

True-False Lumen Segmentation of Aortic Dissection using Multi-Scale Wavelet Analysis and Generative-Discriminative Model Matching

Noah Lee^a, Huseyin Tek^b, and Andrew F. Laine^a

^aDepartment of Biomedical Engineering, Columbia University, New York, NY, USA

^bDepartment of Imaging & Visualization, Siemens Corporate Research, Princeton, NJ, USA

ABSTRACT

Computer aided diagnosis in the medical image domain requires sophisticated probabilistic models to formulate quantitative behavior in image space. In the diagnostic process detailed knowledge of model performance with respect to accuracy, variability, and uncertainty is crucial. This challenge has led to the fusion of two successful learning schools namely generative and discriminative learning. In this paper, we propose a generative-discriminative learning approach to predict object boundaries in medical image datasets. In our approach, we perform probabilistic model matching of both modeling domains to fuse into the prediction step appearance and structural information of the object of interest while exploiting the strength of both learning paradigms. In particular, we apply our method to the task of true-false lumen segmentation of aortic dissections an acute disease that requires automated quantification for assisted medical diagnosis. We report empirical results for true-false lumen discrimination of aortic dissection segmentation showing superior behavior of the hybrid generative-discriminative approach over their non hybrid generative counterpart.

Keywords: Quantitative Image Analysis, Segmentation, Generative and Discriminative Learning, Wavelet Analysis, Model Matching, X-ray CT.

1. INTRODUCTION

Probabilistic modeling for computer aided diagnosis in the medical image domain is an active area of research [7-10]. Computer aided diagnostic processes require detailed knowledge of statistical model behavior with respect to accuracy, variability, and model uncertainty for quantitative diagnostic reasoning. This challenge has led to the fusion of two successful learning schools namely generative [11-13] and discriminative [14-16] learning. While both approaches were successfully applied to medical imaging problems in isolation [31-33], they have different strengths and weaknesses inherent in their theoretical formulation that speaks for selective ways of combining both learning paradigms. Further, it is known that each learning approach alone lacks the ability to completely address the problem domain due to the complexity of the modeling process in question [20, 21]. It is evident to find ways for combining both learning frameworks to fuse their strength for computational efficient learning in large knowledge and data domains. Such a hybrid provides superior applicability, adaptability, and performance over the individual learning frameworks in separation.

For instance, generative models enable the natural integration of prior knowledge into the probabilistic model by working on a refined probability density through Bayes rule. Given expert knowledge one can condition and marginalize over the generative joint density in a Bayesian data-driven way to perform inferences. This perspective of modeling enables us to reuse relational information and coherences of the given problem setting while being able to refine our model with new incoming knowledge. On the other hand, discriminative learning has its effectiveness in its simple formulation omitting the need to model a complete joint distribution of the problem phenomenon. In contrary, a discriminative approach directly optimizes a maximum-margin criterion for a domain specific task such as classification or regression. Through maximum-margin learning discriminative methods aim for simple task specific classification rules that optimize generalization performance on unseen data. In combining selective parts of both learning frameworks it is possible to directly address the problem domain of specific practical systems while keeping the flexibility of the generative model. In practice we often face changing environments requiring the need for a modeling approach that can adapt to the new environment in a quantitative probabilistic way. It is this symbiotic relationship that makes such a hybrid approach suitable.

In this paper we propose a generative-discriminative learning approach to predict object boundaries in medical image datasets. We apply our approach on CT medical datasets to perform true-false lumen segmentation of aortic dissection.

The segmentation problem is solved through generative-discriminative model matching exploiting object appearance and structural information cues in the learning process. For structural object information we use dyadic multi-scale Wavelet analysis to extract multi-scale edge information of the object of interest. Appearance information is modeled with a generative mixture model. Discriminative model knowledge aids in the model selection process of the generative model through weighted model matching. In comparison to deterministic or heuristic approaches our approach brings as a byproduct the following key advantages: i) easy integration of a priori knowledge through Bayesian data driven probabilistic modeling ii) optimized generalization error bounds through discriminative classification rules, and iii) model adaptability and natural refinement to changing data and knowledge environments through weighted model matching.

The remainder of the paper is organized as follows: in section 2 we outline the medical background of aortic dissection an acute disease that requires immediate treatment. Next we continue with prior work in the domain of generative-discriminative learning approaches closely related to the spirit of our work. Section 4 describes our approach using dyadic multi-scale Wavelet analysis and generative-discriminative model matching. Experimental results are shown in section 5. Finally we conclude with future work and open issues.

2. MEDICAL BACKGROUND

Life threatening aortic dissection occurs in two types (A and B) after the Stanford classification system, where a tear of the inner artery wall leads to the separation of the whole lumen into two channels. One, where blood continues to flow (true lumen) and the other, where the blood flow is zero (false lumen). Aortic dissection is a medical emergency and can quickly lead to death, even with optimal treatment. Aortic dissections resulting in rupture have an 80% mortality rate and 50% of patients die before they even reach the hospital [6]. The accurate discrimination of both lumen volumes is crucial for immediate treatment procedures. Treatment of aortic dissection requires automated quantification for assisted medical diagnosis of the complete aortic lumen structure. Manual quantification is time consuming and prone to inter- and intra observer variability. Figure 1 shows examples of cross sectional planes of the aortic dissection lumen for various pathological cases.

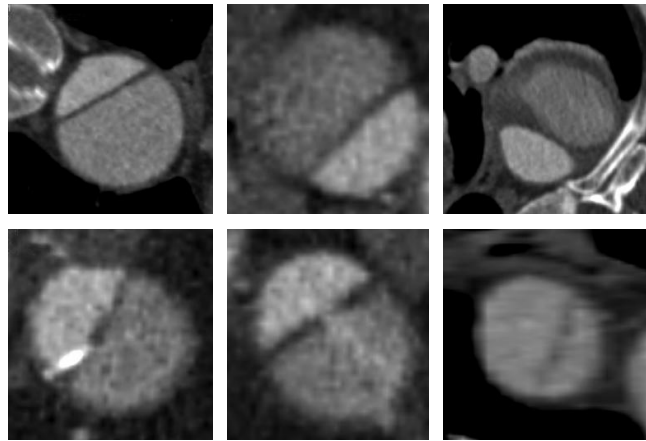


Fig. 1. Input samples of aortic dissection cross sectional planes showing various forms of the true and false lumen separated by the aortic dissection membrane. These samples are the input of our algorithm.

3. PRIOR ART

We provide two nodes of related work to situate our work. Closely related to the spirit of fusing both learning frameworks are learning approaches using i) variational maximum conditional likelihood [23, 24], ii) discriminative maximum entropy (MED) [25, 26], iii) probabilistic kernel methods [27,28], iv) multi-task probabilistic discrimination [29], and v) latent mixture and non-stationary discrimination [30]. Various approaches exist in realizing a hybrid learning framework to combine generative and discriminative modeling.

Li et al. [17] proposed a generative/discriminative learning algorithm for image classification. They presented a sequential two phase learning step using EM for generative learning to normalize the description length of the images. The output serves as the classifier input for the discriminative learning step. They outlined their method for single and multi-dimensional feature spaces. Everingham and Zisserman [18] applied their learning framework to face recognition. In their work they first use discriminative learning to classify visual features of complete head models using decision tree classifiers. The second step then verifies the decision accuracy within a generative framework employing edge features and chamfer matching. The authors reported excellent results. Zhang et al. [19] presented an integrated learning approach of both frameworks. They proposed a Random Attributed Relational Graph for generative modeling, where discriminative classifiers are used at the graph nodes to perform multi-view object detection. Other hybrid learning frameworks combining the two learning approaches can be found in [20-22]. Since we apply our framework to the task of true-false lumen segmentation in aortic dissection we briefly outline recent state of the art. A method has been reported by Kovacs et al. for automatic segmentation of the aortic dissection membrane [1]. The authors use eigenvalue analysis of the Hessian matrix to compute a probability score that a pixel belongs to the membrane.

The philosophy of our approach is a different one. Rather than targeting the extraction of the dissection membrane we perform segmentation of the true and false lumen using dyadic multi-scale Wavelet analysis coupled with generative and discriminative model matching in a machine learning framework.

4. METHODOLOGY

A - Dyadic Multi-Scale Wavelet Analysis

The Wavelet transform finds its use in many computer vision applications such as texture discrimination [35] or other frequency measures in image modeling. The Wavelet transform W can be seen as decomposing a signal $f \in L^2(\mathbb{R})$ onto a set of basis functions ψ_s , which are scaled and shifted versions of a mother wavelet. To construct a translation-invariant wavelet representation, the scale s is discretized while keeping the translation parameter u fixed. The discrete wavelet transform (DWT) can be implemented as a perfect reconstruction filter bank, where transform coefficients are obtained by successive applications of discrete filters. Dyadic wavelet transforms are scale samples of wavelet transforms sampled along a dyadic sequence of $\{s = 2^j\}_{j \in \mathbb{Z}, j = 1, \dots, n}$. The dyadic wavelet transform is defined as

$$W[f(u, s)] = \int_{-\infty}^{+\infty} f(t) \frac{1}{\sqrt{s}} \psi\left(\frac{t-u}{s}\right) dt = f * \overline{\psi}_s(u),$$

$$\overline{\psi}_s(t) = \psi_s(-t) = \frac{1}{\sqrt{s}} \psi\left(\frac{-t}{s}\right). \quad (4.1)$$

In the two dimensional case $f \in L^2(\mathbb{R}^2)$ we have

$$W_s^1 f(x, y) = f * \psi_s^1(x, y) \text{ and } W_s^2 f(x, y) = f * \psi_s^2(x, y),$$

$$\psi_s^1(x, y) = \frac{\partial \theta(x, y)}{\partial x} \text{ and } \psi_s^2(x, y) = \frac{\partial \theta(x, y)}{\partial y}. \quad (4.2)$$

Here $\theta(x, y)$ is a smoothing kernel dilated by a dyadic scale factor s and $\psi_s^{1,2}(x, y)$ denote the partial derivatives of θ . As mentioned we can compute eq. 4.2 with a fast filter bank algorithm. We use spline dyadic Wavelets satisfying the inequality constraint with frame bounds $\{A > 0, B\}$, s.t.

$$\forall (\omega_x, \omega_y) \in \mathbb{R}^2 - \{(0, 0)\}, A \leq \sum_{k=1}^K \sum_{j=-\infty}^{+\infty} |\psi^k(2^j \omega_x, 2^j \omega_y)|^2 \leq B. \quad (4.3)$$

To extract structural object information we focus on object edges at multiple scales obtained through modulus maxima representations of the discrete filter responses of $\psi_s^{1,2}(x, y)$. Translation-invariance provides us with an over-complete

frame representation suitable for edge localization. The Lipschitz regularity of edge points is derived from the decay of wavelet modulus maxima across scales. To achieve good edge detection results under noisy environments $f(x, y)$ is smoothed before its derivative is computed. Eq. 4.2 implies that the Wavelet transform coefficients are proportional to the coordinates of the gradient vector of f convoluted with a smoothing kernel. Thus the modulus of the gradient vector is proportional to the Wavelet transform modulus. The sharp variation points of the signal smoothed at different scales can be detected by finding the local maxima of the Wavelet transform magnitude. For a more detailed view on the underlying theory we refer to [34]. A point is declared as an edge point if the magnitude $|f|_2$ has a local maximum in the direction \angle of the gradient computed as

$$|f|_2 = \sqrt{|W_s^1 f(x, y)| + |W_s^2 f(x, y)|},$$

$$\angle = \arctan \frac{W_s^1 f(x, y)}{W_s^2 f(x, y)}. \quad (4.4)$$

Rather than comparing the magnitude with its neighbors closest to the gradient direction we interpolate the two values in a kernel neighborhood using bilinear interpolation. After edges are detected we group together neighboring maximum magnitude responses. We compute local maxima of a five level dyadic Wavelet decomposition for multi-scale edge detection

$$W_{2^5}^1 f(x, y) = f * \psi_{2^5}^1(x, y)$$

$$W_{2^5}^2 f(x, y) = f * \psi_{2^5}^2(x, y). \quad (4.5)$$

We discard the fine detail edges of level $\{s = 2^j\}_{1..3}$, since they do not give meaningful edge information. The level of scale identified depends on selected basis wavelets adopted for a DWT and structure size relative to the support of each wavelet. In our case multi-scale edges of level $s = 2^4$ were chosen since they show most of the coherent structural information of both lumens (see fig 2). From figure 2 one can observe false positive edges within the lumen structure. We apply a discriminative learning approach to reduce the false positive rate by injecting expert knowledge into the framework.

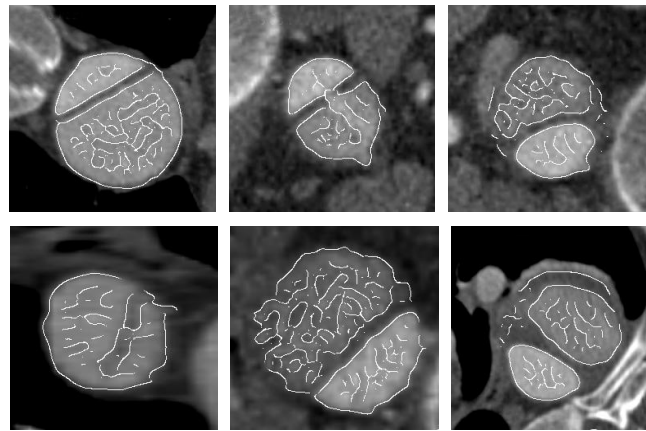


Fig. 2. Multi-scale Wavelet edge detection by local maxima detection in the Wavelet domain. Level 4 multi-scale edges overlaid on the original image data.

B – Learning the Discriminative Object Model

Starting from section A) we perform discriminative learning based edge refinement to reduce the false positive rate. Let us consider an input output mapping $\{(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)\}$, $x_i \in \mathbb{R}^m$, $y_i \in \{\pm 1\}$, where y_i denotes the binary class label (edge point label) of our observation x_i (feature vector). The feature space consists of vectorized circular patches ($1 \times m$) encoding intensity and texture covariance cues for each edge point on the true-false lumen boundary. For all edge points we build a training instance matrix A ($n \times m$) for all n edge points. For our discriminative learning task we use the Support Vector Machine (SVM) [14]. The goal is to find an optimal classification rule $f: \mathbb{R}^m \rightarrow \{\pm 1\}$ for unseen test cases, i.e. a rule that provides an optimal generalization error bound through maximum-margin classification between the two classes. Vapnik's learning theory suggests that in order to increase generalization performance the classifier should have low VC dimension and thus a large margin $M = 2/\|w\|$. For an in depth review of the underlying theory we refer to [16]. For the case of non-separable data the SVM formulation is given by

$$\begin{aligned} \min_{w, b, \xi} \quad & \frac{1}{2} \|w\|^2 + C \sum_{i=1}^n \xi_i \\ \text{Subject To} \quad & y_i(\mathbf{w}^T \mathbf{x}_i + b) \geq 1 - \xi_i, \forall 1 \leq i \leq n \\ & \xi_i \geq 0, \forall 1 \leq i \leq n. \end{aligned} \quad (4.6)$$

Here $\xi_i = 1 - y_i(\mathbf{w}^T \mathbf{x}_i + b)$ is a penalty term providing a measure for the miss-classification distance between the margin bounds and the sample pair. The parameter C is a weighting factor for the miss-classification penalty. We note that the solution obtained is always unique. Eq. 4.6 is the primal version of the SVM. In the dual case we have the data represented as a dot product. The Lagrangian of eq. 4.6 results in

$$L(\mathbf{w}, b, \xi, \alpha, \beta) = \frac{1}{2} \|\mathbf{w}\|^2 + C \sum_{i=1}^n \xi_i - \sum_{i=1}^n y_i(\mathbf{x}_i^T \mathbf{x}_i + b) - 1 + \xi_i - \sum_{i=1}^n \beta_i \xi_i \quad (4.7)$$

and leads us to the dual by maximizing L in terms of α_i, β_i

$$\begin{aligned} \max_{\alpha, \beta} \quad & -\frac{1}{2} C \sum_{i=1}^n \sum_{j=1}^n \alpha_i \alpha_j y_i y_j \mathbf{x}_i^T \mathbf{x}_j + \sum_{i=1}^n \alpha_i \\ \text{Subject To} \quad & 0 \leq \alpha_i \leq C, \forall 1 \leq i \leq n. \end{aligned} \quad (4.8)$$

Replacing the inner product with a kernel function $k(\mathbf{x}_i, \mathbf{x}_j)$ the solution of eq. 4.8 can be obtained through convex quadratic optimization. The final decision function of our kernelized discriminative classifier is given by

$$f(x) = \text{sign} \left(\sum_{i=1}^n \alpha_i y_i k(\mathbf{x}_i, \mathbf{x}) + b \right). \quad (4.9)$$

In our implementation we use a Gaussian RBF kernel $k(\mathbf{x}_i, \mathbf{x}) = \exp(-\gamma|\mathbf{x}_i - \mathbf{x}|^2)$ with $\gamma = 8e - 4$ and a penalty weighting factor of $C = 10$. Both parameters were obtained through cross-validation of our training set. Once our discriminative classifier is trained we refine our results obtained from A) to reduce the false positive rate. Most prominent lumen edge points are then predicted by computing the maximum posterior probability from eq. 4.9 as described in [36]. The next step is to model the appearance information of the object of interest.

C – Learning the Generative Appearance Model

Given initial vessel cross section boundaries we estimate the intensity distribution inside the lumen by decomposing it into a set of K mixture components. The density of each data point in a mixture model can be written as

$$p(\mathbf{y}|\theta) = \sum_{k=1}^K \pi_k p(\mathbf{y}|\theta_k), \quad (4.10)$$

where each of the K components of the mixture is, for instance, a Gaussian with differing means and covariances $\theta_k = (\mu_k, \Sigma_k)$ and π_k is the mixing proportion for component k , such that $\sum_{k=1}^K \pi_k = 1$ and $\pi_k > 0, \forall k$. The goal is to maximize the log-likelihood of our mixture model given unlabeled data. Therefore we introduce a latent variable for the unknown class distribution. Expressed as a latent variable model we can write eq. 4.10 as

$$p(\mathbf{y}|\theta) = \sum_{k=1}^K P(s = k|\pi) p(\mathbf{y}|s = k, \theta), \quad (4.11)$$

where $P(s = k|\pi) = \pi_k$ is the prior for the latent variable taking on value k , and $p(\mathbf{y}|s = k, \theta) = p(\mathbf{y}|\theta_k)$ is the density under component k . The hidden variable is a multinomial binary K element vector with $\sum_{i=1}^K s^i = 1$. To estimate our mixture model we use the EM algorithm. Dependent on the disease state and progress the intensity distributions of the true and false lumen vary such that a fixed mixture model cannot be assumed a priori. However, the true and false lumen intensity lies within a range of mixture components. To perform model selection we perform a model matching between the discriminative and the generative model to fuse structural and appearance information into the learning framework.

D –Generative and Discriminative Model Matching

We perform model matching at a high-level representation to keep model independence in our learning framework. To perform matching we optimize the generative conditional density by taking the derivative of the distribution and maximize for high similarity of the discriminative probabilistic output mapping. Weighting parameters between both frameworks ensure adaptability for task specific modeling assumptions. We note that in that way both learning frameworks interact in a careful systematic way to share and communicate model performance for the final decision task. Below we give a brief summary of our matching algorithm.

Algorithm: Generative-Discriminative Model Matching

1| Train discriminative object model of true-false lumen boundary using expert knowledge $\{\mathbf{x}_i, y_i\}_{i=1}^n$

- Solve dual SVM using QP:

$$\max_{\alpha, \beta} -\frac{1}{2} C \sum_{i=1}^n \sum_{j=1}^n \alpha_i \alpha_j y_i y_j \mathbf{x}_i^T \mathbf{x}_j + \sum_{i=1}^n \alpha_i$$

Subject To $0 < \alpha_i < C, \forall 1 < i < n.$

2 | For each cross sectional plane f compute eq. 4.5 to obtain edge feature map $\{\mathbf{x}_i, y_i\}_{i=1}^n$ of the object model:

3 | Use discriminative model from step 1) and classify prominent edge point after:

$$f(x) = \text{sign} \left(\sum_{i=1}^n \alpha_i y_i k(\mathbf{x}_i, \mathbf{x}) + b \right).$$

- Label the prominent edge point set as $\{\bar{\mathbf{x}}_i, \bar{y}_i\}_{i=1}^n$

4 | Compute the generative mixture model $p(\mathbf{y}|\theta) = \sum_{k=1}^K P(s = k|\pi)p(\mathbf{y}|s = k, \theta)$ on f (step 3) and 4) interchangeable) :

- Maximize lower bound of log-likelihood using:

$$\begin{aligned} L(\theta) &= \log p(\mathbf{y}|\theta) = \log \int p(\mathbf{x}, \mathbf{y}|\theta) d\mathbf{x} \\ &= \log \int q(\mathbf{x}) \frac{p(\mathbf{x}, \mathbf{y}|\theta)}{q(\mathbf{x})} d\mathbf{x} \\ &\geq \int q(\mathbf{x}) \log \frac{p(\mathbf{x}, \mathbf{y}|\theta)}{q(\mathbf{x})} d\mathbf{x} \end{aligned} \quad (4.12)$$

- Estimate ML of θ using iterative optimization:

E-Step: Compute updated cond. expectation $q_k(\mathbf{x})$ at each step:

$$q_k(\mathbf{x}) = \operatorname{argmax}_{q(\mathbf{x})} \int q(\mathbf{x}) \log \frac{p(\mathbf{x}, \mathbf{y}|\theta_{k-1})}{q(\mathbf{x})} d\mathbf{x} \quad (4.13)$$

M-Step: Compute ML parameters on new $q_k(\mathbf{x})$:

$$\theta_k = \operatorname{argmax}_{\theta} \int q_k(\mathbf{x}) \log \frac{p(\mathbf{x}, \mathbf{y}|\theta)}{q_k(\mathbf{x})} d\mathbf{x} \quad (4.14)$$

5 | For each k : mixture component compute conditional density and perform model selection:

$$l(\theta_k)_j = \max_k \frac{\exp\left(-\frac{1}{2}(\mathbf{y}_j - \mu_k)(\Sigma_k)^{-1}(\mathbf{y}_j - \mu_k)^T\right)}{|\Sigma_k|^{-\frac{1}{2}}} \quad (4.15)$$

$$\theta_k = \operatorname{argmax}_k \sum_{\bar{y}=1}^K \left(\lambda_1 \sum_{\bar{y}=1} \{\bar{\mathbf{x}}_i, \bar{y}_i\} + \lambda_2 \sum_{\partial l(\theta)_j > 0} l(\theta)_j \right) \quad (4.16)$$

In our experiments we build a decomposition of at least $K = 4$ mixture components and apply a matching to the discriminative shape model to detect the appropriate K mixture model. Model training in step 1 was performed for $n = 10$ datasets randomly selected from our database, where each dataset consists of approximately 200-300 individual cross sections. Matching is performed between the conditional density of the mixture model and the discriminative object model in a least square cost functional. Our matching approach has the flexibility that generative and discriminative learning can be performed independently. Additionally, weighting parameters control the influence of each model. In that way our approach can adapt to different application scenarios and changing data domains. If we have high confidence that the generative model is a good representation for the problem phenomenon we favor the generative prediction over the discriminative model.

5. EXPERIMENTS AND RESULTS

Our approach integrates into an existing vessel segmentation platform [4]. In this platform the complete aorta lumen was segmented using a tracking approach, where successive cross sectional segmentations build up the aortic mesh structure of the lumen boundary. Cross sections were segmented from previously extracted centerline points. To find the whole lumen structure of the aorta the user defined start and end seeds of the aorta. Our algorithm works on individual cross sections as shown in the left part of below figure. Each cross sectional plane had a resolution of 256x256. All cross sectional planes were normalized to the intensity range [0-1]. We sampled individual cross section planes along the aortic medial axis in 5mm distances.



Fig. 3. Aorta segmentation of aortic dissection. The left part shows the segmented cross section boundaries in blue. One cross section boundary is shown on the left. The right view shows the aorta boundary mesh (red) as an iso-surface.

To evaluate our approach we have tested our algorithm on various cases of aortic dissection covering a variety of disease phenotypes to address variations in segmentation complexity. We performed validation compared to expert grading for the hybrid generative-discriminative approach and a non-hybrid generative approach. The hybrid approach outperformed the generative approach with a mean sensitivity of 0.83 and a mean specificity of 0.87 with standard deviations of ± 0.098 and ± 0.15 respectively. The generative approach alone performed with a mean sensitivity of 0.7 and a mean specificity of 0.85 with standard deviations of ± 0.148 and ± 0.253 . Figure 4 shows the ROC statistics using our learning framework for 14 randomly selected datasets. Figure 5 shows examples of detected true-false lumen boundaries overlaid on individual cross sectional planes.

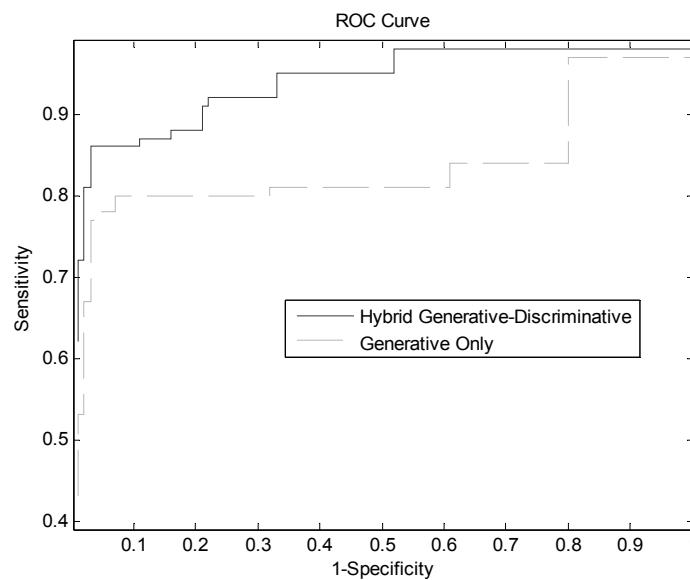


Fig. 4. ROC curve comparing segmentation performance between the hybrid generative-discriminative approach and generative only approach.

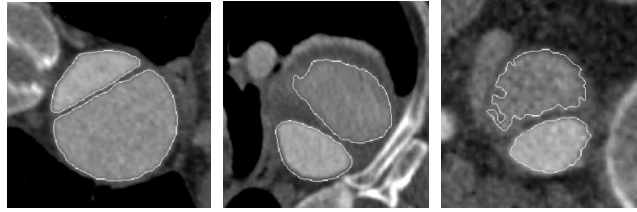


Fig. 5. Examples of true-false lumen segmentation boundaries overlaid in yellow on cross section image data using our generative-discriminative learning approach.

6. CONCLUSION

We have presented a hybrid learning approach combining generative and discriminative learning through multi-scale Wavelet analysis and probabilistic model matching. Obtained results seem promising. To our knowledge no such system was reported for the application of true-false lumen segmentation in aortic dissection datasets. The segmentation boundaries can be used for visualization of the two lumen channels and estimation of image based physiologic function such as cardiac blood flow. These measures are clinically relevant to track disease progress as well as quantitative assessment for acute disease diagnosis. Future work is devoted to an in depth evaluation of our learning approach to other object targets as well as research towards a generalization of our learning algorithm to other application domains. Furthermore clinical validation studies will follow.

ACKNOWLEDGEMENTS

This work was supported by the Imaging and Visualization Department at Siemens Corporate Research.

REFERENCES

- [1] T. Kovacs et al., "Automatic Segmentation of the Aortic Dissection Membrane from 3D CTA Images," In *Medical Imaging and Augmented Reality*, vol. 4091, pp. 317-324, (2006).
- [2] I. Koren, A. F. Laine, et al., "Edge Detection in Echocardiographic Image Sequences by 3-D Multi-Scale Analysis," In *Proceedings IEEE International Conference on Image Processing*, vol. 1, pp. 288-292, (1994).
- [3] A. Laine and X. Zong, "Border Identification of Echocardiograms via Multi-Scale Edge Detection and Shape Modeling," In *Proceedings of the IEEE International Conference on Image Processing, Lausanne, Switzerland*, vol. 3, pp. 287-290, (1996).
- [4] H. Tek et al., "Multi-Scale Vessel Boundary Detection," In *Computer Vision for Biomedical Image Applications (CVBIA)*, pp. 388-398, (2005).
- [5] T. Jebara, "Discriminative, Generative, and Imitative Learning," PhD Thesis, Media Laboratory MIT, (2001).
- [6] Unknown, "Aortic Dissection," http://en.wikipedia.org/wiki/Aortic_dissection, Wikipedia, (2007).
- [7] M. Andreetto, L. Zelnik-Manor, and P. Perona, "Non-Parametric Probabilistic Image Segmentation," In *IEEE 11th International Conference on Computer Vision (ICCV)*, (2007).
- [8] K. L. Vincken, A. S.E. Koster, M. A. Viergever, "Probabilistic Multiscale Image Segmentation," In *IEEE Transaction on Pattern Analysis and Machine Intelligence (PAMI)*, vol. 19, no. 2, pp.109-120, (1997).
- [9] A. Marco, Y.-M. Lih, P. Pietro, "Non-Parametric Probabilistic Image Segmentation," In *IEEE 11th International Conference on Computer Vision (ICCV)*, pp.1-8, (2007).

- [10] O. Rotem, H. Greenspan, J. Goldberger, "Combining Region and Edge Cues for Image Segmentation in a Probabilistic Gaussian Mixture Framework," In Conference on Computer Vision and Pattern Recognition (CVPR), (2007).
- [11] D. Barber and C. Williams, "Gaussian Processes for Bayesian Classification via hybrid Monte Carlo," In Advances in Neural Information Processing Systems, vol. 9, (1997).
- [12] Z. Ghahramani and M. Beal, "Variational Inference for Bayesian Mixture of Factor Analyzers," In Advances in Neural Information Processing Systems, vol. 12, (1999).
- [13] M.I. Jordan and R. A. Jacobs, "Hierarchical Mixtures of Experts and the EM Algorithm," In Neural Computation, vol. 6, pp. 181-214, (1994).
- [14] C. Cortes and V. Vapnik, "Support-Vector Networks," In Journal of Machine Learning, vol. 20, pp. 273-297, (1995).
- [15] Y. Freund and R. Schapire, "Experiments with a New Boosting Algorithm," In Proceedings of the 13th International Conference on Machine Learning, (1996).
- [16] N. Cristianini and J. S. Taylor, "An Introduction to Support Vector Machines and other Kernel-Based Learning Methods," Cambridge University Press, (2000).
- [17] Y. Li, L. G. Shapiro, J.A. Bilmes, "A Generative/Discriminative Learning Algorithm for Image Classification," In Proceedings of IEEE International Conference on Computer Vision (ICCV), vol. 2, pp. 1605-1612, (2005).
- [18] M. Everingham and A. Zisserman, "Identifying Individuals in Video by Combining Generative and Discriminative Head Models," In Proceedings of the 10th IEEE International Conference on Computer Vision (ICCV), vol. 2, pp. 1103-1110, (2005).
- [19] D.-G. Zhang and S.-F. Chang, "A Generative-Discriminative Hybrid Method for Multi-View Object Detection," In Proceedings of the 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR), vol. 2, pp. 2017-2024, (2006).
- [20] I. Ulusoy, C. M. Bishop, "Generative Versus Discriminative Methods for Object Recognition," Computer Vision and Pattern Recognition (CVPR), vol. 2, pp. 258-265, (2005).
- [21] P. M. Long, R. A. Servedio, H. U. Simon, "Discriminative Learning can Succeed where Generative Learning Fails," Information Processing Letters, vol. 103, no. 4, pp. 131-135, (2007).
- [22] B. M. Kelm, C. Pal, and A. McCallum, "Combining Generative and Discriminative Methods for Pixel Classification with Multi-Conditional Learning," Proceedings of the 18th International Conference on Pattern Recognition (ICPR), vol. 2, pp. 828-832, (2006).
- [23] T. Jebara and A. Pentland, "Maximum Conditional Likelihood via Bound Maximization and the CEM Algorithm," In Neural Information Processing Systems 11, NIPS, (1998).
- [24] T. Jebara and A. Pentland, "On Reversing Jensen's Inequality," In Neural Information Processing Systems 13, NIPS 13, (2000).
- [25] T. Jaakkola, M. Meila, and T. Jebara, "Maximum Entropy Discrimination," In Neural Information Processing Systems 12, NIPS 12, (1999).
- [26] T. Jebara, T. Jaakkola, "Feature Selection and Dualities in Maximum Entropy Discrimination," In 16th Conference on Uncertainty in Artificial Intelligence (UAI), (2000).
- [27] T. Jebara, R. Kondor, A. Howard, "Probability Product Kernels," Journal of Machine Learning Research, vol. 5, pp. 819-844, (2004).
- [28] R. Kondor and T. Jebara, "Gaussian and Wishart Hyperkernels," In Neural Information Processing Systems (NIPS), (2006).
- [29] T. Jebara, "Multi-Task Feature and Kernel Selection for SVMs," In International Conference on Machine Learning (ICML), (2004).
- [30] D. P. Lewis, T. Jebara, and W. S. Noble, "Nonstationary Kernel Combination," In International Conference on Machine Learning (ICML), (2006).

- [31] H. Greenspan, A. Ruf, and J. Goldberger, "Constrained Gaussian Mixture Model Framework for Automatic Segmentation of MR Brain Images," In IEEE Transactions on Medical Imaging (TMI), vol. 25, no. 9, pp. 1233-1245, (2006).
- [32] C. Carson, S. Belongie, H. Greenspan, J. Malik, "Blobworld: Image Segmentation using Expectation-Maximization and Its Application to Image Querying," In IEEE Transactions of Pattern Analysis and Machine Intelligence (PAMI), vol. 24, no. 8, pp. 1026-1038, (2002).
- [33] I. El-Naqa, Y. Yang, M. N. Wernick, "A Support Vector Machine Approach for Detection of Microcalcifications," In IEEE Transactions on Medical Imaging (TMI), vol. 21, no. 12, pp. 1552-1563, (2002).
- [34] S. Mallat and S. Zhong, "Characterization of Signals from Multiscale Edges," In IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI), vol. 14, no. 7, pp. 710-732, (1992).
- [35] A. F. Laine and J. Fan, "Frame Representations for Texture Segmentation," In IEEE Transactions on Image Processing (TIP), vol. 5, no. 5, pp. 771-780, (1996).
- [36] J. Platt, "Probabilistic Outputs for Support Vector Machines and Comparisons to Regularized Likelihood Methods," In Advances in Large Margin Classifiers, MIT Press, (1999).