A probabilistic, hierarchical, and discriminant framework for rapid and accurate detection of deformable anatomic structure

S. Kevin Zhou, F. Guo, J.H. Park, G. Carneiro, and D. Comaniciu
Siemens Corporate Research, Integrated Data Systems
755 College Road East, Princeton NJ 08540

Abstract

We propose a probabilistic, hierarchical, and discriminant *(PHD) framework for fast and accurate detection of deformable anatomic structures from medical images. The PHD framework has three characteristics. First, it integrates distinctive primitives of the anatomic structures at global, segmental, and landmark levels in a probabilistic manner. Second, since the configuration of the anatomic structures lies in a high-dimensional parameter space, it seeks the best configuration via a hierarchical evaluation of the detection probability that quickly prunes the search space. Finally, to separate the primitive from the background, it adopts a discriminative boosting learning implementation. We apply the PHD framework for accurately detecting various deformable anatomic structures from M-mode and Doppler echocardiograms in about a second.*

1. Introduction

Rapid and accurate detection of deformable anatomic structures from medical images is a difficult task. The main reason is that these anatomic structures are deformable, rendering a high-dimensional configuration space to explore. Second, the anatomy appearance variation is large, resulting in a complex model. Finally, typical speed and accuracy requirements for this type of system pose additional challenges.

To illustrate the problem, consider the deformable anatomic structures in M-mode and Doppler echocardiograms [5] shown in Figure 1. The M-mode echocardiogram is a spatial-temporal image slice of the human heart captured by an ultrasound device. Unlike regular B-mode echocardiography that uses multiple interrogation beams, the M-mode echocardiography uses a single interrogation beam and hence achieves an enhanced temporal and spatial (along the single line though) resolution. It is often used in clinical practices to assess the functionality of anatomic structures inside the heart such as left ventricle and aortic root as its high image quality allows accurate measurement and captures subtle motion. The Doppler echocardiograhy, which is widely used to assess cardiovascular functionalities such as valvular regurgitation and stenosis, employs the Doppler effect to determine whether structures (usually blood) are moving towards or away from the ultrasound probe, and its relative velocity. The acquired Doppler echocardiogram is a velocity-time image.

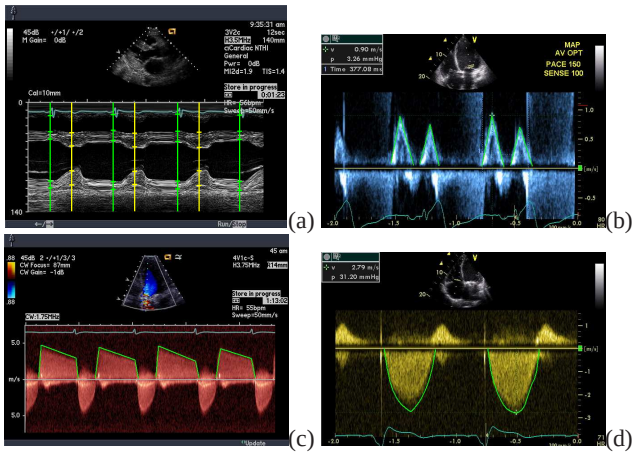

Figure 1. *(a) M-mode echocardiogram and (b,c,d) Doppler echocardiogram: (b) mitral inflow, (c) aortic regurgitation, (d) tricuspid regurgitation.*

As shown in the Figure 1, from the M-mode echocardiogram, we detect (a) a cohort of five/four landmarks on the lines corresponding to the end of diastole (ED)/the end of systole (ES); from the Doppler echocardiogram, we detect (b) triangle(s) from a mitral inflow image, (c) quadrilateral(s) from an aortic regurgitation image, and (d) curve(s) from a tricuspid regurgitation image. Note that the ED/ES line position in a M-mode image and the baseline position y_0 in a Doppler image are given *a priori*. Throughout the paper, we use the following parameterizations:

$$\theta_{a,ED} = (y_1, y_2, y_3, y_4, y_5), \theta_{a,ES} = (y_1, y_2, y_3, y_4), \quad (1)$$

$$\theta_b = (x_{LR}, y_0, x_{PK}, y_{PK}, x_{RR}, y_0), \quad (2)$$

$$\theta_c = (x_{LR}, y_0, x_{LP}, y_{LP}, x_{RP}, y_{RP}, x_{RR}, y_0), \quad (3)$$

$$\theta_d = (x_{LR}, y_0, x_{PK}, y_{PK}, x_{RR}, y_0, \alpha_1, ..., \alpha_n). \quad (4)$$

In (1), y_i is the y-coordinate of the i^{th} landmark position. In (2), we parameterize three points of left root (LR), right root (RR), and peak (PK) using four variables as the baseline is fixed. In (3), we parameterize four points of left root (LR), right root (RR), left peak (LP) and right peak (RP) with six variables. In (4), we first align three points (LR, RR, and PK) and then use n PCA coefficients to model the curve variation. Typically we choose $n = 3$, leading to a 7-D parameterization.

The use of generative models/energy minimization methods to detect deformable structures is widely studied in the literature [3, 6, 11, 12, 16]. Classical deformable models [11, 12] seek a parameterized curve that minimizes the cost function based on the gradient operator, assuming that the edge defines the curve. Felzenzwalb [6] represented a deformable shape using triangulated polygons and fitted the shape via energy minimization. Sclaroff and Liu [16] used model-based region grouping to find a deformable template while Coughlan and Ferreira [3] used loopy belief propagation. The main disadvantage of using the above generative models lies in their need for initialization and slow fitting speed.

In this paper, we pursue a discriminative approach. Motivated by the success of classifier-based detector for rigid object [19, 20], which is able to handle the large appearance variation manifested by the object, we formulate the deformable structure detection problem as a classification task. Given an image $\mathtt{I}$, we aim to discover the best configuration $\hat{\theta}$ (or several isolated ones) that maximizes (or locally maximizes) the detection probability $p(\mathtt{O}|\mathtt{I}, \theta) = p(\mathtt{O}|\mathtt{I}(\theta))$, i.e.,

$$\hat{\theta} = \arg_\theta \max p(\mathtt{O}|\mathtt{I}, \theta), \quad (5)$$

where $\mathtt{I}(\theta)$ is a warped patch extracted from the image $\mathtt{I}$ using the parameter θ. Due to nonrigid deformation, the warping computation becomes a bottleneck. In the M-mode experiment, if we use a global detector trained based on the non-rigidly warped images, during testing typically for an ED line there are over 10^{10} warping possibilities. Performing all these warping operations alone takes more than two months on a standard PC!

In section 2, we propose a *probabilistic, hierarchical, and discriminant* (PHD) framework for detecting anatomic structures from medical images. The PHD framework probabilistically integrates distinctive primitives manifested by the anatomic structure at global, segmental, and landmark levels to give an accurate account of the object. Because the configuration of the anatomic structures lies in a high-dimensional parameter space, the PHD framework seeks the best configuration via a hierarchic evaluation of the detection probability that quickly prunes the search space. Inspired by that argument that "visual processing in cortex is classically modeled as a hierarchy of increasingly sophisticated representations" [15], we build up the hierarchy in a simple-to-complex fashion. To separate the primitives from the background, the PHD framework implements the discriminative boosting learning. In section 3, we applied the proposed framework to detecting various deformable anatomic structures such as a cohort of landmarks, triangles, quadrilaterals, and curves from M-mode and Doppler echocardiograms in about a second per structure. Section 4 concludes the paper.

2. The PHD framework

2.1. Probabilistic

Let $\mathtt{P}$ denote the appearance for a primitive derived from the image. The primitive can be a landmark $\mathtt{L}$, a local segment/"part" $\mathtt{R}$, a perfectly warped global template $\mathtt{T}$. Here the term segment/"part" loosely means some intermediate representation between the landmark and global template; in other words, the segment/"part" uses a partial parameterization of the overall parameter θ. Each primitive $\mathtt{P}$ is parameterized by $\theta^P \subseteq \theta$. Given an image $\mathtt{I}$ and its associated primitives $\{\mathtt{P}_i;\ i = 1, \ldots, N_P\}$, the PHD framework, assuming the conditional independence among the primitives, aims to discover the best configuration $\hat{\theta}$ that maximizes the detection probability $p(\mathtt{O}|\mathtt{I}, \theta)$ defined as the product of primitive detection probabilities:

$$p(\mathtt{O}|\mathtt{I}, \theta) = \prod_{i=1}^{N_P} p(\mathtt{O}|\mathtt{P}_i, \theta^{P,i}), \quad (6)$$

where N_P is the number of primitives. Equivalently,

$$p(\mathtt{O}|\mathtt{I}, \theta) = \prod_{i=1}^{N_L} p(\mathtt{O}|\mathtt{L}_i, \theta^{L,i}) \prod_{j=1}^{N_R} p(\mathtt{O}|\mathtt{R}_j, \theta^{R,j}) p(\mathtt{O}|\mathtt{T}, \theta). \quad (7)$$

where N_L and N_R are the numbers of landmarks and segments, respectively, and $N_P = N_L + N_R + 1$. Note that there is only one perfectly aligned global template.

Part-based object representation [1, 4, 7] has recently gained prevalence. The main idea is to put together multiple local parts into a spatial arrangement using a generative model. Combining the generative and discriminative models in a part-based representation for object detection is also proposed in [10, 21]. Garg *et al.* [9] fused a global ICA representation with part-based SNoW detectors for car detection. Mohan *et al.* [13] first detected the four components of the human body: the head, legs, left arm, and right arm and then further classified these components annexed in the proper geometric configuration with a second classifier as

either a pedestrian or not. The classifiers trained in [13] are based on SVM. However, the above approaches hardly meet the speed requirement posed by medical applications. Also, they tend to measure the detection performance by detection and false alarm rates; but we need to derive accurate clinically meaningful measurements.

For the M-mode case, we learned five landmark detectors, one for each landmark $\mathsf{L}_i(y_i)$ and one global detector for warped template $\mathsf{T}(y_1, y_2, ..., y_5)$ (also called warping detector). For the Doppler case, we learned 2-3 landmark detectors (two root detectors and/or one peak detector), one "part" detector, and one global detector for warped image $\mathsf{T}(\theta)$. The "part" detector used is a box detector that finds the bounding box containing the Doppler structure. For example, in the Doppler aortic regurgitation case, the parameter θ^R associated with the box is $\theta^R = (x_{LR}, y_0, y_{LP}, x_{RR}, y_0)$.

2.2. Hierarchical

Using the product rule in (8) allows an efficient exploration of the parameter space: If any term in the product is zero (or close to zero), then the overall detection probability is zero (or close to zero). This implies the following strategy for computational efficiency: if one of the classifiers fails to recognize the input as positive, we can simply stop evaluating the remaining classifiers.

$$\arg\max_{\theta} p(\mathsf{O}|\mathsf{I}, \theta) = \prod_{i=1}^{N_P} p(\mathsf{O}|\mathsf{P}_i, \theta^{P,i}) \quad subject\ to\ p(\mathsf{O}|\mathsf{P}_i, \theta^{P,i}) > \epsilon_i, \tag{8}$$

where each ϵ_i is a pre-specified threshold close to zero. Each classifier defines a "feasible" region in which lies the parameter. The overall "feasible" region is the intersection of the "feasible" regions of all classifiers. We seek the maximizing configuration in the overall "feasible" region. We implement the above space pruning idea using a progressive detector hierarchy illustrated in Figure 2. The progressive detector hierarchy consists of multiple layers of detectors. Each layer targets detecting a particular primitive or pruning the relevant space to find the "feasible" region.

The proposed detector hierarchy seems similar to the detector cascade [20]. However, there exists a significant difference between them: the hierarchy divides the parameter space for fast exploration of the full space while the cascade always explore the full space. As argued earlier, blindly applying the cascade for detecting deformable structure is computationally prohibitive!

Following [15], we adopt the principle of using the simple models first followed by complex models when designing the progressive detector hierarchy . There are two types of complexity: one is model complexity and the other is computational complexity. The model complexity of a binary classifier is determined by the shape of decision boundary. The computational complexity depends on both the model complexity and scanning procedure. For example, the lef/right root detector is simple to learn and needs only a line scan; on the other hand the warping detector is difficult to learn, rendering a complex model, and it takes longer to search. To build a detector hierarchy that supports fast evaluation, we start with simple models and progressively move to complex ones in terms of computation. Table 1 lists the primitive detectors (along with the number of weak classifiers) used in the progressive detector hierarchies for detecting anatomic structures in the experiments.

2.3. Discriminant

We followed [19] to learn a probabilistic boosting tree (PBT) as a binary object detector. The PBT trains a binary decision tree, with each node of the tree being a strong classifier that combines multiple weak classifiers via a *discriminant* boosting procedure [8]. Because we based the weak classifier on the Haar-like local rectangle features [14, 20], whose rapid evaluation is enabled via the means of integral image, the PBT operates as a feature selector. The PBT also has early exits for fast negative rejection. We also need to compute the posterior probability. A nice property of the PBT is that it allows exactly computing the posterior probability of being positive. Refer to [19] for more details on PBT.

To train the detectors in all layers of the hierarchy, we need positives and negatives. Generating positives is straightforward by using the ground truth annotation (with a slight perturbation). When generating negatives, we take into account the interaction between layers especially when training the detector in the later layers of the hierarchy. For example, when generating negatives for the 2nd layer box detector for quadrilateral detection in the aortic regurgitation image, we used only the positives values of x_{LR} and x_{RR} that pass the 1st layer root detectors; for the y_{LP} variable, we used those a few pixels away from the ground truth position.

2.4. Mode selection

The candidates close to the ground truth position (or highly confusing spots) are likely to fire up due to smoothness, which renders a large number of candidates. Optionally, we may run a mode selection to further reduce the search space by finding isolated local maxima. Below, we illustrate the mode selection scheme using the 1-D example. Given a probability response line, we first smoothed it to find all local maxima. After ranking the local maxima based on their responses, we then performed the following operations to find isolated modes. Let the set of

	M-mode	Mitral inflow	Aortic reg.	Tricuspid reg.
structure	a cohort of landmarks	triangle	quadrilateral	curve
1st layer det.	indep. landmarks	box	left root & right root	left root & right root
parameter	(y_i)	(x_{LR}, y_{PK}, x_{RR})	(x_{LR}) & (x_{RR})	(x_{LR}) & (x_{RR})
# of WCs	$\sim$400	299	61 & 92	243&274
2nd layer det.	warping	peak	box	box
parameter	$(y_1, \ldots, y_5)$	(x_{PK}, y_{PK})	(x_{LR}, y_{LP}, x_{RR})	(x_{LR}, y_{PK}, x_{RR})
# of WCs	$\sim$1000	103	192	739
3rd layer det.	NA	NA	left peak	warping
parameter	-	-	(x_{LP}, y_{LP})	$(x_{LR}, x_{PK}, y_{PK}, x_{RR}, \alpha_1, \alpha_2, \alpha_3)$
# of WCs	-	-	54	550
4th layer det.	NA	NA	warping	NA
parameter	-	-	$(x_{LR}, x_{LP}, y_{LP}, x_{RP}, y_{RP}, x_{RR})$	-
# of WCs	-	-	316	-

Table 1. *The list of primitive detectors in the progressive detector hierarchy.*

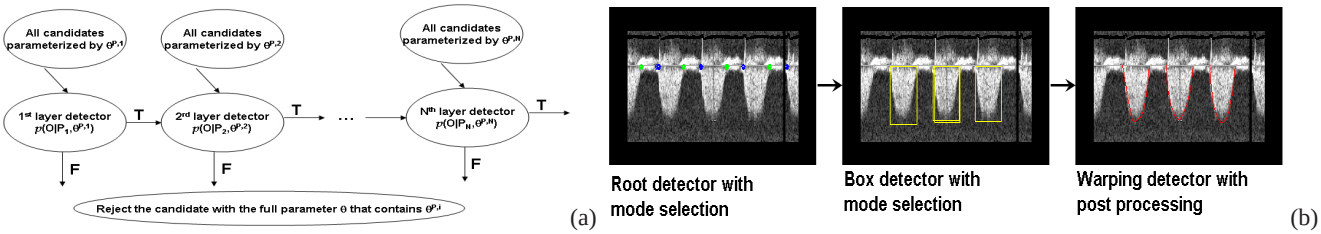

Figure 2. *(a) Graphical illustration of progressive detector hierarchy. (b) A real example of applying the PHD framework for detecting curves in the tricuspid regurgitation image.*

local maxima be $\{y_1, y_2, ..., y_M\}$, $\mathcal{L}$ the final list of selected models initialized as $\mathcal{L} = \emptyset$, and λ a pre-specified threshold. For $m = 1, 2, \ldots, M$, if the minimum distance $min_{x \in \mathcal{L}}(y_n, x) \geq \lambda$, then add it to $\mathcal{L}$: $\mathcal{L} = \mathcal{L} \cup \{y_m\}$; otherwise continue. Figure 3(a) displays a typical probability response map for a landmark and its isolated modes (marked as red circles). For a high-dimensional case, a similar strategy can be designed.

2.5. Image warping

One essential part is to align the global image appearance to place the landmarks in the canonical positions. For 2-D image warping, we used the piecewise local rectangle warping; other methods like thin plate spline (TPS) warping [2] can be used too. Figure 3(f,g) displays several example images after 2D warping of the aortic and tricuspid regurgitation images.

In the M-mode case, we need only the 1-D warping as the scale is fixed along the x-axis. Figure 3(b,c) illustrates the warping process. In Figure 3(b), two synthetic signals with the peaks located at different positions are displayed and in Figure 3(c), the peaks of the two signals are roughly aligned after warping.

We study the 1-D counterpart of TPS warping. Assuming that in the query image, the landmarks are located at $\{y_1, y_2, \ldots, y_N\}$ while in the canonical template, the N landmarks should be positioned at $\{z_1, z_2, \ldots, z_N\}$. We seek a warping function (or interpolator) $y = f(z)$ that satisfies $y_n = f(z_n)$ by assuming

$$f(z) = \sum_{n=1:N} c_n \phi(|z - z_n|) + z \cdot d, \tag{9}$$

where $\phi(r) = r^2 \log(r)$ is the TPS function and $\{c_1, c_2, \ldots, c_N, d\}$ are coefficients. Figure 3(d,e) displays several pairs of images before and after warping. Note the significant variations in the landmark positions and the image intensities.

To determine the coefficients, we express the conditions $\{y_n = f(z_n);\ n = 1, 2, \ldots, N\}$ in a matrix form:

$$\begin{bmatrix} \mathsf{y} \\ d \end{bmatrix} = \begin{bmatrix} \Phi & \mathsf{z} \\ \mathbf{0}^\mathsf{T} & 1 \end{bmatrix} \begin{bmatrix} \mathsf{c} \\ d \end{bmatrix}, \tag{10}$$

where $\mathsf{y} = [y_1, \ldots, y_N]^\mathsf{T}$, $\mathsf{z} = [z_1, \ldots, z_N]^\mathsf{T}$, $\mathsf{c} = [c_1, \ldots, c_N]^\mathsf{T}$, and $\Phi = [\phi(|z_i - z_j|)]$. Solving the above linear system yields a unique solution $\{\mathsf{c}, d\}$, which stays fixed and is hence pre-computable. Another trick to accelerate evaluating (9) is to pre-compute a table of values $\phi(r)$ for all possible integers r within a proper range.

3. Experimental results

3.1. M-mode echocardiogram

The M-mode echocardiogram, from which the ventricular measurements are derived, is captured from two windows [5]: parasternal long axis and parasternal short axis. In each case, the ultrasound scan line first penetrates the chest wall, then the right ventricle, and finally the left ventricle. Seven measurements can be obtained from the M-mode echo:

1. RV internal dimension in diastole (RVIDd);
2. Interventricular septum thickness in diastole (IVSd);
3. LV internal dimension in diastole (LVIDd);
4. LV posterior wall thickness in diastole (LVPWd);

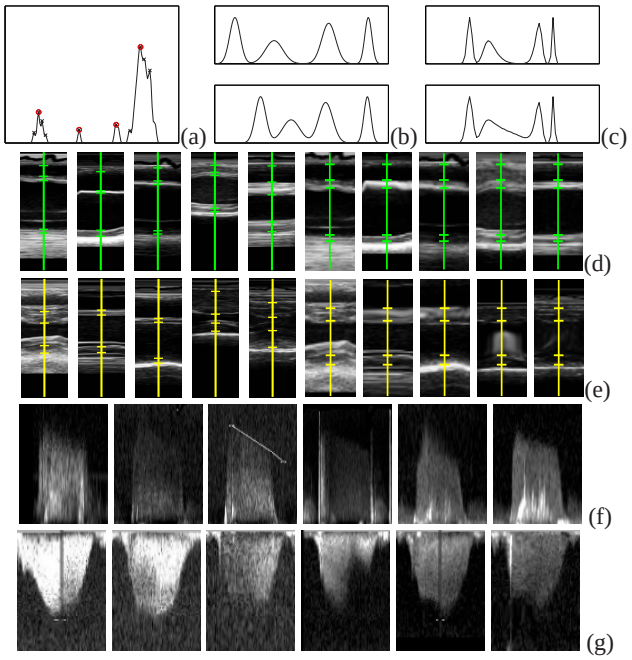

Figure 3. *(a) Graphical illustration of mode selection. (b,c) Graphical illustration of 1D TPS warping: (b) Two signals before warping and (c) two signals after warping. (d,e) M-mode Image examples before warping (the left five) and after warping (the right five) for the (d) ED and (e) ES lines. (f,g) Doppler image examples after warping of (f) aortic regurgitation and (g) tricuspid regurgitation.*

5. Interventricular septum thickness in systole (IVSs);
6. LV internal dimension in systole (LVIDs); and
7. LV posterior wall thickness in systole (LVPWs).

These measurements are derived only on the lines corresponding to the end of diastole (ED) and end of systole (ES), whose positions are either provided beforehand or reliably estimated based on the electrocardiogram (ECG). Therefore, the image analysis task is to accurately detect a cohort of landmarks on the given ED and ES lines: five landmarks on an ED line and four landmarks on an ES line.

We collected a library of 89 M-mode images that were annotated by an experienced sonographer. Depending on the heart rate and temporal sampling rate, each image contains 1-8 cardiac cycles, with each cardiac cycle contributing a pair of ED/ES lines (some might have missing ED or ES line). In total, there are 284 ED lines and 278 ES lines in the database. We randomly selected 70 images for training and the remaining 19 for testing and repeated this random selection three times for the *cross validation* purpose.

Since the heart rate and the temporal sampling rate are known, we normalized the size of each image in such a way that each cardiac rate spans about the same size (125 pixels or so) in the x-direction. We also normalized the y-direction such that each pixel corresponds to 0.5mm in depth; we kept 400 pixels to cover a range from zero to 20cm. After normalization, we also padded the images (5 pixels in each direction) for searching convenience and used the cyan color detector to remove the ECG signal line. Figure 4(a,b) shows an example of size normalization.

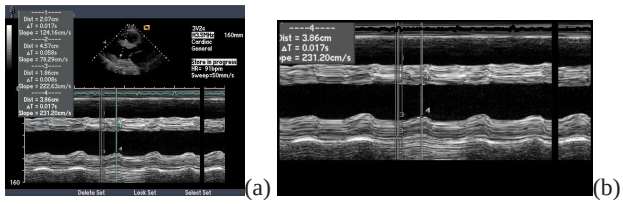

Figure 4. *M-mode echocardiogram: (a) the original image and (b) the size-normalized image. Note that the sizes of the original and normalized images are in proportion.*

The prior search ranges for landmarks on the normalized domain are empirically determined based on the database. To learn the local classifier, we cropped image patches of size 51×51 around the ground truth positions (there is a ± 1 perturbation) as positives and those 5 pixels away but within the search range as negatives. After learning, the PBTs for landmarks use about 400 weak classifiers. To learn the global detector, we again perturbed all landmarks around the ground truth positions and warped them to canonical positions in the template of size 75×51 to generate positives. Negatives were similarly generated by forcing at least one landmark 5 pixels away from the true positions. Here the landmark positions must be positives from the 1st layer landmark detector. The PBTs for global templates include about 1000 weak classifiers.

During detection, we kept a maximum of five top modes for each landmarks, resulting in, on average, about 1000 image warping per ED/ES line. Speed wise, it takes about 300ms to process one image containing about 3 cardiac cycles on a standard PC with 3GHz Xeon CPU and 3GB memory: about 100ms on size normalization, 150ms on local detection, and 50ms on global detection. If an image contains multiple cardiac cycles, we computed the median of measurements from multiple cycles as the final output.

We used the absolute errors in landmark localization and measurements to calibrate the performance. Table 2(a) tallies the experimental results by pooling together three batches of testing sets. Collectively, there are 185 ED lines and 187 ES lines for testing, and 57 data points per measurement. From Table 2, we observe that (i) all results for different landmarks are quite consistent except a few outliers; (ii) the overall median absolute error in landmark localization is 0.0570cm and the mean error is 0.0625cm, amounting to a subpixel error in the original image; and (iii) as we take the median for the measurements, they have less outliers as reflected by their smaller standard deviations. Figure 5 shows the detection results: The detected landmarks are very close to the true ones despite significant variances in landmark

positions and image intensities. Our preliminary clinical evaluation shows a good correlation between the detected results and a consensus ground truth.

For comparison, we implemented the approach that finds the landmarks independently. Table 2(b) tabulates the results, which are much worse than those in Table 2(a). In terms of computation, it is faster than the proposed approach, taking about 50ms less to process one image.

(Unit: cm)	L1d	L2d	L3d	L4d	L5d
median	0.0570	0.0326	0.0586	0.0669	0.0586
mean	0.0698	0.0427	0.0531	0.0905	0.0715
std. dev.	0.1592	0.0526	0.0593	0.1129	0.0804
		RVIDd	IVSd	LVIDd	LVPWd
median		0.0359	0.0628	0.0669	0.0669
mean		0.0679	0.0640	0.0863	0.0841
std. dev.		0.1200	0.0497	0.0816	0.0879
		L1s	L2s	L3s	L4s
median		0.1172	0.0807	0.2278	0.1172
mean		0.2437	0.1906	0.4055	0.1835
std. dev.		0.4543	0.3610	0.5703	0.2367
			IVSs	LVIDs	LVPWs
median			0.0570	0.0586	0.0628
mean			0.0596	0.0648	0.0785
std. dev.			0.0532	0.0708	0.0819

(a)

(Unit: cm)	L1d	L2d	L3d	L4d	L5d
median	0.1757	0.1256	0.0807	0.2008	0.0807
mean	0.3864	0.2959	0.2675	0.3844	0.1841
std. dev.	0.5413	0.6303	0.7109	0.5838	0.4155
		RVIDd	IVSd	LVIDd	LVPWd
median		0.1614	0.1339	0.1883	0.1464
mean		0.2779	0.3459	0.5755	0.2678
std. dev.		0.3754	0.6089	0.9248	0.3580
		L1s	L2s	L3s	L4s
median		0.1614	0.1339	0.1883	0.1464
mean		0.2779	0.3459	0.5755	0.2678
std. dev.		0.3754	0.6089	0.9248	0.3580
			IVSs	LVIDs	LVPWs
median			0.1506	0.2422	0.2278
mean			0.2850	0.4588	0.4071
std. dev.			0.4551	0.5457	0.5296

(b)

Table 2. *The absolute testing errors in landmark localization and measurements obtained by (a) the proposed algorithm and (b) the local approach.*

3.2. Doppler echocardiogram

The goal is to derive automated measurements of Doppler spectra of the blood flow in the heart. The algorithm developed should be robust and general enough so that it performs well despite the large variation of spectral shapes observed in everyday clinical practice. There are only a few methods in the literature focused on the velocity envelop extraction problem [17, 18]. All these method relied on image processing/filtering techniques, whose robustness is not guaranteed.

Specifically, we dealt with three types of flows so far: (a) mitral inflow, (b) aortic regurgitation, and (c) tricuspid regurgitation. The same framework can be easily applied for detecting doppler structures associated with other types of flows, such as tricuspid inflow, mitral outflow, pulmonic regurgitation, given that they possess similar visual patterns to those we processed.

The inflow patterns through the mitral and tricuspid valves are similar, consisting of the E and A waves. A trace of the envelope will be required as well as identification of the peaks and the trough of the structure. For our purpose, it is sufficient to represent the E/A wave envelope using a triangle [5] (p. 170). The regurgitation jets from aortic and pulmonary valves have similar appearance. The measurements do not use the full trace but only a fit to a straight line of the sloping part of the spectrum [5] (p. 300). Nevertheless, we decided to detect the quadrilateral. The regurgitation jets from the mitral and tricuspid valves have a different appearance from aortic and pulmonary valves. The trace of these regurgitant jets is complicated when portions of the jet are not visible which is quite common [5] (p. 163).

We collected 153 mitral inflow, 43 aortic regurgitation and 147 tricuspid regurgitation images for training and 46 mitral inflow, 6 aortic regurgitation and 28 tricuspid regurgitation images for testing. The number of doppler structures varies significantly from image to image: 2 to 20 triangles per image, 1 to 7 quadrilaterals per image, and 3 to 5 curves per image. Table 3 provides the data statistics of our training and testing data sets.

We performed the size normalization only along the x-direction to compensate the discrepancies in the heart rate and the temporal sampling rate. After normalization, we also padded the images (50 pixels in each direction) and removed the ECG signal line.

The list of primitive detectors along with their number of weak classifiers is given in Table 1. As expected, the root detector is the simplest while the warping detector is the most complicated. When designing the hierarchy, the main concern is the computation. Landmark/root scanning is both reliable (except the mitral inflow case) and fast, so we used it as the first layer. Since the warping is the most time consuming part, we always left it as the last layer, if used. In addition, mode selection is always turned on to accelerate the computation.

In order to further reduce the number of warping candidates in testing, we stored a code book of all possible warping possibilities (using the relative parameterization with respect to the bounding box) in the memory. We also added slight perturbations of the parameter values to increase robustness. For example, we stored 460 prior warping parameters for the aortic regurgitation case even though there are only 93 structures. This way, we avoided a full-range search of the parameter used for warping.

It is likely to have a cluster of overlapping detection results close to the ground truth. Among the cluster, we singled out the detection result with the maximum detection probability as the output. Even after cluster removal, it is still possible to have severely overlapping results. If this happens, we heuristically selected the one with the maximum peak velocity as the final result.

To quantify the performance of our algorithm, we extracted two key clinical measurements from the deformable structure: peak velocity (PV) and velocity time integral (VTI). The PV is the maximum velocity achieved by the flowing blood; the VTI is the area under the velocity curve of one cardiac cycle. We also measure the standard area overlapping ratio (OR) to gauge the detection accuracy, i.e., $OR = 2 * area(A \cap B)/(area(A) + area(B))$. Table 3 reports the testing results. In terms of the area overlapping ratio, we achieved above 90% in the median for all three Doppler structures. We are currently doing clinical validation and preliminary study shows that the detected results are within the user variability with respect to a consensus ground truth. We also benchmarked the computational speed on the same machine and recorded the target of less than one second. Figure 5 displays the detection results.

	Mitral inflow	Aortic reg.	Tricuspid reg.
structure	triangle	quadrilateral	curve
# of training images	153	43	147
# of training structures	698	93	367
# of test images	46	6	28
# of test structures	176	15	53
# of structures/image	2-20	1-7	1-5
mean of PV (m/s)	1.02±0.50	2.90±0.69	2.72± 0.91
med. $\|dPV\|$ (m/s)	0.035	0.031	0.090
mean $\|dPV\|$ (m/s)	0.041	0.063	0.194
std. dev. $\|dPV\|$ (m/s)	0.049	0.098	0.260
mean of VTI (m)	0.19±0.21	1.04±0.18	0.86±0.36
med. $\|dVTI\|$ (m)	0.016	0.021	0.059
mean $\|dVTI\|$ (m)	0.021	0.033	0.133
std. dev. $\|dVTI\|$ (m)	0.024	0.043	0.218
med. area OR	90.1%	97.2%	93.0%
mean area OR	89.3%	96.9%	89.0%
std. dev. area OR	7.6%	5.2%	9.3%
# of false alarms	0/46	0/6	1/28
# of miss	0/46	2/6	1/28
avg. det. time (ms)	482	672	971

Table 3. *Data statistics and detection performance for the Doppler echocardiogram.*

4. Conclusion

We have presented a generic PHD framework for detecting deformable anatomic structure from medical images. The probabilistic framework integrates evidence from different primitive levels via a progressive detector hierarchy, consisting of a series of discriminative classifiers. The PHD framework, if its hierarchy is carefully designed, supports the two contradictory tasks of fast evaluation and accurate detection. We have demonstrated the effectiveness of the framework on various heterogenous tasks of detecting a cohort of landmark, triangles, quadrilaterals, and curved from M-mode and Doppler echocardiograms.

References

[1] S. Agarwal, A. Awan, and D. Roth. Learning to detect objects in images via a sparse, part-based representation. *IEEE Trans. PAMI*, 26:1475 – 1490, 2004. 2

[2] F. L. Bookstein. Principal warps: Thin-plate splines and the decomposition of deformation. *IEEE Trans. PAMI*, 11(6):567–585, 1989. 4

[3] J. Coughlan and S. Ferreira. Finding deformable shapes using loopy belief propagation. In *European Conf. Computer Vision*, 2002. 2

[4] D. Crandall, P. Felzenszwalb, and D. Huttenlocher. Spatial priors for part-based recognition using statistical models. In *Proc. of CVPR*, 2005. 2

[5] H. Feigenbaum, W. Armstrong, and T. Ryan. *Feigenbaum's Echocardiography*. Lippincott Williams & Wilkins, 2005. 1, 4, 6

[6] P. Felzenszwalb. Representation and detection of deformable shapes. *IEEE Trans. PAMI*, 27, 2005. 2

[7] R. Fergus, P. Perona, and A. Zisserman. Object class recognition by unsupervised scale-invariant learning. In *Proc. of CVPR*, 2003. 2

[8] Y. Freund and R. Schapire. A decision-theoretic generalization of online leaning and an application to boosting. *J. Computer and System Sciences*, 55(1):119. 3

[9] A. Garg, S. Agarwal, and T. Huang. Fusion of global and local information for object detection. In *Proc. Int'l Conf. Pattern Recognition*, 2002. 2

[10] A. Holub and P. Perona. A discriminative framework for modeling object class. In *Proc. of CVPR*, 2005. 2

[11] M. Kass, A. Witkin, and D. Terzopoulos. Snakes: Active contour models. *Int. J. Computer Vision*, 1:321–331, 1988. 2

[12] T. McInerney and D. Terzopoulos. Deformable models in medical image analysis: A survey. *Medical Image Analysis*, 1:91–108, 1996. 2

[13] A. Mohan, C. Papageorgiou, and T. Poggio. Example-based object detection in images by components. *IEEE Trans. PAMI*, 23:349. 2, 3

[14] C. Papageorgiou, M. Oren, and T. Poggio. A general framework for object detection. In *Proc. ICCV*, 1998. 3

[15] M. Riesenhuber and T. Poggio. Hierarchical models of object recognition in cortex. *Nature Neuroscience*, 2, 1999. 2, 3

[16] S. Sclaroff and L. Liu. Deformable shape detection and description via model-based region grouping. *IEEE Trans. PAMI*, 23:475. 2

[17] O. Shechner, M. Scheinowitz, M. Feinberg, and H. Greenspan. Automated method for doppler echocardiography image analysis. In *Proc. IEEE Convention of Electrical and Electronics Engineers in Israel*, pages 177 – 180, 2004. 6

[18] J. Tschirren, R. Lauer, and M. Sonka. Automated analysis of doppler ultrasound velocity flow diagrams. *IEEE Transactions on Medical Imaging*, 20:1422, 2001. 6

[19] Z. Tu. Probabilistic boosting-tree: Learning discriminative models for classification, recognition, and clustering. In *Proc. of ICCV*, 2005. 2, 3

[20] P. Viola and M. Jones. Rapid object detection using a boosted cascade of simple features. In *Proc. of CVPR*, 2001. 2, 3

[21] D. Zhang and S.-F. Chang. A generative-discriminative hybrid method for multi-view object detection. In *Proc. of CVPR*, 2006. 2

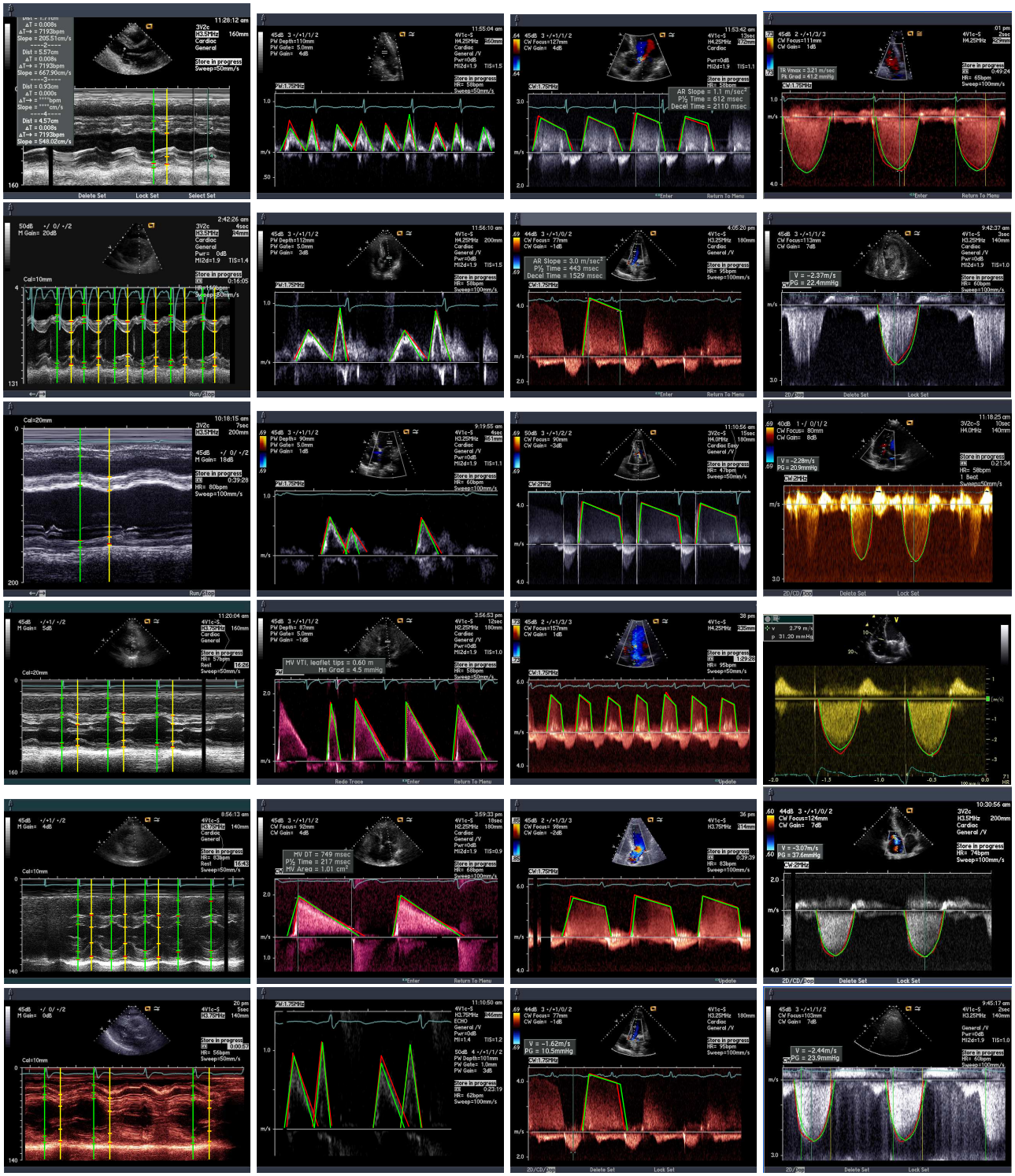

Figure 5. *The detection results (in green/yellow) versus the ground truth (in red). The 1st row: M-mode, the 2rd row: mitral inflow, the 3rd row: aortic regurgitation and the last column: tricuspid regurgitation.*