Database-Guided Segmentation of Anatomical Structures with Complex Appearance

B. Georgescu X. S. Zhou D. Comaniciu A. Gupta

Integrated Data Systems Department
Siemens Corporate Research
Princeton, NJ 08540

Computer Aided Diagnosis
Siemens Medical Solutions
Malvern PA, 19355

Abstract

The segmentation of anatomical structures has been traditionally formulated as a perceptual grouping task, and solved through clustering and variational approaches. However, such strategies require the a priori knowledge to be explicitly defined in the optimization criterion, e.g., "high-gradient border", "smoothness", or "similar intensity or texture". This approach is limited by the validity of underlying assumptions and cannot capture complex structure appearance. This paper introduces ***database-guided segmentation*** *as a new data-driven paradigm that directly exploits expert annotation of interest structures in large medical databases. Segmentation is formulated as a two-step learning problem. The first step is structure detection where we learn how to discriminate between the object of interest and background. The resulting classifier based on a boosted cascade of simple features also provides a global rigid transformation of the structure. The second step is shape inference where we use a sample-based representation of the joint distribution of appearance and shape annotations. To learn the association between the complex appearance and shape we propose a feature selection mechanism and the corresponding metric. We show that the selected features are better than using directly the appearance and illustrate the performance of the proposed method on a large set of ultrasound heart images.*

1. Introduction

Accurate localization of complex structures is important in many computer vision applications ranging from facial feature detection [5, 7] to segmentation of anatomical structures in medical images [3, 13, 19, 21]. Availability of large databases with expert annotation of the interest structures makes a learning approach more attractive than classical approaches of solving perceptual grouping tasks [17] through clustering or variational formulations [12, 14]. This is especially important when the underlying image structures do not have clear border definition, show complex appearance with large amounts of noise, or when there is a relatively large variation between expert's own annotations.

The difficulty of the segmentation task is illustrated by the images in Figure 1. They represent ultrasound images of the heart and the goal is to delineate the left ventricle border (endocardium) [3, 13, 23]. Automated segmentation of echocardiographic images has proved to be challenging due to large amount of noise, signal drop-out and also due to large variations between the appearance, configuration and shape of the left ventricle.

Segmentation is one of the most important low-level image processing methods and has been traditionally approached as a grouping task based on some homogeneity assumption. For example clustering methods have been used to group regions based on color similarity [4] or graph partitioning methods have been used to infer global regions with coherent brightness, color and texture [18]. Alternatively the segmentation problem can be cast in an optimization framework as the minimization of some energy function. Concepts such as "high-gradient border", "smoothness", or "similar intensity or texture" are encoded as region or boundary functionals in the energy function and minimized through variational approaches [12, 14].

However as the complexity of the targeted segmentation increases, it is more difficult to encode prior-knowledge into the grouping task. Learning has become more important for segmentation and there are methods that infer rules for the grouping process conditioned by the user input [2, 17].

On a different approach, active appearance models [6] use registration to infer the shape associated with the current image. However, modeling assumes a Gaussian distribution of the joint shape-texture space and requires initialization close to the final solution. Alternatively, characteristic points can be detected in the input image [7] by learning a classifier through boosting [8, 20].

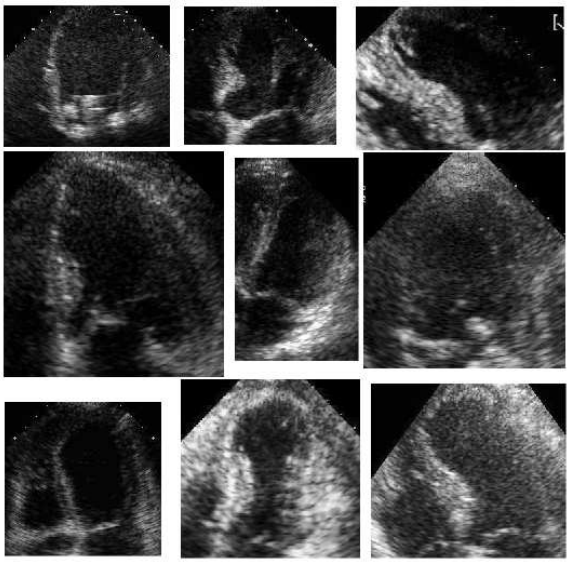


Figure 1. Samples of ultrasound heart images. Note the large variation in appearance and shape.

We propose a method to directly exploit expert annotation of the interest structure in large databases by formulating the segmentation as a two-step learning problem. The first step is to learn a discriminative function between the appearance of the object of interest and the background. The second step is to learn the discriminative features that best associates the shapes to different appearances of the object, and infer the most likely shape. The advantages are that complex prior knowledge is implicitly encoded and the resulting procedure is able to process the input images under real-time constraints.

The paper is structured as follows. Section 2 gives an overview of our approach and Section 3 presents the first learning step for structure detection. We also propose a principled solution for handling invalid image regions. Section 4 presents our shape inference procedure and the associated feature selection. Experimental results on a large number of ultrasound heart images are illustrated in Section 5 and we conclude in Section 6.

2. Database-Guided Segmentation

We use the term **database-guided segmentation** to underline the process of implicitly encoding the prior knowledge embedded in expert annotated databases. We decompose this process into two learning tasks. The first is *structure detection* where we learn to discriminate between the appearance of the interest object and the background. The second is *shape inference* where we learn to discriminate between appearances corresponding to different shapes and derive the most likely shape given an object appearance.

Both tasks use the same pool of a large set of simple features for appearance representation. For structure detection we select the features to solve a two-class problem using a boosted cascade of weak-classifiers. As a result we find the global rigid transformation for the possible locations of the interest object. For shape inference we propose a feature selection procedure to encode the joint distribution of appearance and shape. The local nonrigid shape deformation and the final segmentation is derived through a nearest-neighbor approach by using a sample based representation of the joint distribution.

3. Structure Detection

Some of the most successful real-time object detection methods are based on boosted cascade of simple features [1, 15, 20]. By combining the response of a selected number of simple classifiers through boosting [8], the resulting strong classifier is able to achieve high detection rates and is capable to process images in real-time. The advantage of boosting as oppose to traditional Gaussian appearance models is that it can deal with complex distributions such as multi-modal distributions, which is the case for our application. Boosting is also much faster than other non-linear alternatives such as kernel support vector machines [11].

The simple features used in our method are rectangle features similar to Haar basis functions [16] and we use AdaBoost to learn a two-class classifier able to distinguish between the set of positive appearance examples (containing the object) and the set of negative examples. For complete technical details of this boosting approach to detection we refer the reader to [20].

Here we address two of the problems that directly affects the stability of the object appearance representation. First we propose a weighted structure alignment to increase the influence of stable landmark points. Second we introduce a solution to eliminate the influence of invalid image regions in feature computation.

3.1. Weighted Structure Alignment

As a data preprocessing step the location parameters associated with the detection have to be eliminated from the object appearance training set. To generate the set of positive examples we first eliminate the variations due to global rigid transformations through Procrustes alignment [5, 9]. Hence, the object appearance is normalized with respect to translation, rotation and scale.

An important issue that has been largely overlooked by the existing work on Procrustes shape alignment is the varying stability (or detectability) of the landmark points. Intuitively, points that are more stable or more detectable should receive a higher weight during the least-square alignment process. Some work has addressed this issue “sub-

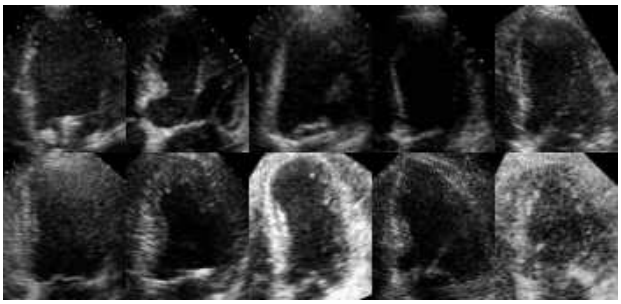


Figure 2. Shape normalized images used for training.

consciously". For example, for face detection, some align training faces by eye corners only, instead of aligning all facial features including points from mouth and nose, which are much less stable. We argue that the optimal solution should be a weighted Procrustes alignment process, with the weights reflecting feature stability.

In this work, we quantify the stability of control points based on their "detectability" using local detectors that learn from the neighboring appearance of each control point. In our case, of left ventricle detection in echocardiography, local detectors perform much better near the basal region of the heart than those near the apical or lateral wall regions. This is in agreement with the nature of ultrasound images: in apical views the reflection from the basal region is much stronger and more stable than the reflection from the apical or lateral wall, where signal dropout is most likely to happen. Nevertheless, how to *optimally* select the weights remains an open question: one can envision an optimization formulation incorporating also the down-stream detection algorithm. This is among our future research efforts.

With a weight matrix W, the minimized criterion for aligning shapes is given by the Mahalanobis distance:

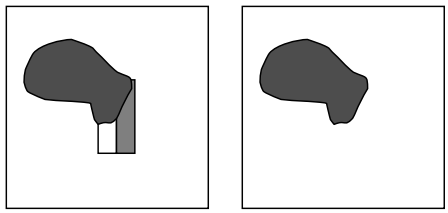
$$\mathcal{J}_{GPA} = ||s_i \mathrm{R}_i \mathbf{c}_i + \mathbf{t}_i - \overline{\mathbf{c}}||_{\mathrm{W}} \qquad (1)$$

where $\mathbf{c}_i$ represents the i^{th} shape control points and $s_i, \mathrm{R}_i, \mathbf{t}_i$ represents scale, rotation and translation; $\overline{\mathbf{c}}$ is the mean shape. This is solved iteratively through weighted least squares.

Figure 2 illustrate some of the aligned positive appearance examples used for training. Compared with Figure 1 the global rigid shape transformations are canceled out and the object appearance started to have some "order" as all instances share the same mean shape. The negative set is generated randomly from the same images by varying the parameters of the global transformations.

3.2. Feature Computation for Invalid Image Regions

The simple features used for the weak classifiers are rectangle features which are similar to Haar basis functions [16]



(a) (b)

Figure 3. (a) Occluded rectangle feature; (b) Invalid mask

and have been proved to provide a rich image representation for object detection [15, 20].

However the existing methods do not address the detection problem under the presence of invalid image regions (occlusions). The erroneous response of a simple (weak) classifier will influence negatively the detection outcome. We propose a method to eliminate the influence of known invalid regions in the object detection process. Our approach has minimal added computation and correctly estimate the simple classifiers response using only the valid image information.

Rectangle features provide an over complete basis, for example for a base region size of 24x24 pixels the number of features is 180,000 [20]. One of the advantages of rectangle features is computational speed. By using an intermediate representation known as the "integral image", a feature value can be calculated through a fixed number of operations (for example a two-rectangle feature requires six array references). The integral image contains at each location the sum of intensities of the pixels above and to the left and it can be computed in one pass over the input image.

It is clear that an invalid intensity value for a pixel will yield an incorrect estimate for the feature using that pixel (Figure 3a). If the valid image mask is available we can use it to eliminate the contribution of the invalid pixels to the feature value. The mask is available when images are taken in controlled environments or it can be inferred from the data (for example in surveillance applications the static background is known, in ultrasound images the fan location can be computed or analysis of time variations can yield the static regions). If we set to zero the intensity for the invalid pixels, the rectangle sum will no longer be influenced by incorrect values. However due to the missing data the sum will be "unbalanced". If there are no missing values, the rectangle sum is proportional to the mean intensity value, therefore we can approximate the mean value if we know the number of valid intensities (when occlusions are present). The number of valid pixels can be easily found by first computing an equivalent map: the "integral mask". Given the valid pixels mask M with boolean values (1 for valid pixel, 0 for invalid or occluded) then the integral mask IM contains the number of valid pixels above and to the

left of the current location (x_0, y_0)

$$IM(x_0, y_0) = \sum_{x \leq x_0, y \leq y_0} M(x, y) \; . \tag{2}$$

Similarly to the integral image the number of valid pixels in a rectangle can be computed from the integral mask in the same number of operations.

The equivalent feature value will be given by a weighted difference between the sum of the intensities I in the "positive" and "negative" image regions. If we denote by R_+ the region where the pixels intensities contribute with a positive value and by R_- with a negative value, the feature value f is

$$f = \frac{n_-}{N} \sum_{(x,y) \in R_+} I(x, y) - \frac{n_+}{N} \sum_{(x,y) \in R_-} I(x, y) \; , \tag{3}$$

where n_-, n_+ denote the number of valid pixels for negative and positive regions respectively, each containing N pixels (note that a similar formulation exist for regions containing a different number of pixels). It can be easily checked that when all pixels are valid, the feature value is equal to the original and the value goes to 0 if one of the region becomes more occluded.

Compared to the original algorithm, the proposed method requires the additional computation of the integral mask and two more multiplications/feature. However, the operations do not modify the complexity of the algorithm and no singularities appear in feature value computation. The advantage is that by using only the valid information, the strong classifier is not influenced by incorrect data, thus increasing the detection performance.

4. Shape Inference

The result of the first classification task is a set of possible locations of the structure of interest and the likelihood of a particular appearance instance is measured by the detection score. We also can say that so far, the associated shape is the mean shape used in alignment, deformed by the corresponding rigid global transformation.

The problem that we try to solve now is: given an appearance of the interest structure, what is the most likely associated shape? For this task we propose to directly use the expert's structure annotations by maintaining a sample based representation of the joint distribution of appearance and shape. We do not use a mixture of Gaussian approach due to the large variations in the joint space of appearance and shape.

To infer the shape we use a nearest-neighbor approach by finding the closest prototypes in the database. To measure similarity we can use the distance between the image intensities directly or we can even use the features selected for the detection classifier and infer the shape through the nearest-neighbor method. The similarity distance is equivalent to the probability of observing that appearance instance given the training set, and therefore is combined with the detection score to yield the most likely segmentation.

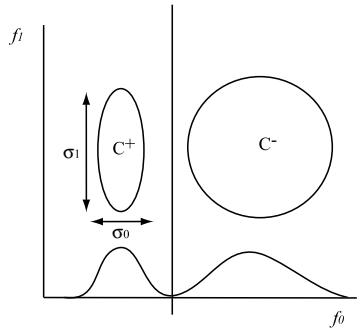


Figure 4. Feature importance: f_0 better for classification, f_1 better for within-class variance.

However because the detection classifier was trained to distinguish between the positive and the negative examples, the selected features for detection are the best to maximize the classes separability and they do not necessarily express the "within class" variability. This is illustrated in Figure 4 where feature f_1 represents better the variability of the positive class than f_0 which is better for detection.

We propose a simple feature selection procedure from the same feature pool used in the detection stage, to better capture the within-class variability. The approach can be actually seen as a boosting approach for improving nearest-neighbor classification.

4.1. Forward Sequential Feature Selection

The problem is to select the features that best associate the respective appearance with the corresponding shape. Note that the number of initial features can be quite large ($\sim$200.000) and also the number of samples is large ($\sim$5000), therefore the selection procedure has to be simple and the evaluation criterion fast to compute.

At run-time we are given the features and we have to infer the associated shape from the joint distribution $(\mathbf{f}, \mathbf{c})$ where we denote by $\mathbf{f}$ the appearance feature vector and $\mathbf{c}$ the corresponding shape. The feature selection criteria in this case is the one that *minimizes the distance between the inferred shape and the real shape*. In other words the distance between shapes

$$d(\mathbf{c}_q, \mathbf{c}_r) = (\mathbf{c}_q - \mathbf{c}_r)^\top (\mathbf{c}_q - \mathbf{c}_r) \tag{4}$$

is emulated through the distance between the feature vectors:

$$d(\mathbf{f}_q, \mathbf{f}_r) = (\mathbf{f}_q - \mathbf{f}_r)^\top \Sigma (\mathbf{f}_q - \mathbf{f}_r) \tag{5}$$

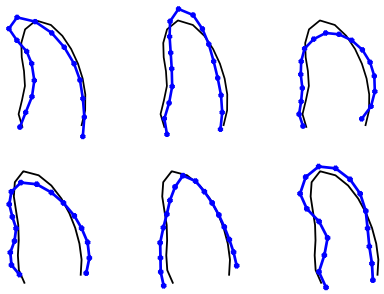


Figure 5. Three of the modes for the A4C set (top) and for the A2C set (bottom) relative to the mean shape.

where $(\mathbf{f}_q, \mathbf{c}_q), (\mathbf{f}_r, \mathbf{c}_r)$ represent the vector of the query and respectively the reference, and Σ is the linear metric associated with the feature vector space.

We propose a simple selection procedure based on forward sequential feature selection with the criteria based on class separability. We want to emulate in the feature space as close as possible the distance in the shape space. Thus, we cluster the data in the shape space in an representative number of shape clusters K. Because our criterion to be minimized is the euclidean distance between shapes, we can use a simple K-means algorithm for clustering, which assumes an isotropic Gaussian distribution in the shape space. This will partition the original feature vectors in K classes. The number of clusters is not critical because is only used to impose the shape space metric to the feature space. Figure 5 illustrates some of the shape modes obtained through clustering relative to the mean shape for two datasets (A4C and A2C) that we used for training.

The problem now is to find the best subset of the original feature set that best separate the detected classes. To measure class separability we use the well known criteria based on the between class and within class variance

$$\mathcal{J}_{sel} = \mathrm{trace}\left(\mathrm{S}_w^{-1}\mathrm{S}_b\right) \tag{6}$$

where S_w is the within class variance and S_b is the between class variance.

Ideally we would like to compute this matrices nonparametrically because the points belonging to one cluster might not be grouped in the feature space, but the class still is separable from the others (multiple modes).

However because of the large number of features and a potential large number of training samples, the nonparametric computation is not feasible. Thus, under a normal distribution assumption, we compute the matrices as:

$$\mathrm{S}_b = \sum_{k=1}^{K} \pi_k (\bar{\mathbf{f}}_k - \bar{\mathbf{f}})(\bar{\mathbf{f}}_k - \bar{\mathbf{f}})^\top \tag{7}$$

and

$$\mathrm{S}_w = \sum_{k=1}^{K} \pi_k \Sigma_\mathrm{k} \tag{8}$$

where $\pi_k, \bar{\mathbf{f}}_k, \Sigma_k$ are the probability, mean and covariance of class k and $\bar{\mathbf{f}}$ the global feature mean.

The standard forward sequential feature selection approach is used to determine the relevant features. The procedure starts with an empty set. At each step each feature is tested and the one yielding the largest increase in the criterion function (6) is added to the current set. The selection is stopped when no significant change in the criterion occurs. Note that in our case it would be impossible to use backward feature selection which would be more effective in discovering combination of features. This is because both the number of features and the number of samples is very large ($\sim$200.000 features and $\sim$5000 samples corresponding to a size of $\sim 10^{10}$ for the variance matrices) .

The shape of the discriminating metric matrix Σ will be given [10, pp. 429] by the within and between-class covariance matrices as

$$\begin{aligned} \Sigma &= \mathrm{S}_w^{-1/2}\left(\mathrm{S}_w^{-1/2}\mathrm{S}_b\mathrm{S}_w^{-1/2} + \epsilon\mathrm{I}\right)\mathrm{S}_w^{-1/2} \\ &= \mathrm{S}_w^{-1/2}\left(\mathrm{S}_b^{*} + \epsilon\mathrm{I}\right)\mathrm{S}_w^{-1/2} \end{aligned} \tag{9}$$

which spheres the space with respect to S_w and then it stretches the space in the null-space of S_b^*. The parameter ϵ rounds the neighborhood.

4.2. Nonrigid Shape Inference

Segmentation starts with an input image sequence on which the appearance candidates (detections) are determined through a hierarchical search in the discretized rigid transformation parameter space (translation, scale, rotation and image frame). The search is refined for parameters corresponding to positive responses with a large error margin of the detection classifier. We maintain multiple hypothesis for the appearance candidates for which we infer the shape. The shape $\hat{\mathbf{c}}$ is computed through a kernel smoother given by the Nadaraya-Watson kernel-weighted average [10, pp. 166]

$$\hat{\mathbf{c}}(\mathbf{f}) = \frac{\sum_{i=1}^{N} K_k(\mathbf{f}, \mathbf{f}_i)\mathbf{c}_i}{\sum_{i=1}^{N} K_k(\mathbf{f}, \mathbf{f}_i)} \tag{10}$$

where $(\mathbf{f}_i, \mathbf{c}_i)$ is the i^{th} sample of the N prototypes and $\mathbf{f}$ the query feature vector. For the kernel K_k we use the Epanechnikov quadratic kernel

$$K_k(\mathbf{f}, \mathbf{f}_i) = \begin{cases} 3/4\left[1 - \frac{d(\mathbf{f},\mathbf{f}_i)}{d\left(\mathbf{f},\mathbf{f}_{[k]}\right)}\right] & \text{if } \frac{d(\mathbf{f},\mathbf{f}_i)}{d\left(\mathbf{f},\mathbf{f}_{[k]}\right)} \leq 1 \\ 0 & \text{otherwise} \end{cases} \tag{11}$$

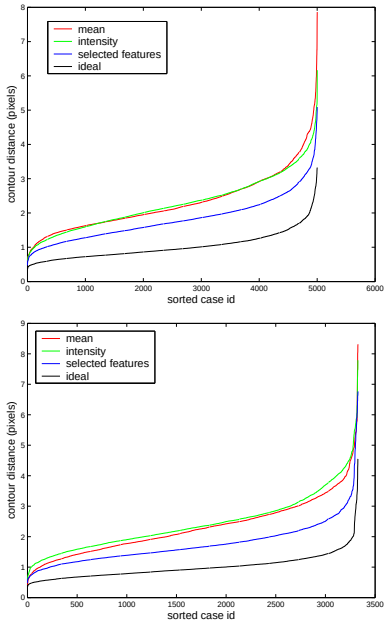


Figure 6. Point error between the predicted shape and real shape by using the mean shape, image intensity and selected features relative to the minimum error for the A4C set (top) and the A2C set (bottom).

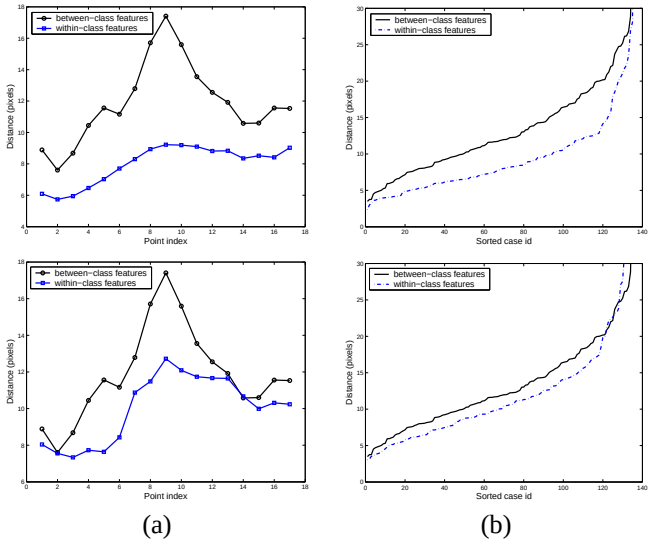
where the distance is given by (5) and $\mathbf{f}_{[k]}$ is the k^{th} prototype closest to the query.

The effect of using a kernel smoother is that it decreases the estimate variance, which is high for nearest-neighbor approach, at the expense of a higher bias. The final selected candidate is the one with a minimum detection score and small neighbor distance.

5. Experimental Results

The performance of the proposed method was tested on two annotated sets of ultrasound heart image sequences. The A4C set contains apical 4 chamber views of the heart (206 videos) and the A2C set has apical 2 chamber views of the heart (136 videos). The database has 5007 samples for the A4C set and 3330 samples for the A2C set. We characterize the associated shapes by a number of 17 control points.

The first experiment shows the effectiveness of the selected features relative to using directly the image appearance, or using the features selected by boosting for detection. For this experiment we consider only the joint appearance-shape distribution, that is, the images are rigidly aligned. In Figure 6 we plot the distance between the inferred shape and the true shape by leave-one-out method. Note that we exclude from the set *all* the images that belong to the same video and no two videos are from the same



(a) (b)

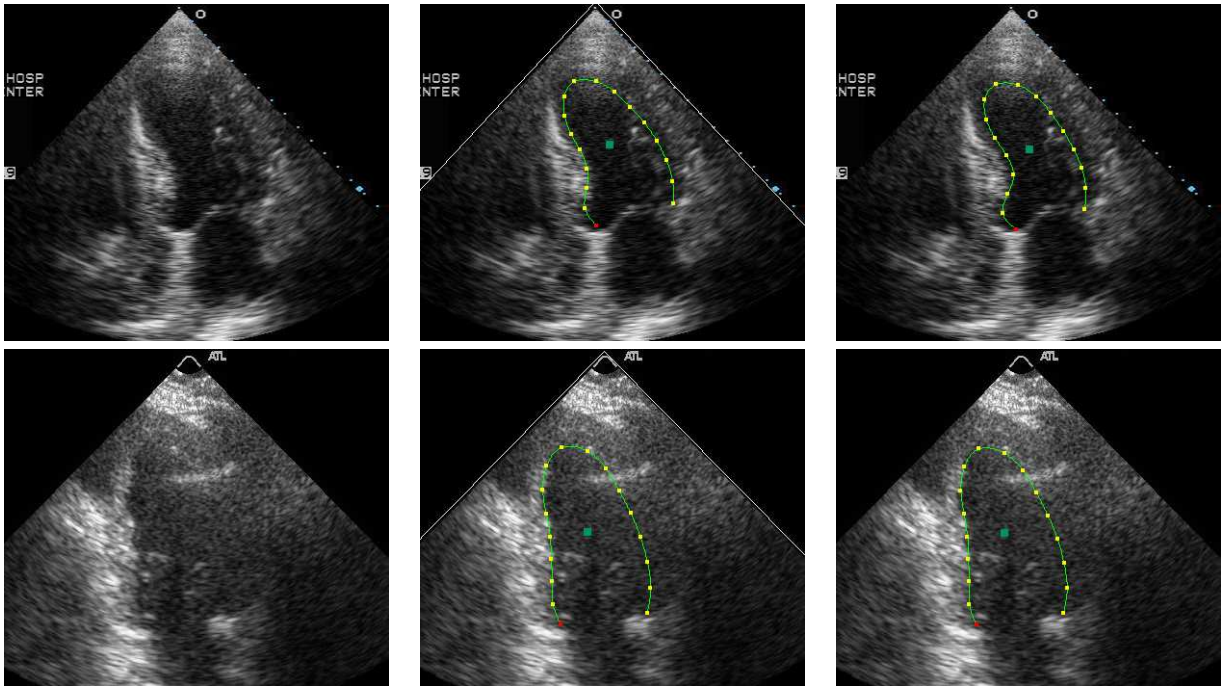
Figure 7. Median error for each control point (a) and for each case (b) for the A4C set (top) and for the A2C set (bottom) relative to the expert shape by using the between-class features and within class features.

patient. To visualize the errors, they are sorted for each curve, thus vertically they do not correspond to the same image. The top two curves in each graph represent the error between the true contour and the mean shape and the true contour and the one inferred using the normalized appearance. Thus, using the appearance is the same on average than using the mean, this is true also by using the detection features. The middle curve is the error by using the selected features and for reference the bottom curve is the nearest neighbor available in the shape space.

In the second experiment we test the error of the entire segmentation procedure. After detection and shape inference, Figure 7a shows the median error for each of the 17 control points computed using the features used for detection (top curve, between-class features) and the selected features (bottom curve, within-class features). Figure 7b illustrates the sorted global contour error for each case for all frames, where again lower error is obtained by using the within-class features than using the between-class features.

Figure 8 compares the completely automatic segmentation result (Figure 8b) to a contour drawn by an expert (Figure 8c). The difficulties of the problem are illustrated in (Figure 8a) where the input images are affected by speckle noise, there is not a clear border definition, there is signal dropout and imaging artifacts.

Additional segmentation results are shown in Figure 9 on a variety of new input images. Without occlusion handling for feature computation it is difficult to detect shapes close to the ultrasound fan (Figure 9, top right image). Note also the large variations in appearance of the interest structure.



(a) (b) (c)

Figure 8. Left ventricle endocardial border detection on an image from the A4C set (top) and A2C set (bottom). (a) input image; (b) automatic shape; (c) expert drawn contour.

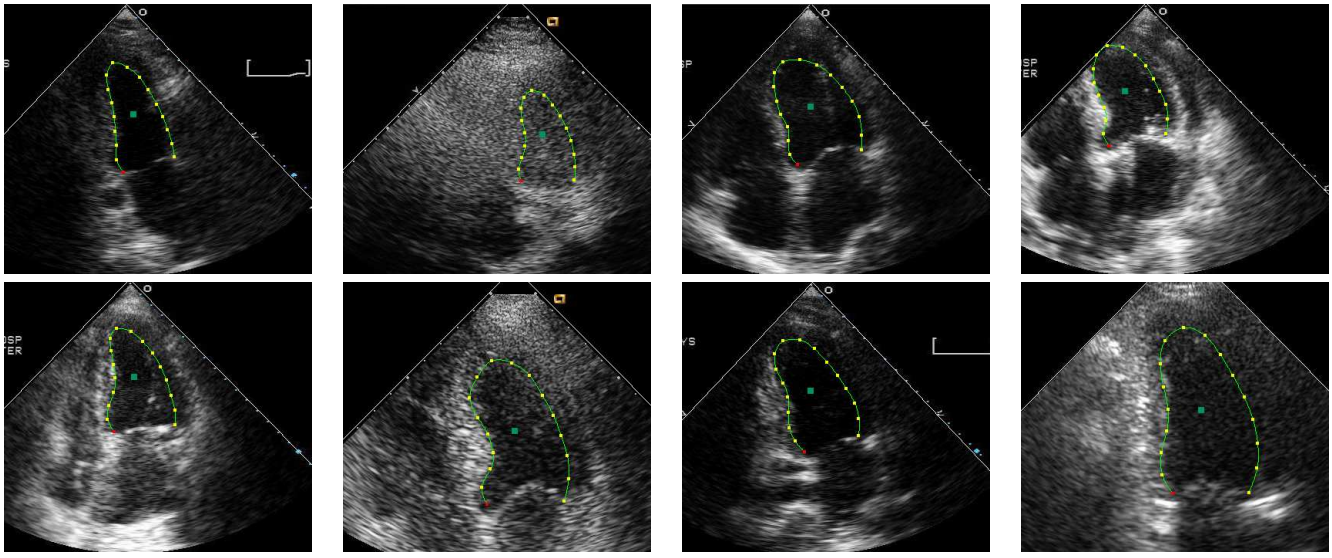


Figure 9. Left ventricle endocardial border detection.

6. Summary and Discussion

We have introduced **database-guided segmentation** as a new paradigm that directly exploits expert annotation of interest structures in large medical databases by formulating it as a two-step learning problem. The proposed method does not explicitly encode the a-priori knowledge and works under real-time constraints (under 1sec.).

We shown the performance of the method on a variety of ultrasound heart images. As illustrated in the examples, the images are corrupted with large amounts of noise, have signal drop-out in some regions and the correct segmentation regions does not correspond always to a strong edge. For example the contour should cut the right wall (papillary muscle) in images such as one in Figure 8(top row). In this cases it is difficult to formulate explicit local constraints and traditional segmentation might fail. To solve the segmentation problem for an entire video sequence (in 2D+Time), the detection and shape inference steps are integrated with our existing robust shape tracking algorithm [22] in a maximum-likelihood framework. Please see the accompanying videos for the performance of our algorithm for several sequences.

Our approach is general and can be used on a variety of segmentation problems. The requirement is that the range of nonrigid deformations in shape and the associated appearance to be possible to be captured through boosted learning using a large set of simple features. In the case of large deformations or articulated structure the task can be divided into several manageable subproblems where the appearance variations can be learned. We are currently investigating more complex learning methods for appearance-shape association such as for example using a regression setting. The difficulty is the very large dimensionality of the space in which we have to solve both the feature selection and the data fitting problems.

References

[1] M. S. Bartlett, G. Littlewort, I. Fasel, and J. R. Movellan. Real time face detection and facial expression recognition: Development and applications to human computer interaction. In *IEEE Workshop on Computer Vision and Pattern Recognition for Human Computer Interaction*, Madison, WI, June 2003.

[2] E. Borenstein and S. Ullman. Class-specific, top-down segmentation. In *Proc. European Conf. on Computer Vision*, Copenhagen, Denmark, pages 109–122, 2002.

[3] J. G. Bosch, S. C. Mitchell, P. F. Lelieveldt, F. Nijland, O. Kamp, M. Sonka, and J. H. C. Reiber. Automatic segmentation of echocardiographic sequences by active appearance motion models. *IEEE Trans. Medical Imaging*, 21(11):1374–1383, 2002.

[4] D. Comaniciu and P. Meer. Mean shift analysis and applications. In *Proc. Intl. Conf. on Computer Vision*, Kerkyra, Greece, pages 1197–1203, September 1999.

[5] T. Cootes and C. Taylor. Active shape models - Their training and application. *Computer Vision and Image Understanding*, 61(1):38–59, 1995.

[6] T. F. Cootes, G. J. Edwards, and C. J. Taylor. Active appearance models. In *1998 European Conf. on Computer Vision*, volume 2, pages 484–498, Berlin, 1998.

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

[8] Y. Freund and R. E. Schapire. Experiments with a new boosting algorithm. In *International Conference on Machine Learning*, pages 148–156, 1996.

[9] J. C. Gower. Generalized procrustes analysis. *Psychometrika*, 40(1):33–51, 1975.

[10] T. Hastie, R. Tibshirani, and J. Friedman. *The Elements of Statistical Learning*. Springer Verlag, 2001.

[11] B. Heisele, T. Serre, M. Pontil, and T. Poggio. Component-based face detection. In *CVPR01*, pages I:657–662, 2001.

[12] M. Kass, A. Witkin, and D. Terzopoulos. Snakes: Active contour models. *Intl. J. of Computer Vision*, 1:321–332, 1988.

[13] S. Mitchell, B. P. Lelieveldt, R. van der Geest, H. G. Bosch, J. H. Reiber, and M. Sonka. Time-continuous segmentation of cardiac MR image sequences using active appearance motion models. In *Proc. SPIE Medical Imaging*, San Diego, CA, USA, volume 4322, pages 249–256, 2001.

[14] D. Mumford and J. Shah. Optimal approximations by piecewise smooth functions, and associated variational problems. *Comm. Pure Math.*, pages 577–684, 1989.

[15] K. Okuma, A. Taleghani, and N. de Freitas. A boosted particle filter: Multitarget detection and tracking. In *2004 European Conf. on Computer Vision*, volume 1, pages 28–39, Prague, Czech Republic, May 2004.

[16] C. Papageorgiou and T. Poggio. A trainable system for object detection. *IJCV*, 38(1):15–33, June 2000.

[17] X. Ren and J. Malik. Learning a classification model for segmentation. In *2003 International Conf. on Computer Vision*, volume 1, pages 10–17, Nice, France, 2003.

[18] J. Shi and J. Malik. Normalized cuts and image segmentation. *IEEE Trans. Pattern Anal. Machine Intell.*, 22(8):888–905, 2000.

[19] L. H. Staib and J. S. Duncan. Model-based deformable surface finding for medical images. *IEEE Trans. Medical Imaging*, 15(5):720–731, 1996.

[20] P. Viola and M. Jones. Rapid object detection using a boosted cascade of simple features. In *Proc. IEEE Conf. on Computer Vision and Pattern Recognition*, Hawaii, 2001.

[21] J. Yang and J. S. Duncan. 3d image segmentation of deformable objects with shape appearance joint prior models. In *Proc. of Medical. Image Computing and Computer Assisted Intervention (MICCAI)*, volume 2878, pages 573–580, Montreal, Canada, November 2003.

[22] X. S. Zhou, D. Comaniciu, and A. Gupta. An information fusion framework for robust shape tracking. *IEEE Trans. Pattern Anal. Machine Intell.*, 27:115–129, 2005.

[23] X. S. Zhou, D. Comaniciu, B. Xie, R. Cruceanu, and A. Gupta. A unified framework for uncertainty propagation in automatic shape tracking. In *2004 IEEE Conf. on Computer Vision and Pattern Recog.*, volume 1, pages 872–879, Washington, DC, 2004.