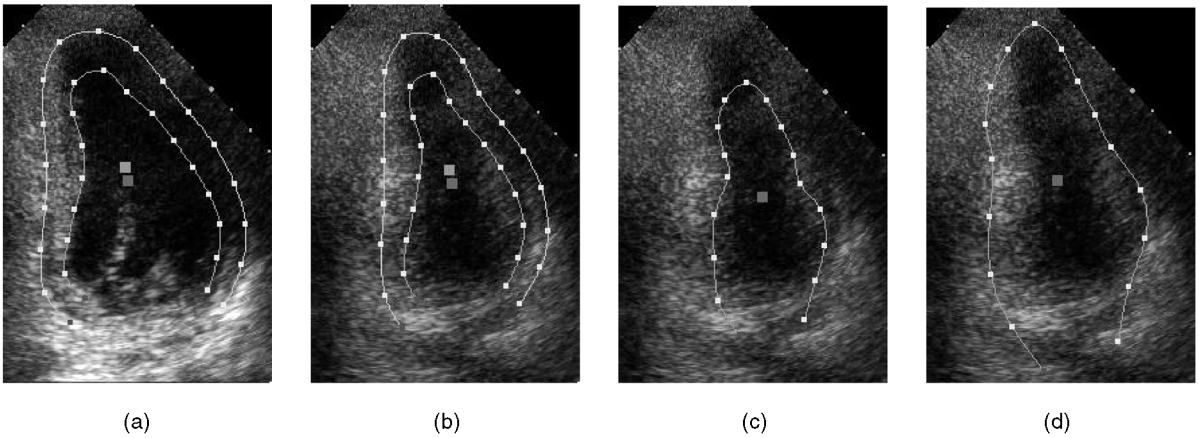


Fig. 15. Double versus single contours. (a) The initial contours, (b) seventh frame tracked by the double-contour, (c) seventh frame with the inner contour, and (d) seventh frame with the outer contour.

our framework directly for tracking the silhouette of a moving human head. As the shape model, we just borrowed the model of Fig. 8a, updated online through SAPCA. Initial testing showed that our proposed framework is more robust when the uncertainty information is fully exploited. Fig. 16 shows tracking results on two frames with partial occlusions.

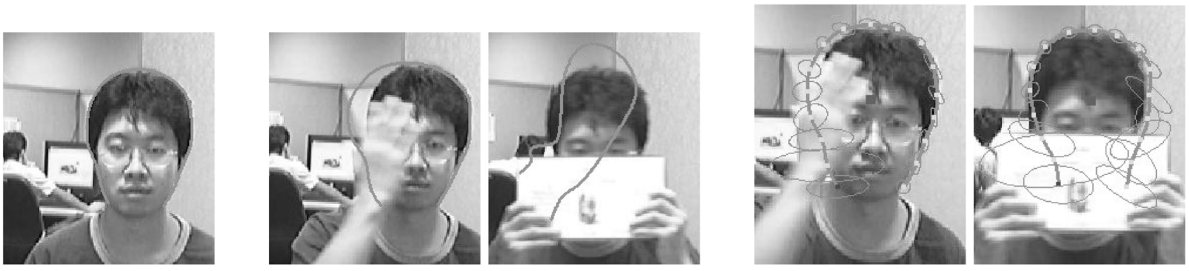


Fig. 16. Tracking head silhouette with partial occlusions using the proposed framework. (a) The initial frame. (b) and (c) "ShapeSpace," (d) and (e) "Proposed" where the ellipses indicate matching uncertainties, estimated using (24).

This example indicates that our framework is not only suitable for ultrasound signals with regional drop-out, but also potentially applicable for dealing with partial occlusions. Investigations are ongoing for application of the proposed framework to this and other vision problems.

Extensions to 3D or 2D+T(time) are natural. Our method might benefit from the addition of robust statistics tools [26]. Another natural extension may be to exploit the subspace constraint together with the heteroscedastic model in a particle filter framework. Finally, there are extensive research efforts on variational and level-set approaches and Bayesian approach to shape learning, representation, and extraction (e.g., [39]). An interesting future research topic is the unification of different approaches to provide more robust and efficient solutions.

ACKNOWLEDGMENTS

The authors would like to thank Visvanathan Ramesh and Bogdan Georgescu from Siemens Corporate Research, and Bharat Rao, Sriram Krishnan, and David Gustafson of Siemens Medical Solutions for insightful discussions on the subject. They are also grateful for the many enjoyable discussions and invaluable guidances from Dr. Randolph P. Martin from the Emory Heart Center.

REFERENCES

[1] Y. Akgul and C. Kambhamettu, "A Coarse-to-Fine Deformable Contour Optimization Framework," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 25, no. 2, pp. 174-186, Feb. 2003.
[2] P. Anandan, "A Computational Framework and an Algorithm for the Measurement of Visual Motion," *Int'l J. Computer Vision,* vol. 2, pp. 283-310, 1989.
[3] B. Anderson and J. Moore, *Optimal Filtering.* Prentice Hall, 1979.
[4] Y. Bar-Shalom and L. Campo, "The Effect of the Common Process Noise on the Two-Sensor Fused Track," *IEEE Trans. Aerospace and Electronic Systems,* vol. 22, pp. 803-805, 1986.
[5] Y. Bar-Shalom and T. Fortmann, *Tracking and Data Association.* Academic Press, 1988.
[6] P. Batchelor and J. Fitzpatrick, "A Study of the Anisotropically Weighted Procrustes Problem," *Proc. IEEE Workshop Math. Methods in Biomedical Image Analysis,* pp. 212-218, 2000.
[7] M. Black and P. Ananadan, "The Robust Estimation of Multiple Motions: Parametric and Piecewise-Smooth Flow Fields," *Computer Vision and Image Understanding,* vol. 63, no. 1, pp. 75-104, 1996.
[8] M. Black and A. Jepson, "Eigentracking: Robust Matching and Tracking of Articulated Objects Using a View-Based Representation," *Proc. European Conf. Computer Vision,* pp. 610-619, 1996.
[9] A. Blake and M. Isard, *Active Contours.* Springer Verlag, 1998.
[10] M. Brand and R. Bhotika, "Flexible Flow for 3D Nonrigid Object Tracking and Shape Recovery," *Proc. IEEE Conf. Computer Vision and Pattern Recognition,* vol. I, pp. 315-322, 2001.
[11] C. Bregler, A. Hertzmann, and H. Biermann, "Recovering Non-Rigid 3D Shape from Image Streams," *Proc. IEEE Conf. Computer Vision and Pattern Recognition,* vol. II, pp. 690-696, 2000.
[12] C. B. Burckhardt, "Speckle in Ultrasound b-Mode Scans," *IEEE Trans. Sonics and Ultrasonics,* vol. 25, no. 1, pp. 1-6, 1978.
[13] P. Clarysse, D. Friboulet, and I.E. Magnin, "Tracking Geometrical Descriptors on 3-D Deformable Surfaces: Application to the Left-Ventricular Surface of the Heart," *IEEE Trans. Medical Imaging,* vol. 16, no. 4, pp. 392-404, 1997.
[14] D. Comaniciu, "Bayesian Kernel Tracking," *Proc. Ann. Conf. German Soc. for Pattern Recognition,* pp. 438-445, 2002.
[15] D. Comaniciu, "An Algorithm for Data-Driven Bandwidth Selection," *IEEE Trans. Pattern Analysis Machine Intelligence,* vol. 25, no. 2, pp. 281-288, Feb. 2003.
[16] D. Comaniciu, "Nonparametric Information Fusion for Motion Estimation," *Proc. IEEE Conf. Computer Vision and Pattern Recognition,* vol. I, pp. 59-66, 2003.
[17] D. Comaniciu and P. Meer, "Mean Shift: A Robust Approach toward Feature Space Analysis," *IEEE Trans. Pattern Analysis and and Machine Intelligence,* vol. 24, no. 5, pp. 603-619, May 2002.
[18] T. Cootes and C. Taylor, "Active Shape Models-'Smart Snakes'," *Proc. British Machine Vision Conf.,* pp. 266-275, 1992.
[19] T. Cootes and C. Taylor, "Statistical Models for Appearance for Computer Vision," unpublished manuscript, http://www.wiau.man.ac.uk/Models/app_model.ps.gz, 2001.
[20] I.L. Dryden and K.V. Mardia, *Statistical Shape Analysis.* John Wiley & Sons, Ltd., 1998.
[21] R. Goldenberg, R. Kimmel, E. Rivlin, and M. Rudzsky, "Cortex Segmentation: A Fast Variational Geometric Approach," *IEEE Trans. Medical Imaging,* vol. 21, no. 2, pp. 1544-1551, Dec. 2002.
[22] P. Hall, D. Marshall, and R. Martin, "Merging and Splitting Eigenspace ZModels," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 22. no. 9, pp. 1042-1048, Sept. 2000.
[23] B. Han, D. Comaniciu, and L. Davis, "Sequential Kernel Density Approximation through Mode Propagation: Applications to Background Modeling," *Proc. Asian Conf. Computer Vision,* 2004.
[24] A. Hoerl and R. Kennard, "Ridge Regression: Biased Estimation for Nonorthogonal Problems," *Technometrics,* vol. 12, no. 1, pp. 55-67, 1970.
[25] B. Horn and B. Schunck, "Determining the Optical Flow," *Artificial Intelligence,* vol. 17, pp. 185-203, 1981.
[26] P. J. Huber, *Robust Statistics.* Wiley, 1981.
[27] M. Irani, "Multi-Frame Optical Flow Estimation Using Subspace Constraints," *Proc. Int'l Conf. Computer Vision,* pp. 626-633, Sept. 1999.
[28] M. Irani and P. Anandan, "Factorization with Uncertainty," *Proc. Sixth European Conf. Computer Vision,* pp. 539-553, 2000.
[29] 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.
[30] S. Julier and J. Uhlmannn, "A Non-Divergent Estimation Algorithm in the Presence of Unknown Correlations," *Proc. Am. Control Conf.,* 1997.
[31] Y. Kanazawa and K. Kanatani, "Do We Really Have to Consider Covariance Matrices for Image Features?" *Proc. Int'l Conf. Computer Vision,* vol. II, pp. 586-591, 2001.

[32] D.G. Kendall, D. Barden, T.K. Carne, and H. Le, *Shape and Shape Theory.* John Wiley & Sons, Ltd., 1999.
[33] Y. Leedan and P. Meer, "Heteroscedastic Regression in Computer Vision: Problems with Bilinear Constraint," *Int'l J. Computer Vision,* vol. 37, no. 2, pp. 127-150, 2000.
[34] X. Li, Y. Zhu, and C. Han, "Unified Optimal Linear Estimation Fusion-Part I: Unified Models and Fusion Rules," *Proc. Third Int'l Conf. Information Fusion,* pp. MoC2-10-MoC2-17, 2000.
[35] B. Lucas and T. Kanade, "An Iterative Image Registration Technique with an Application to Stereo Vision," *Proc. DARPA Imaging and Understanding Workshop,* pp. 121-130, 1981.
[36] G.E. Mailloux, F. Langlois, P.Y. Simard, and M. Bertrand, "Restoration of the Velocity Field of the Heart from Two-Dimensional Echocardiograms," *IEEE Trans. Medical Imaging,* vol. 8, no. 2, pp. 143-153, 1989.
[37] J. McEachen and J. Duncan, "Shape-Based Tracking of Left Ventricular Wall Motion," *IEEE Trans. Medical Imaging,* vol. 16, pp. 270-283, 1997.
[38] I. Mikic, 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.
[39] W. Mio, A. Srivastava, and X. Liu, "Learning and Bayesian Shape Extraction for Object Recognition," *Proc. European Conf. Computer Vision,* 2004.
[40] K. Nickels and S. Hutchinson, "Estimating Uncertainty in SSD-Based Feature Tracking," *Image and Vision Computing,* vol. 20, pp. 47-58, 2002.
[41] J.K. Oh, J.B. Seward, and A.J. Tajik, *The Echo Manual.* Philadelphia: Lippincott Williams & Wilkins, 1999.
[42] N. Paragios, "A Variational Approach for the Segmentation of the Left Ventricle in MR Cardiac Images," *Proc. IEEE Workshop Variational and Level Set Methods,* 2001.
[43] J. Rehg and T. Kanade, "Visual Tracking of High DoF Articulated Structures: An Application to Human Hand Tracking," *Proc. European Conf. Computer Vision,* pp. 35-46, 1994.
[44] M. Rogers and J. Graham, "Robust Active Shape Model Search," *Proc. European Conf. Computer Vision,* pp. 517-530, 2002.
[45] L.L. Scharf, *Statistical Signal Processing.* Reading, Mass.: Addison Wesley, 1991.
[46] C. Schmid, R. Mohr, and C. Bauckhage, "Evaluation of Interest Point Detectors," *Int'l J. Computer Vision,* vol. 37, no. 2, pp. 151-172, 2000.
[47] P.H. Schonemann, "A Generalized Solution of the Orthogonal Procrustes Problem," *Psychometrika,* vol. 31, pp. 1-10, 1966.
[48] J. Shi and C. Tomasi, "Good Features to Track," *Proc. IEEE Conf. Computer Vision and Pattern Recognition,* pp. 593-600, June 1994.
[49] E. Simoncelli, E. Adelson, and D. Heeger, "Probability Distributions of Optical Flow," *Proc. IEEE Conf. Computer Vision and Pattern Recognition,* pp. 310-315, June 1991.
[50] A. Singh, *Optical Flow Computation: A Unified Perpective.* Los Alamitos, Calif. : IEEE CS Press, 1992.
[51] A. Singh and P. Allen, "Image-Flow Computation: An Estimation-Theoretic Framework and a Unified Perspective," *CVGIP: Image Undestanding,* vol. 56, no. 2, pp. 152-177, 1992.
[52] M. Suhling, M. Arigovindan, P. Hunziker, and M. Unser, "Motion Analysis of Echocardiograms Using a Local-Affine, Spatio-Temporal Model," *Proc. IEEE Int'l Symp. Biomedical Imaging: Macro to Nano,* pp. 573-576, 2002.
[53] E. Trucco and A. Verri, *Introductory Techniques for 3-D Computer Vision.* Prentice Hall, 1998.
[54] M.A. Turk and A.P. Pentland, "Face Recognition Using Eigen-Face," *Proc. IEEE Conf. Computer Vision and Pattern Recognition,* pp. 586-591, 1991.
[55] S. Wang, X. Ji, and Z.-P. Liang, "Landmark-Based Shape Deformation with Topology-Preserving Constraints," *Proc. Int'l Conf. Computer Vision,* 2003.
[56] B. Xie, D. Comaniciu, V. Ramesh, M. Simon, and T. Boult, "Component Fusion for Face Detection in the Presence of Heteroscedastic Noise," *Proc. Ann. Conf. German Soc. for Pattern Recognition,* pp. 434-441, 2003.
[57] X.S. Zhou, D. Comaniciu, and S. Krishnan, "An Information Fusion Framework for Robust Shape Tracking," *Proc. Int'l Workshop Statistical and Computational Theories of Vision,* 2003.
[58] X.S. Zhou, D. Comaniciu, and S. Krishnan, "Coupled-Contour Tracking through Non-Orthogonal Projections and Fusion for Echocardiography," *Proc. European Conf. Computer Vision,* 2004.

Xiang Sean Zhou received the PhD degree in electrical engineering from the University of Illinois at Urbana Champaign (UIUC) in 2002. Previously, he received the bachelor's degree in automation and in economics and management (minor) and studied economics for two years in a PhD program at Tsinghua University, China. Since 2002, he has been with Siemens Corporate Research in Princeton, New Jersey. His research interests include computer vision and machine learning, object detection, tracking, and recognition, multimedia analysis, representation, understanding, and retrieval. He was the program cochair for the 2003 International Conference on Image and Video Retrieval. He is the coauthor of the book *Exploration of Visual Data* (Kluwer 2003). He was the recipient of eight scholarships and awards from Tsinghua University from 1988 to 1995. In 2001, he received the M.E. Van Valkenburg Fellowship Award, an award given to one or two PhD students in the ECE department of UIUC each year "for demonstrated excellence in research in the areas of circuits, systems, or computers. " He is a member of the IEEE and the IEEE Computer Society.

Dorin Comaniciu received PhD degrees in electrical engineering from the Polytechnic University of Bucharest in 1995 and from Rutgers University in 1999. Since 1999, he has been with Siemens Corporate Research in Princeton, New Jersey, as a senior member of technical staff and manager of the Statistical Methods for Vision Systems Program. He is currently the head of the Integrated Data Systems Department. His research interests include robust methods for autonomous computer vision, motion estimation, nonparametric analysis, robust information fusion, medical imaging, content-based access to visual data, personalized medicine, and integrated information modeling. He has coauthored more than 80 papers, conference papers, and book chapters in the area of visual information processing. He received the Best Paper Award at the IEEE Conference on Computer Vision and Pattern Recognition 2000 and is a coauthor of the upcoming book *Nonparametric Analysis of Visual Data: The Mean Shift Paradigm*. He is an associate editor of *Pattern Analysis and Applications* and a senior member of the IEEE.

Alok Gupta received the B.Tech. degree in computer science and engineering from the Indian Institute of Technology in Kharagpur, India, in 1986, the MS and PhD degrees in computer and information science from GRASP laboratory at the University of Pennsylvania in Philadelphia, in 1988 and 1991, respectively, and the MBA degree from the Wharton School of the University of Pennsylvania in 2002. He was with Siemens Corporate Research in Princeton, New Jersey, from 1991 to 2002, first as a member of the technical staff and a project manager and, from 1996 to 2002, as the department head for the Imaging & Visualization Department, where he was responsible for image processing and computer vision technologies for Siemens worldwide. In May 2002, he launched Siemens Medical's internal venture on Computer-Aided Diagnosis and Therapy (CAD), with the responsibility for building Siemens' CAD business with strong R&D competencies in medical image processing, data fusion, and machine learning. His current research and commercialization interests are in the wide area of automation of medical data processing for clinical decision support. He has published numerous refereed papers in the field of computer vision, image processing and medical imaging, and holds seven patents, with several patents pending.