Learning Coupled Prior Shape and Appearance Models for Segmentation

Xiaolei Huang, Zhiguo Li, and Dimitris Metaxas

Center for Computational Biomedicine Imaging and Modeling,
Division of Computer and Information Sciences, Rutgers University, NJ, USA

Abstract. We present a novel framework for learning a joint shape and appearance model from a large set of un-labelled training examples in arbitrary positions and orientations. The shape and intensity spaces are unified by implicitly representing shapes as "images" in the space of distance transforms. A stochastic chord-based matching algorithm is developed to align photo-realistic training examples under a common reference frame. Then dense local deformation fields, represented using the cubic B-spline based Free Form Deformations (FFD), are recovered to register the training examples in both shape and intensity spaces. Principal Component Analysis (PCA) is applied on the FFD control lattices to capture the variations in shape as well as on registered object interior textures. We show examples where we have built coupled shape and appearance prior models for the left ventricle and whole heart in short-axis cardiac tagged MR images, and used them to delineate the heart chambers in noisy, cluttered images. We also show quantitative validation on the automatic segmentation results by comparing to expert solutions.

1 Introduction

Learning shape and appearance prior representations for an anatomical structure of interest has been central to many model-based medical image analysis algorithms. Although numerous methods have been proposed in the literature, most are hampered by the automated alignment and registration problem of training examples. In the seminal work of Active Shape and Appearance Models (ASM [3] and AAM [4]), models are built from analyzing the shape and appearance variabilities across a set of labelled training examples. Typically landmark points are carefully chosen and manually placed on all examples by experts to assure good correspondences. This assumption leads to a natural framework for alignment and statistical modeling, yet it also makes the training process time-consuming. Yang & Duncan [15] proposed a shape-appearance joint prior model for Bayesian image segmentation. They did not deal with registration of the training examples, however, and assumed the training data are already aligned.

A number of automated shape registration and model building methods have been proposed [5], [6], [7], [2]. These approaches either establish correspondences between geometric features, such as critical points of high curvature [7]; or find the "best" corresponding parametrization model by optimizing some criterion,

C. Barillot, D.R. Haynor, and P. Hellier (Eds.): MICCAI 2004, LNCS 3216, pp. 60–69, 2004.

such as minimizing accumulated Euclidean Distance [6], [2], Minimum Description Length [5], or Spline Bending Energy [2]. Both geometric feature based and explicit parameterization based registration methods are not suitable for incorporating region intensity information. In [10], the implicit shape representation using level sets is considered, and shape registration algorithms using this representation have been proposed [11,8].

Non-rigid registration is a popular approach to build statistical atlas and to model the appearance variations [14,1]. The basic idea is to establish dense correspondences between textures through non-rigid registration. However, few of the existing methods along this line are able to register training examples in arbitrary poses or to be coupled with shape registration.

In this paper, we introduce a new framework for learning statistical shape and appearance models that addresses efficiently the above limitations. This framework is an extension of our work on *MetaMorphs*, a new class of deformable models that have both shape and interior texture statistics [9], to incorporate prior information. We work in a unified shape and intensity space by implicitly representing shapes as “images” in the space of distance transforms. A novel stochastic chord-based matching algorithm efficiently aligns training examples through a similarity transformation (with rotation, translation and isotropic scaling), considering both shape and gray-level intensity information. Then the complementary local registration is performed by deforming a Free Form Deformations (FFD) control lattice to maximize mutual information between both “shape” and intensity images. we apply principal component analysis on the deformed FFD control lattices to capture variations in shape and on registered object interior textures to capture variations in intensity. This learning framework is applied to build a statistical model of the left ventricle as well as an articulated model of the whole heart in short-axis cardiac tagged MR images, then the prior models are used for automated segmentation in novel images.

2 Data Description and Algorithm Outline

2.1 Description of the Training Data

The training data are from spatial-temporal short-axis cardiac tagged MR images. A 1.5T GE MR imaging system is used to acquire the images, and an EGG-gated tagged gradient echo pulse sequence. Every 30ms, 2 sets of parallel short axis (SA) images are acquired; one with horizontal tags and one with vertical tags. These images are perpendicular to an axis through the center of the LV. A complete systole-diastole cardiac cycle is divided into 24 phases. We collected 180 images from 20 phases, discarding the beginning and ending two phases. An expert is asked to segment the epicardium (Epi), the left ventricle (LV) endocardium and the right ventricle (RV) endocardium from the images.

2.2 Learning and Segmentation Algorithm Outline

Our overall learning and segmentation framework is outlined by the flow-chart in Fig. (1). There are two major components in the framework. The procedures

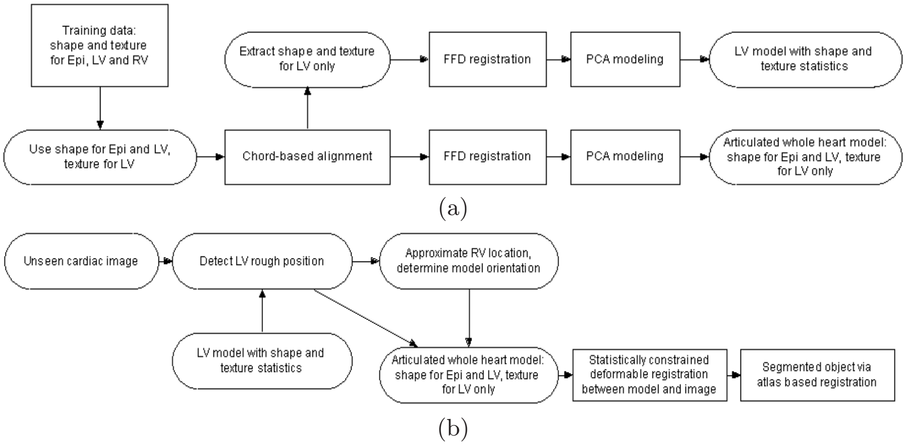
Training data: shape and texture for Epi, LV and RV
Extract shape and texture for LV only
FFD registration
PCA modeling
LV model with shape and texture statistics
Use shape for Epi and LV, texture for LV
Chord-based alignment
FFD registration
PCA modeling
Articulated whole heart model: shape for Epi and LV, texture for LV only
(a)
Unseen cardiac image
Detect LV rough position
Approximate RV location, determine model orientation
LV model with shape and texture statistics
Articulated whole heart model: shape for Epi and LV, texture for LV only
Statistically constrained deformable registration between model and image
Segmented object via atlas based registration
(b)

Fig. 1. (a) Learning framework. (b)Segmentation. (Rectangular boxes) Generic algorithmic steps. (Oval boxes) Specific designs for the cardiac segmentation problem.

described in the rectangular boxes are the algorithmic steps that are generic to all learning and segmentation problems. Additional procedures described in the oval boxes involve specific domain knowledge about the heart anatomy and the characteristics of the tagged MR images.

We utilize prior knowledge of the segmentation problem in devising the domain-specific procedures. First, since the images are acquired perpendicular to an axis through the center of the LV, the LV shapes appear relatively stable and near circular, and the LV interior intensities are also relatively homogeneous. Thus we learn a joint shape and texture model for the LV, which can be used for automated detection as well as segmentation. Second, for the alignment of training examples however, the LV's near-circular shape and homogeneous interior become unreliable in estimating the transformations. Thus we do the alignment based on an articulated heart model with both the epicardium (shape only) and LV endocardium (both shape and texture). Third, during segmentation in an unseen cardiac image, the LV shape and appearance model is used for automatically detecting the rough position of the heart. This position constraint and a Gabor-filter bank based method [12] are used to approximate the position of the RV. The positions of LV and RV centers determine the rough orientation of the whole heart model, which is thus transformed and further registered to the image using our statistically constrained deformable registration algorithm. The converged registration result defines the final segmentation of the heart chambers.

In the next two sections, we focus on presenting the generic algorithmic steps in our framework.

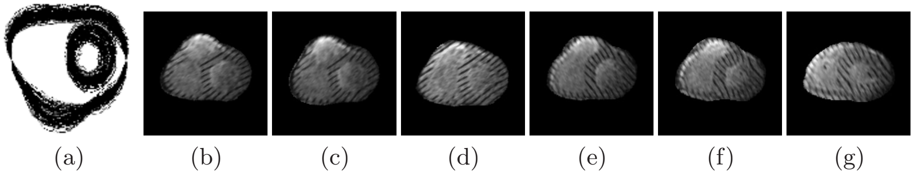

Fig. 2. Chord-based global alignment. (a) All aligned contours overlaid together. (b-g) Some examples of the globally aligned textures.

3 Learning the Shape and Appearance Statistical Model

3.1 Unified Shape and Intensity Feature Space

Within the proposed framework, we represent each shape using a Euclidean distance map. In this way, shapes are implicitly represented as “images” in the space of distance transforms where shapes correspond to the zero level set of the distance functions. The level set values in the shape embedding space is analogous to the intensity values in the intensity (appearance) space. As a result, for each training example, we have two ”images” of different modalities, one representing its shape and another representing its intensity (grey-level appearance). The shape and intensity spaces are conveniently unified this way.

We use the Mutual Information criterion as the similarity measure to be optimized. Suppose A and B are two training examples. Let us denote their level set value random variables in the shape space as X_S^A and X_S^B, and their intensity random variables in the intensity space as X_I^A and X_I^B. Then the similarity between the two examples in the joint shape and intensity space can be defined using a weighted form of Mutual Information:

$$\begin{aligned}\mathcal{M}_J(A,B) &= \mathcal{M}_S(A,B) + \alpha\mathcal{M}_I(A,B)\\ &= \mathcal{H}(X_S^A) + \mathcal{H}(X_S^B) - \mathcal{H}(X_S^A, X_S^B) + \alpha\big[\mathcal{H}(X_I^A) + \mathcal{H}(X_I^B) - \mathcal{H}(X_I^A, X_I^B)\big]\end{aligned} \quad (1)$$

where $\mathcal{H}$ represents the differential entropy and α is a constant balancing the contributions of shape and intensity in measuring the similarity. In our experiments, we have set the values for α between $[0.2, 0.6]$. For brevity, we will use $\mathcal{M}_J$ to represent the mutual information in the joint space, $\mathcal{M}_S$ in the shape space, and $\mathcal{M}_I$ in the intensity space.

3.2 Chord-Based Global Alignment

When aligning the training examples under a common reference frame, we pursue an alignment that is ”optimal” in the sense that the mutual information criterion in the joint feature space is optimized. Our solution is a novel alignment algorithm based on the correspondences between chords. Given a training example, A, suppose its un-ordered set of boundary points is $\{P_i^A = (x_i^A, y_i^A)\}, i = 1, ..., m$, a chord is a line segment joining two distinct boundary points. Our observations here are: **(i)** each of the total $\frac{1}{2}m(m-1)$ chords defines an internal, normalized reference frame for the example, in which the midpoint of the chord is the origin, the chord is aligned with the x axis, and the chord length is scaled to be of unit length 1.0; **(ii)** One pair of chord correspondences between two examples is sufficient to recover an aligning similarity transformation. So the basic idea of our algorithm is that, instead of finding correspondences between individual feature points as in most other matching algorithms, we find correspondences between chords, hence the correspondences between internal reference frames of two examples, and align the examples by aligning the best matching pair of internal reference frames.

Suppose we have an example A, as describe above, and a second example B with unordered set of boundary points $\{P^B_{i'} = (x^B_{i'}, y^B_{i'})\}, i' = 1, ..., n$. Let us denote a chord joining two points P_i and P_j ($i \neq j$) as c_{ij}. The matching algorithm can be outlined as follows:

1. For every chord c^A_{ij} on example A,
   Find its corresponding chord $c^B_{i'j'}$ on example B as:
   $$c^B_{i'j'} = argmax_{c^B_{kl}} \left[\mathcal{M}_S\big(A_{ij}, B_{kl}(c^B_{kl})\big) + \alpha \mathcal{M}_I\big(A_{ij}, B_{kl}(c^B_{kl})\big) \right] \tag{2}$$
   where A_{ij} is the representation of A in its internal reference frame F^A_{ij} defined by the chord c^A_{ij}; $B_{kl}(c^B_{kl})$ represents B in its internal reference frame F^B_{kl} defined by the chord c^B_{kl}.
2. Among all hypothesized alignments between A and B, suppose the one based on a pair of corresponding chords, c^A_{IJ} and $c^B_{I'J'}$, gives rise to the maximal mutual information in the joint shape and intensity space: $\big[\mathcal{M}_S\big(A_{IJ}, B_{I'J'}\big) + \alpha \mathcal{M}_I\big(A_{IJ}, B_{I'J'}\big)\big]$, then the internal reference frames defined by this pair of chords, F^A_{IJ} and $F^B_{I'J'}$, are chosen to be the best matching reference frames.
3. Align examples A and B into a common reference frame by aligning the two reference frames F^A_{IJ} and $F^B_{I'J'}$ using a similarity transformation.

In practice, we find the chord correspondences using a stochastic algorithm based on the Chord Length Distribution (CLD) [13]. The algorithm is very efficient by considering only those chords with lengths greater than a certain percentile in the CLD of each example. On average, the computation time for aligning two examples on a $3GHz$ PC is around $15ms$ using the $85th$ percentile in our experiments. Furthermore, the algorithm can handle structures of arbitrary topology since it does not require the explicit parameterization of shapes. It is also invariant to scaling, rotation and translation, thus the training examples can be aligned robustly regardless of their initial poses. In Fig. 2, we show the aligned examples for our articulated whole heart model. Here we randomly pick one example as the atlas, and align all other examples to it.

3.3 Local Registration Using FFD and Mutual Information

After global alignment, the next step towards building a statistical model is to solve the dense correspondences problem. We proposed a nonrigid shape registration framework for establishing point correspondences in [8]. In this paper, we

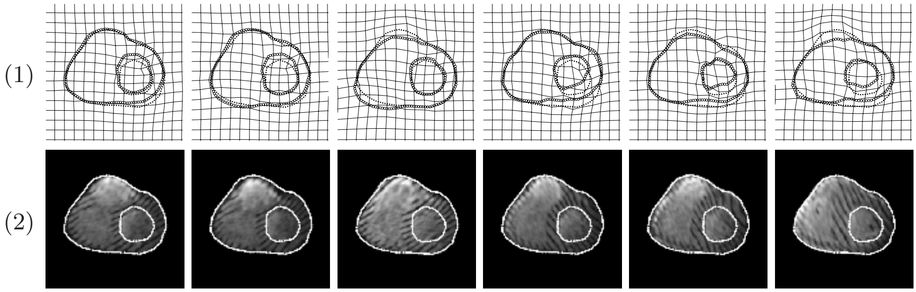

Fig. 3. Local FFD registration between training examples. (1) Each training shape (points drawn in circles) deforms to match a target mean atlas (points drawn in dots). The FFD control lattice deformations are also shown. (2) The registered textures. Note that each training texture is non-rigidly deformed based on FFD and registered to a mean texture. All textures cover a same area in the common reference frame.

extend this framework to perform nonrigid registration in the unified shape and intensity space, thus achieving simultaneous registration on both shapes and textures of the training examples. This joint registration provides additional constraints on the deformation field for the large area inside the object.

We use a space warping technique, the Free Form Deformations (FFD), to model the local deformations. The basic idea of FFD is to deform an object by manipulating a regular control lattice overlaid on its volumetric embedding space. We consider an Incremental cubic B-spline FFD in which dense registration is achieved by evolving a control lattice P according to a deformation improvement δP. Let us consider a regular lattice of control points $P_{m,n} = (P^x_{m,n}, P^y_{m,n})$; $(m,n) \in [1,M] \times [1,N]$ overlaid to a region $\Gamma_c = \{\mathbf{x}\} = \{(x,y) | 1 \leq x \leq X, 1 \leq y \leq Y\}$ that encloses a training example. Suppose the initial configuration of the control lattice is P^0, and the deforming control lattice is $P = P^0 + \delta P$. Then the incremental FFD parameters are the deformations of the control points in both x and y directions: $\boldsymbol{\Theta} = \{(\delta P^x_{m,n}, \delta P^y_{m,n})\}$. The incremental deformation of a pixel $\mathbf{x} = (x,y)$ given the deformation of the control lattice from P^0 to P, is defined in terms of a tensor product of Cubic B-spline: $\delta L(\boldsymbol{\Theta}; \mathbf{x}) = \sum_{k=0}^{3} \sum_{l=0}^{3} B_k(u) B_l(v) (\delta P_{i+k,j+l})$, where $i = \lfloor \frac{x}{X} \cdot (M-1) \rfloor + 1,\ j = \lfloor \frac{y}{Y} \cdot (N-1) \rfloor + 1$; $\delta P_{i+k,j+l}$ consists of the deformations of pixel $\mathbf{x}$'s sixteen adjacent control points; $B_k(u)$ is the k^{th} basis function of the cubic B-spline.

To register an atlas T and a rigidly aligned training example R, we consider a sample domain Ω in the common reference frame. The mutual information criterion defined in the joint shape and intensity space can be considered to recover the deformation field $\delta L(\boldsymbol{\Theta}; \mathbf{x})$ that registers R and T:

$$\mathcal{M}_J\big(R, T(\delta L(\boldsymbol{\Theta}))\big) = \mathcal{M}_S\big(R(\Omega), T(L(\boldsymbol{\Theta}; \Omega))\big) + \alpha \mathcal{M}_I\big(R(\Omega), T(L(\boldsymbol{\Theta}; \Omega))\big) \qquad (3)$$

In the equation, $L(\boldsymbol{\Theta}; \Omega)$ represents the deformed domain of the initial sample domain Ω, i.e. $L(\boldsymbol{\Theta}; \mathbf{x}) = \mathbf{x} + \delta L(\boldsymbol{\Theta}; \mathbf{x})$, for any $\mathbf{x} \in \Omega$. A gradient descent optimization technique is used to maximize the mutual information criterion, and

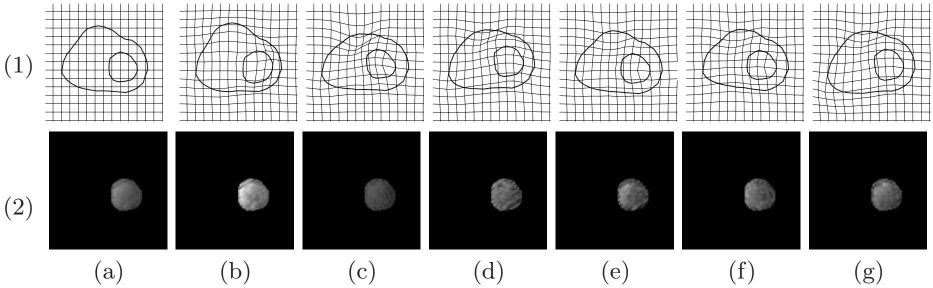

Fig. 4. PCA modeling. (1.a) The mean FFD control lattice configuration and mean shape. (1.b-c) Varying first mode of FFD deformations: -2σ reconstruction in (b) and 2σ in (c). (1.d-e) Second mode of FFD. (1.f-g) Third mode of FFD. (2.a) The mean LV texture (based on pixel-wise correspondences). (2.b-c) Varying first mode of LV texture. (2.d-e) Second mode of LV texture. (2.f-g) Third mode of LV texture.

to recover the parameters of the smooth, one-to-one registration field δL. Then dense pixel-wise correspondences can be established between each point $\mathbf{x}$ on example R, with its deformed position $\hat{L}(\mathbf{x})$ on the atlas T. The correspondences are valid on both the "shape" images and the intensity images. We show some example results using this local registration algorithm in Fig. (3).

3.4 Statistical Modeling of Shape and Appearance

After registration in the joint shape and intensity space, we apply Principal Component Analysis (PCA) on the deformed FFD control lattices to capture variations in shape. The feature vectors are the coordinates of the control points in x and y directions in the common reference frame. We also use PCA on the registered textures to capture variations in intensity. Here the feature vectors are the image pixel intensities from each registered texture. Fig. 4 illustrates the mean atlas and three primary modes of variation for both the shape deformation fields (Fig. (4).1) and intensities (Fig. (4).2). The shape model uses the articulated heart model with Epi and LV, and the texture model is for the LV interior texture only (due to tagging lines in heart walls and RV irregularity).

4 Segmentation via Statistically Constrained Registration

Given an unseen image, we perform segmentation by registering the learned prior model with the image based on both shape and texture. The mutual information criterion to be optimized is the same as Equation 3, except that here R consists of the new intensity image and a "shape" image, which is derived from the unsigned distance transform of the edge map (computed by the Canny edge detector). Another difference from the learning process is that, during optimization, instead of using directly the recovered FFD parameter increments to deform the prior model, we back-project the parameter increments to the PCA-based feature

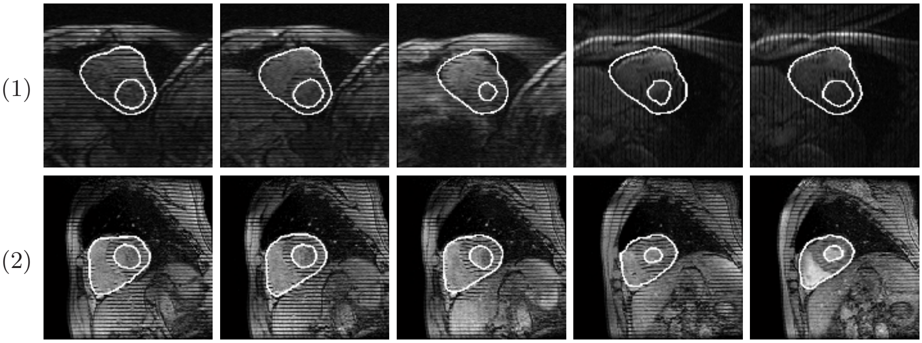

Fig. 5. Coupled prior based segmentation results on two novel tagged MR image sequences. (1) Example results from sequence 1. (2) Example results from sequence 2.

space, and magnitudes of the allowed actual parameter changes are constrained to have a 2σ upper bound. This scheme is similar to that used in AAM.

4.1 Results and Validation

Using the statistical model learned as shown in Fig. 4, we conduct automated segmentation via statistically constrained registration on two novel sequences of 4D spatial-temporal tagged MR images. Each sequence consists of 24 phases, with 16 slices (images) per phase. Since we do not use the first and last two phases in the new sequences, we have 320 images for testing from each sequence. The segmentation framework is depicted in Fig. 1.b. Example segmentation results are shown in Fig. 5. In all the experiments, following the LV detection and rough model pose estimation, the registration-based segmentation process takes less than 2 seconds to converge for each image on a $3GHz$ PC workstation.

Quantitative validation is performed by comparing the automated segmentation results with expert solutions. Denote the expert segmentation in the images as ℓ_{true}, and the results from our method as ℓ_{prior}. We define the false negative fraction (FNF) to indicate the fraction of tissue that is included in the true segmentation but missed by our method: $FNF = \frac{|\ell_{true} - \ell_{prior}|}{|\ell_{true}|}$. The false positive fraction (FPF) indicates the amount of tissue falsely identified by our method as a fraction of the total amount of tissue in the true segmentation: $FPF = \frac{|\ell_{prior} - \ell_{true}|}{|\ell_{true}|}$. And the positive fraction (TPF) describes the fraction of the total amount of tissue in the true segmentation that is overlapped with our method: $TPF = \frac{|\ell_{true} \cap \ell_{prior}|}{|\ell_{true}|}$. On the novel tagged MR sequence 1, our segmentation results produce the following average statistics: $FNF = 2.4\%, FPF = 5.1\%, TPF = 97.9\%$. On the novel sequence 2, the average statistics are: $FNF = 2.9\%, FPF = 5.5\%, TPF = 96.2\%$.

5 Discussion and Conclusions

In this paper, we have proposed a novel, generic algorithm for learning coupled prior shape and appearance models. Our main contributions in this paper are three folds. First, we work in a unified shape and intensity feature space. Second, we develop a novel stochastic chord-based matching algorithm that can efficiently align training examples in arbitrary poses, considering both shape and texture information. Third, a local registration algorithm based on FFD and mutual information performs registration both between shapes and between textures simultaneously. In our future work, we will learn a 3D coupled prior shape and texture model for the heart in tagged MR images. It is also important to explore the use of other learning techniques, such as Independent Component Analysis, in our framework.

Acknowledgements. The authors are grateful to Dr. Leon Axel for providing the MR cardiac images. We also would like to acknowledge the stimulating discussions with Dr. Nikos Paragios. Support for this research was provided by the NSF-0205671 grant.

References

1. M. Chen, T. Kanade, D. Pomerleau, and J. Schneider. 3-D deformable registration of medical images using a statistical atlas. In *MICCAI 1999*, pp. 621-630, 1999.
2. H. Chui, and A. Rangarajan. Learning an atlas from unlabeled point-sets. In *IEEE Workshop on Mathematical Methods in Biomedical Image Analysis*, 2001.
3. T. F. Cootes, C. J. Taylor, D. H. Cooper, and J. Graham. Active Shape Models - their training and application. In *Computer Vision and Image Understanding*, Vol. 61, No. 1, pp. 38-59, 1995.
4. T. F. Cootes, G. J. Edwards, and C. J. Taylor. Active Appearance Models. In *Proc. European Conf. on Computer Vision*, Vol. 2, pp. 484-498, Springer, 1998.
5. R.H. Davies, T.F. Cootes, J.C. Waterton, and C.J. Taylor. An efficient method for constructing optimal statistical shape models. In *MICCAI 2001*, pp. 57-65, 2001.
6. N. Duta, A. K. Jain, and M.-P. Dubuisson-Jolly. Learning 2D shape models. In *IEEE Conf. on Computer Vision and Pattern Recognition*, Vol. 2, pp. 8-14, 1999.
7. A. Hill, C. J. Taylor, and A.D. Brett. A framework for automatic landmark identification using a new method of non-rigid correspondences. In *IEEE Trans. on Pattern Analysis and Machine Intelligence*, Vol. 22, No. 3, pp. 241-251, 2000.
8. X. Huang, N. Paragios, and D. Metaxas. Establishing local correspondences towards compact representations of anatomical structures. In *MICCAI 2003*, LNCS 2879, pp. 926-934, 2003.
9. X. Huang, D. Metaxas, and T. Chen. MetaMorphs: deformable shape and texture models. In *IEEE Conf. on Computer Vision and Pattern Recognition*, 2004.
10. M.E. Leventon, E.L. Grimson, and O. Faugeras. Statistical shape influence in Geodesic Active Contours. In *IEEE Conf. on Computer Vision and Pattern Recognition*, Vol. 1, pp. 1316-1323, 2000.
11. N. Paragios, M. Rousson, and V. Ramesh. Matching distance functions: a shape-to-area variational approach for global-to-local registration. In *European Conf. on Computer Vision*, pages II:775–790, 2002.

12. Z. Qian, A. Montillo, D. Metaxas, and L. Axel. Segmenting cardiac MRI tagging lines using gabor filter banks. In *25th Int'l Conf. of IEEE EMBS*, 2003.
13. S. P. Smith, and A. K. Jain. Chord distribution for shape matching. In *Computer Graphics and Image Processing*, Vol. 20, pp. 259-271, 1982.
14. D. Rueckert, A.F. Frangi, and J.A. Schnabel. Automatic construction of 3D statistical deformation models using non-rigid registration. In *MICCAI 2001*, LNCS 2208, pp. 77-84, 2001.
15. J. Yang, and J.S. Duncan. 3D image segmentation of deformable objects with shape-appearance joint prior models. In *MICCAI 2003*, pp. 573-580, 2003.