Simultaneous Detection and Registration for Ileo-Cecal Valve Detection in 3D CT Colonography

Le Lu[1,2] Adrian Barbu[1] Matthias Wolf[2] Jianming Liang[2]
Luca Bogoni[2] Marcos Salganicoff[2] Dorin Comaniciu[1]

[1]Integrated Data Systems Dept., Siemens Corporate Research, Princeton, NJ 08540
[2]Computer Aided Diagnosis Group, Siemens Medical Solutions USA, Malvern, PA 19355

Abstract. Object detection and recognition has achieved a significant progress in recent years. However robust 3D object detection and segmentation in noisy 3D data volumes remains a challenging problem. Localizing an object generally requires its spatial configuration (i.e., pose, size) being aligned with the trained object model, while estimation of an object's spatial configuration is only valid at locations where the object appears. Detecting object while exhaustively searching its spatial parameters, is computationally prohibitive due to the high dimensionality of 3D search space. In this paper, we circumvent this computational complexity by proposing a novel framework capable of incrementally learning the object parameters (IPL) of location, pose and scale. This method is based on a sequence of binary encodings of the projected true positives from the original 3D object annotations (i.e., the projections of the global optima from the global space into the sections of subspaces). The training samples in each projected subspace are labeled as positive or negative, according their spatial registration distances towards annotations as ground-truth. Each encoding process can be considered as a general binary classification problem and is implemented using probabilistic boosting tree algorithm. We validate our approach with extensive experiments and performance evaluations for Ileo-Cecal Valve (ICV) detection in both clean and tagged 3D CT colonography scans. Our final ICV detection system also includes an optional prior learning procedure for IPL which further speeds up the detection.

1 Introduction

Detecting and segmenting human anatomic structures in a 3D medical image volume (*e.g.,* CT, MRI) is very challenging. It demonstrates different aspects of difficulties as 2D counterparts of occlusion, illumination and camera configuration variations (for instance, rotation-invariant, single-view or multi-view 2D face detection [9, 15, 4, 6, 10]). Human anatomic structures are highly deformable by nature, which leads to large intra-class shape, appearance and pose variation. However only a limited number of patient image volumes are available for training. Another important issue is that the pose of the anatomic structure for detection is generally unknown in advance. If we knew the pose as a prior, the detection problem would be easier because we can train a model for anatomic structures under a fixed pose specification and pre-align all testing data (w.r.t. the known pose) to then evaluate their fitness values using the learned model.

However we always face a chicken-and-egg problem in practice. When estimating the pose configuration, the structure itself must be first detected and localized because pose information is only meaningful in the area where the object exists. In this paper, our goal is to localize and segment an anatomic structure using a bounding box under a full 3D spatial configuration (i.e., 3D translation, 3D scaling and 3D orientation)

Exhaustive search for 3D object detection and segmentation is infeasible, due to the prohibitive computational time required in 9D space. Naturally one would consider restricting the search space by concatenated subspaces. Since the global optima projections are not necessarily optima in the projected subspaces, such *naïve* projection strategies cannot guarantee to find the global optima. In this paper, we propose a novel learning framework to tackle this problem. In training, we encode the projections of "global optima" in the global parameter space to a sequence of subspaces as optima for learning. Thus the obtained classifiers can direct the searching sequentially back to "global optima" in testing.

Our encoding process is iterative. At each stage of encoding, we extract new training samples by scanning the object's configuration parameters in the current projected subspace, based on previously detected candidates/hypotheses from the preceding step. The distances of these extracted samples w.r.t. their corresponding labeled object annotations are then utilized to separate these training samples into positive or negative set. This ensures the projections of the global optima represented by positives in the subspace for training, so that the global optima can be sequentially detected through subspaces in testing. We repeat this process until the full object configuration parameter spaces are explored. Each encoding process is a general binary classification problem, and is specifically implemented using probabilistic boosting tree algorithm (*PBT*) [12].

We demonstrate the validity of our approach with the application on 3D object detection: fully automated Ileo-Cecal Valve[1] (ICV) detection in 3D computed tomography (CT) volumes. However our technique is generally applicable to other problems as 3D object extraction in range-scanned data [3] or event detection in spatial-temporal video volumes [7, 1]. For event detection [7, 1], only subvolumes with very pre-constrained scales and locations in video are scanned for evaluation due to computational feasibility. Our 3D detection method allows full 9 degree-of-freedoms (DOF) of searching to locate the object/event with optimal configurations (3D for translation, 3D for rotation and 3D for scales).

Comparing with our previous empirical approach for cardiac heart segmentation [19], this paper develops an explicit, formal mathematical formulation for the core object detection and parameter learning algorithm (see section 2). It also presents a more intuitive interpretation, theoretical insights and convergence analysis in section 4. The task of ICV detection in 3D colonography is more challenging than the organ localization in [19], without considering its boundary delineation. The rest of this paper is organized as follows. We give the mathematical formulation of proposed incremental parameter learning (IPL) algorithm in section 2 followed by the application on ICV

[1] Ileo-Cecal Valve (ICV) is a small, deformable anatomic structure connecting the small and large intestine in human body. In addition to its significant clinical value, automated detection of ICV is of great practical value for automatic colon segmentation and automatic detection of colonic cancer in CT colongraphy (CTC) [11, 17, 5]

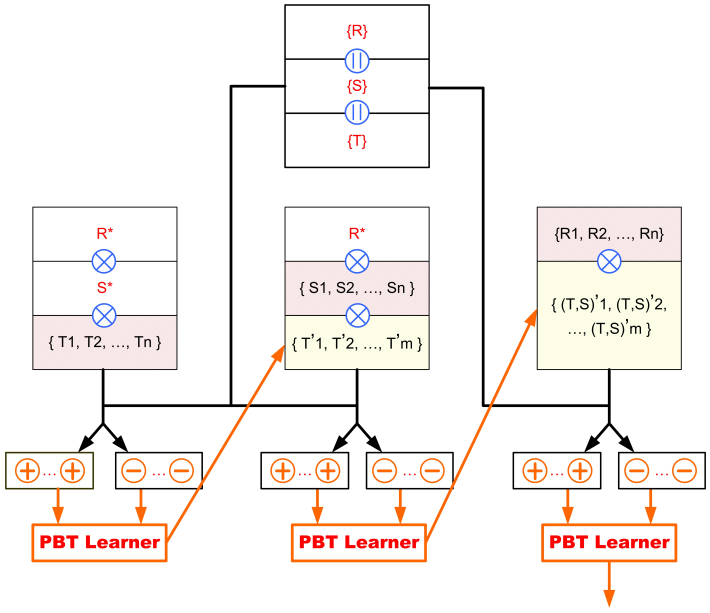


Fig. 1. Algorithm framework of incremental parameter learning (IPL) by projections in a full 3D space including 3D translations, 3D rotations (poses) and 3D scales. The parameter box on the top row represents the ground truth, or the global optimal solution in searching. In the second row, left, center and right boxes show how the object spatial parameters are incrementally learned from translation, scale, to rotation. $\|$ means one-to-one corresponding parameter augmentation, and $\times$ means *Cartesian* product in $\Omega_T, \Omega_S, \Omega_R$ parameter spaces.

detection in section 3 and its evaluation in section 4. We conclude the paper with discussion in section 5.

2 Incremental Parameter Learning

For noisy 3D medical data volumes, the scanning or navigation processes of finding interested objects can be very ambiguous and time-consuming for human experts. When the searched target is partially or fully coated by other types of noisy voxels (such as colonic objects embedded within stool, or tagging materials in CT), 3D anatomic structure detection by human experts becomes extremely difficult and sometimes impossible. These characteristics make it very necessary to solve the type of problems using computer aided detection and diagnosis (CAD) system for clinic purpose. This is the main motivation for our paper.

The diagram of our proposed incremental parameter learning (IPL) framework is shown in figure 1, by taking a full 3D object detection problem as an illustrative example. We define the detection task as finding a 3D bounding box including the object in 3D data volume as closely as possible. The object's (or the box's) spatial configuration space Ω can be uniquely determined by its 3D (center) position (Ω_T), 3D size (Ω_S) and

3D pose (rotation angles Ω_R). However the prohibitive computational expense makes it impossible for the direct searching (ie. scanning and verifying) strategy in this total 9D space[2]. To address the computational feasibility, we decompose the 9D parameter searching or learning process into three 3D steps: location finding (Ω_T), followed by size adjustment (Ω_S) and orientation estimation (Ω_R). The general searching strategy in sequentially decomposed subspaces can cause undesirable, sub-optimal solutions because the global optima are not necessary to be optimal in the decomposed dimensions as well. In this paper, we propose an incremental parameter learning framework to tackle this problem with guaranteed training performance using ROC curves analysis of multiple steps. In each step a "detection (using the detector from previous step)-sampling-registration-training (the detector in the current step)" scheme is applied, as explained later. In more detail, we formulate the following incremental parameter subspaces

$$\Omega_1 : \{\Omega_T\} \subset \Omega_2 : \{\Omega_T, \Omega_S\} \subset \Omega_3 : \{\Omega_T, \Omega_S, \Omega_R\} \tag{1}$$

where $\Omega_3 = \Omega$, or

$$\Omega_1 \subset \Omega_2 \subset ... \subset \Omega_n = \Omega \tag{2}$$

more generally. In equation 1, the order of Ω_S, Ω_R is switchable, but Ω_T needs to be first learned. The object's size and pose configurations can only be optimized where object is found.

For training, a set of 3D objects are labeled with their bounding boxes $\{T, S, R\}$. Without loss of generality, we assume that there is only one true object in each 3D data volume. In the first step, we search into Ω_T by scanning n samples $\{T_1, T_2, ...T_n\}$ around the true object positions $\{T\}$ and set parameters Ω_S, Ω_R with the mean values S^*, R^* of $\{S\}$ and $\{R\}$ as priors. Prior learning itself is a general and important computer vision problem. The mean-value (or median) prior setting is the simplest but not necessary the only or optimal choice of formulation, which is selected for representation clarity in this section. For example, a more natural option is prior sampling from the distribution formed by annotation parameters. In this paper, as an optional, more problem-specific treatment, the prior configuration of ICV detection can be learned from its informative orifice surface profiles and other side information using the same training/detection strategy.

First, we compute the distances $dist((T_i, S^*, R^*), (T_t, S_t, R_t)), i = 1, 2, ..., n$ between each of the sampled box candidates $\{(T_1, S^*, R^*); (T_2, S^*, R^*); ...; (T_n, S^*, R^*)\}$ and the annotated object bounding box (T_t, S_t, R_t) as its corresponding ground truth in the same volume. The translational distance metric $dist((T_i, S^*, R^*), (T_t, S_t, R_t))$ is computed as the center-to-center *Euclidean* distance

$$dist((T_i, S^*, R^*), (T_t, S_t, R_t)) = \parallel C_i - C_t \parallel \tag{3}$$

where C_i is the geometrical center of the sampling box (T_i, S^*, R^*) and C_t for the ground truth box (T_t, S_t, R_t). Then the box samples $\{(T_1, S^*, R^*); (T_2, S^*, R^*); ...;$

[2] Assume that the searching step is M in each dimension, and the overall cost will be M^9. If $M = 20$, the searching cost will be 512 billion times! Our target gain is M^6 here.

$(T_n, S^*, R^*)\}$ are divided into positive Φ_T^+ if

$$dist((T_i, S^*, R^*), (T_t, S_t, R_t)) < \theta_1 \tag{4}$$

or negative training set Φ_T^- if

$$dist((T_i, S^*, R^*), (T_t, S_t, R_t)) > \theta_2 \tag{5}$$

where $\theta_2 > \theta_1$. Φ_T^+ and Φ_T^- are learned using our implementation of a boosting based probabilistic binary learner (*PBT* [12]). Steerable features [19] are computed from each 3D bounding box and its including volume data for PBT training. After this, the output classifier $\mathbb{P}_T$ is able to distinguish sampled (in training) or scanned (in testing) object boxes: higher positive-class probability values (close to 1) for boxes which are close to their respective labeled object boxes, lower values (close to 0) for boxes that are distant. For computational efficiency, only top M candidates are retained as $\{(T'_1, S^*, R^*); (T'_2, S^*, R^*); ...; (T'_m, S^*, R^*)\}$ with highest output probabilities. If there is only one existing object per volume (such as ICV) and the training function can be perfectly learned by a classifier, $M = 1$ is sufficient to achieve the correct detection. In practice, we set $M = 50 \sim 100$ for all intermediate detection steps to improve robustness. It means that we maintain multiple detected hypotheses until the final result.

We then use these M intermediate detections as a basis to search in the next step. Each candidate $(T'_i, S^*, R^*), i = 1, 2, ..., M$ is augmented as n samples: $\{(T'_i, S_1, R^*); (T'_i, S_2, R^*); ...; (T'_i, S_n, R^*)\}$. Overall $M \times n$ box candidates are obtained. Similarly, they are divided into positive Φ_S^+ if

$$dist((T'_i, S_j, R^*), (T'_t, S_t, R_t)) < \tau_1 \tag{6}$$

or negative training set Φ_S^- if

$$dist((T'_i, S_j, R^*), (T'_t, S_t, R_t)) > \tau_2 \tag{7}$$

for $i = 1, 2, ..., M$ and $j = 1, 2, ..., n$. $dist((T'_i, S_j, R^*), (T_t, S_t, R_t))$ is defined as a box-to-box distance function which formulates 3D box differences in both Ω_T and Ω_S. More generally,

$$dist(box_1, box_2) = \sum_{i=1,2,...,8} \{\parallel v_1^i - v_2^i \parallel\}/8 \tag{8}$$

where v_1^i is one of the eight vertices of box_1 and v_2^i is its according vertex of box_2.$\parallel v_1^i - v_2^i \parallel$ is the *Euclidean* distance between two 3D vectors v_1^i, v_2^i. Again PBT algorithm and steerable features are used for training to get $\mathbb{P}_S$.

In the third step, $\mathbb{P}_S$ is employed to evaluate the positive-class probabilities for $M \times n$ samples $\{T'_i, S_j, R^*)\}$,i=1,2,...,M; $j = 1, 2, ..., n$, and keep a subset of M candidates with the highest outputs. We denote them $\{(T'_i, S'_i, R^*)\}, i = 1, 2, ..., M$, which are further expanded in Ω_R as $\{(T'_i, S'_i, R_j)\}, i = 1, 2, ..., M; j = 1, 2, ..., n$. After this, all the process is the same for training dataset construction and classifier training $\mathbb{P}_R$, as step 2. Box-to-box distance is employed and the two distance thresholds are denoted as η_1 and η_2. Finally we have $\{(T'_k, S'_k, R'_k)\}, k = 1, 2, ...M$ returned by our whole

algorithm as the object detection result of multiple hypotheses. In testing, there are three searching steps in Ω_T, Ω_S and Ω_R, according to the training procedure. In each step, we can scan and detect 3D object box candidates which are close to the global optimum (i.e., the object's true spatial configuration) in the current parameter subspace ($\Omega_1 \rightarrow \Omega_2 \rightarrow \Omega_3$), using the learned classifier ($\mathbb{P}_T$, $\mathbb{P}_S$ or $\mathbb{P}_R$) respectively. The output candidates are used as seeds of propagation in the next stage of incremental, more accurate parameter optimization. The training samples at each step are expanded and bootstrapped using the detection results at its previous step (and the global annotations as reference). Note that we set smaller threshold margins,

$$(\theta_2 - \theta_1) > (\tau_2 - \tau_1) > (\eta_2 - \eta_1) \tag{9}$$

for more desirable object detection/registration accuracy as steps of detection proceed.

The above incremental parameter learning process for 3D object detection is illustrated in figure 1. The parameter spaces (Ω_T, Ω_S and Ω_R) before search (prior), during search (learning/optimizing) and after search (optimized) are displayed in red, yellow and white shadows respectively. The mean parameter values T^*, S^*, R^* estimated from the labeled object annotations, are used as prior by default.

3 Ileo-Cecal Valve (ICV) Detection in 3D CT Colonography

Detecting Ileo-Cecal Valve (ICV) in 3D CT volumes is important for accurate colon segmentationand colon polyp false positive reduction [11, 17, 5] that are required by colon CAD system. Nevertheless, it is very challenging in terms of ICV's huge variations in its internal shape/appearance and external spatial configurations: $(X, Y, Z; S_x, S_y, S_z; \psi, \phi, \omega)$, or $(\Omega_T; \Omega_S; \Omega_R;)$. ICV is a relatively small-scaled (compared with heart, liver, even kidney) and deformable human organ which opens and closes as a valve. The ICV size is sensitive to the weight of patient and whether ICV is diseased. Its position and orientation also vary of being a part of colon which is highly deformable. To address these difficulties, we develop a two-staged approach that contains the prior learning of IPL to prune ICV's spatial configurations in position and orientation, followed by the position, size and orientation estimation of incremental parameter learning. Figure 2 shows the diagram of our final system. To validate the proposed incremental parameter learning of Ω_T Ω_S Ω_R, an ICV detection system without prior learning is also experimentally evaluated.

3.1 Features

In the domain of 3D object detection, 3D Haar wavelet features [13] are designed to capture region-based contrasts which is effective to classification. However 3D Haar features are inefficient for object orientation estimation because they require a very time-consuming process of rotating 3D volumes for integral volume computation. In steerable features [19], only a sampling grid-pattern need to be translated, rotated and re-scaled instead of data volumes. It allows fast 3D data evaluation and has shown to be effective for object detection tasks [19]. It is composed by a number of sampling

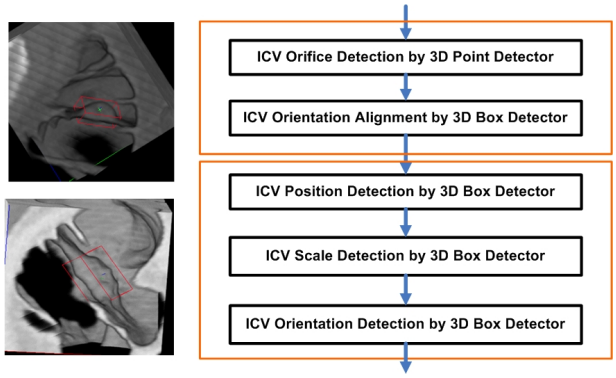


Fig. 2. System diagram of Ileo-Cecal Valve detection. The upper block is prior learning and the lower block is incremental parameter learning for ICV spatial parameter estimation. Examples of the annotated ICV bounding boxes are shown in red.

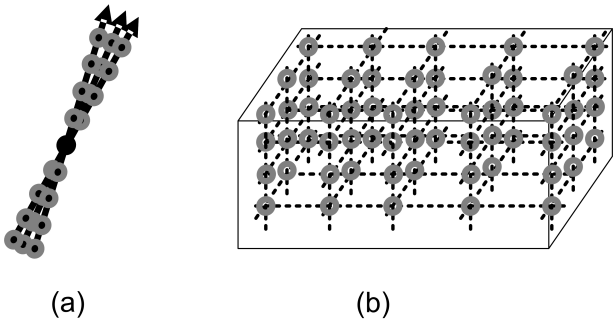
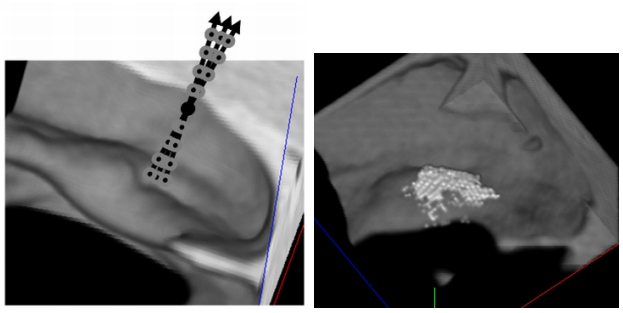


Fig. 3. Steerable sampling grid patterns for (a) 3D point detector and (b) 3D box detector.

grids/points where 71 local intensity, gradient and curvature based features are computed at each grid. The whole sampling pattern models semi-local context. For details, refer to [19].

In this paper, we design two specific steerable patterns for our ICV detection task as shown in figure 3. In (a), we design an axis-based pattern for detecting ICV's orifice. Assume that the sampling pattern is placed with its center grid at a certain voxel v. It contains three sampling axes as the gradient directions averaged in v's neighborhoods under three scales respectively. Along each axis, nine grids are evenly sampled. This process is repeated for halfly and quarterly downsampled CT volumes as well. Altogether we have $M = 81 = 3 \times 9 \times 3$ grid nodes which brings $71 \times 81 = 5751$ features. In (b), we fit each box-based pattern with evenly $7 \times 7 \times 5$ sampling grids. The total feature number is 52185 by integrating features from three different scales. This type of feature is used for all Ω_T Ω_S Ω_R detection. The detector trained with axis pattern and PBT is named 3D point detector; while the detector with box pattern and PBT is noted as 3D box detector.



(a) (b)

Fig. 4. (a) ICV orifice sampling pattern of three sampling axes and nine sampling grids along each axis; (b) detected ICV voxel/orifice candidates shown in white.

3.2 Prior Learning in Ω_T and Ω_R of IPL

If likely hypotheses ICV's orifice can be found, its position in Ω_T can be constrained, then no explicitly exhaustive searching of position is needed. The ICV orifice has an informative, but far from fully unique, surface profile that can possibly indicates ICV location as multiple hypotheses. It also allows very efficient detection using a 3D point detector which involves less feature computation (5751 vs. 52185 for training) than a box detector. Further more, it is known that ICV orifice only lies on the colon surface that is computed using a 3D version of Canny edge detection. Thus we can prune all voxel locations inside the tissue or in the air for even faster scanning. An illustrative example of the orifice sampling pattern and detection result is shown in figure 4. Note that multiple clusters of detection may occur often in practice. From the annotated ICV orifice positions in our training CT volume set, we generate the positive training samples for surface voxels within α_1 voxel distance and negatives out of α_2 voxel distance. We set $\alpha_2 > \alpha_1$, so the discriminative boosting training [12] will not focus on samples with distances $[\alpha_1, \alpha_2]$ which are ambiguous for classifier training but not important for target finding. The trained classifier $\mathbb{P}_O$ is used to exhaustively scan all surface voxels, prune the scanned ICV orifice candidates and only a few hypotheses (eg. $N = 100$) are preserved. In summary, 3D point detector for ICV orifice detection is efficient and suitable for exhaustive search as the first step.

Given any detected orifice hypothesis, we place ICV bounding boxes centering at its location and with the mean size estimated from annotations. In the local 3D coordinates of an ICV box, XY plane is assumed to be aligned with the gradient vector of the orifice as its Z-axis. This is an important domain knowledge that we can use to initially prune ICV's orientation space Ω_R in 2 degrees of freedom (DOF). Boxes are then rotated around Z-axis with 10^o interval to generate training samples. Based on their box-to-box distances against the ground truth of ICV box[3] and β_1, β_2 threshold as above, our routine process is: (1)generating positive/negative training sets by distance

[3] The ground truth annotations are normalized with the mean size to count only the translational and orientational distances.

thresholding; (2) training a PBT classifier $\mathbb{P}_{R'}$ using the box-level steerable features; (3) evaluating the training examples using the trained classifier, and keeping top 100 hypotheses of probabilities ($\rho^i_{R'}, i = 1, 2, ..., 100$). In our experiments, we show results with $\alpha_1 = 4$, $\alpha_2 = 20$ (normally out of the ICV scope), $\beta_1 = 6$ and $\beta_2 = 30$.

3.3 Incremental Parameter Learning in Ω_T Ω_S Ω_R

In this section, we search for more accurate estimates of ICV position, scale and orientation parameter configurations. Incremental parameter learning method described in section 2 is implement. The box-level steerable features (as shown in figure 3(b)) and PBT classifier are employed for all three steps. From section 3.2 we obtain 100 ICV box hypotheses per volume with their positions and orientations pruned. Therefore we select the order of incremental parameter learning as $\Omega_T \rightarrow \Omega_S \rightarrow \Omega_R$, where Ω_T is always the first step to locate itself and Ω_S proceeds before aligned Ω_R.

First, the position of each of the N hypotheses is shifted every one voxel in the range of $[-20, 20]$ of all X, Y and Z coordinates (ie. $\Omega_T + \Delta_T$). This set of synthesized ICV box samples is then splitted into the positive ($< \theta_1 = 5$ voxel distance) and negative ($> \theta_2 = 25$ voxel distance) training sets for the PBT training of $\mathbb{P}_T$. Again the top 100 ICV box candidates in each CT volume (with the largest probability outputs ρ^i_T using $\mathbb{P}_T$) are maintained. Next, the optimal estimates of ICV box scales are learned. We set the size configuration of each survived hypotheses in Ω_S, evenly with 2 voxel intervals from the range of $[23, 51]$ voxels in X, $[15, 33]$ voxels in Y and $[11, 31]$ voxels in Z coordinates. The ranges are statistically calculated from the annotated ICV dataset.

In the same manner, we train the classifier $\mathbb{P}_S$ and use it to obtain the top N candidates of ρ^i_S with more accurate estimates of Ω_S. The distance thresholds are $\tau_1 = 4$ and $\tau_2 = 20$ for positive/negative training respectively. Last, we adaptively add disturbances from the previously aligned orientation estimates in prior learning (ie. $\Omega_R + \Delta_R$). Δ_R varies with 0.05 intervals in $[-0.3, 0.3]$ radians, 0.1 in $([-0.9, -0.3), (0.3, 0.9])$ and 0.3 in $([-1.8, -0.9), (0.9, 1.8])$. This strategy provides a finer scale of searching when closer to the current orientation parameters (retained from $\mathbb{P}_{R'}$ in prior learning), to improve the Ω_R detection accuracy. $\mathbb{P}_R$ is learned with the distance thresholds as $\eta_1 = 4$ and $\eta_2 = 15$. After all steps of incremental parameter learning, the top one box candidate of the highest probability value from $\mathbb{P}_R$ is returned as the final ICV detection result by default.

Incremental parameter learning of $\Omega_T, \Omega_S, \Omega_R$ is equivalent to exhaustive search in $\Omega_T \cup \Omega_S \cup \Omega_R$ if we can train mathematically perfect classifiers (100% recall at 0% false positive rate) at all steps. This causes large positive within-class variations at early learning steps (e.g., detecting object location while tolerating unestimated poses and scales), which decreases trainability in general. Classifiers with intrinsic "divide-and-conquer" scheme as PBT [12] or cluster based tree [14] can be applied. In short, explicit exhaustive searching for parameter estimation is traded by implicit within-class variation learning using data-driven clustering [12, 14]. It also relaxes the requirement for training accuracy by keeping multiple hypotheses during detection. In case of multiple object detection, selecting top N candidates simply based on their class-conditional probabilities can not guarantee to find all objects since a single target may cause many detec-

tions. Possible approaches are to exploit cluster based sampling [8] or *Non-Maximum Suppression* by using the spatial locations of detected hypotheses.

4 Evaluation & Results

Convergence Analysis: The convergence analysis of incremental parameter learning method is first based on the property of Receiver Operating Characteristic (ROC) curves during five stages of training. The training scale for our PBT classifier ranges over $10K \sim 250K$ positives and $2M \sim 20M$ negatives. The ROC curves are shown in figure 5 (a). From the evidence of these plots, our training process are generally well-performed and gradually improves for later steps. We then discuss the error distribution curves between the top 100 ICV hypotheses maintained for all five stages of detection and the ground truth, using five-fold cross-validation. The error curves, as shown in figure 5 (b), also demonstrate that more accurate ICV spatial configurations can be obtained as the detection process proceed through stages. *This convergence is bounded by the good training performance of ROC curves with positive-class distance boundaries that are gradually more close to the global optima (or ground-truth) as* $6, 5, 4, 4$*, and decreasing distance margins between positive and negative classes (eg.* $\beta_2 - \beta_1 = 24$*;* $\theta_2 - \theta_1 = 20$*;* $\tau_2 - \tau_1 = 16$ *and* $\eta_2 - \eta_1 = 11$*) over stages.*

ICV Detection Evaluation: Our training set includes 116 ICV annotated volumes from the dataset of clean colon CT volumes using both Siemens and GE scanners. With a fixed threshold $\rho_R > 0.5$ for the final detection, 114 ICVs are found with the detection rate of 98.3%, under five-fold cross-validation. After manual examination, we find that the two missed ICVs have very abnormal shape from the general training pool which is probably heavily diseased. The ICV detection accuracy is first measured by a symmetric overlapping ratio between a detected box Box_d and its annotated ground truth Box_a

$$\gamma(Box_a, Box_d) = \frac{2 \times Vol(Box_a \bigcap Box_d)}{Vol(Box_a) + Vol(Box_d)} \tag{10}$$

where $Vol()$ is the box-volume function (eg. the voxel number inside a box). The accuracy distribution over 114 detected ICV examples is shown in 5 (c). The mean overlap ratio $\gamma(Box_a, Box_d)$ is 74.9%. This error measurement is directly relevant with our end goal of removing polyp-like false findings in my CAD system. Addtionally the mean and standard deviation of orientational detection errors are 5.89^o, 6.87^o, 6.25^o; and 4.46^o, 5.01^o, 4.91^o respectively for three axes. The distribution of absolute box-box distances (ie. equation 8) has 4.31 voxels as its mean value, and 4.93 voxels for the standard deviation. Two missed cases are further verified by clinician as heavily diseased ICVs which are rare in nature. Our trained classifiers treat them as outliers.

Next we applied our detection system to other previously unseen clean and tagged CT datasets. For clean data, 138 detections are found from 142 volumes. After manual validation, 134 detections are true ICVs and 4 cases are Non-ICVs. This results a detection rate of 94.4%. We also detected 293 ICVs from 368 (both solid and liquid) tagged colon CT volumes where 236 detections are real ICVs with 22 cases for Non-ICVs and 35 cases unclear (which are very difficult even for expert to make decision). Tagged CT data are generally much more challenging than clean cases, under low-contrast imaging

and very high noise level of tagging materials. Some positive ICV detections are illustrated in figure 6. The processing time varies from $4 \sim 10$ seconds per volume on a P4 3.2G machine with 2GB memory.

Without prior learning for ICV detection, our system can achieve comparable detection performance as with prior learning. However it requires about 3.2 times more computation time by applying a $3D$ box detector exhaustively on translational search, not a cheaper $3D$ point detector as in prior learning. Note that prior learning is performed in the exact same probabilistic manner as the incremental 3D translation, scale and orientation parameter estimation. It is not a simple and deterministic task, and multiple (e.g., 100) detection hypotheses are required to keep for desirable results.

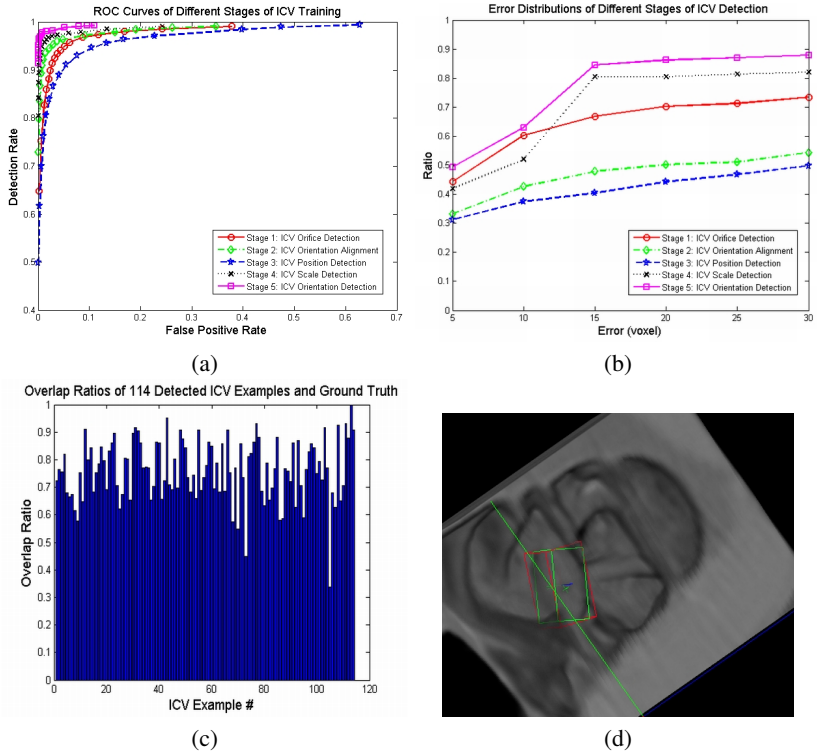


Fig. 5. (a) Receiver operating characteristic curves of different stages of training in our Ileo-Cecal Valve detection system. (b) Error ratio curves of top 100 ICV hypotheses of different stages of detection. Each curve show the ratios of hypotheses (Y axis) under the particular error readings (X-axis) against ground truth. All numbers are averaged over the testing sets of volumes, under five-fold cross-validation of 116 total labeled ICV examples. (c) Overlap ratios between 114 detected ICV examples and their ground truth. (d) A typical example of 3D ICV detection in CT Colonography, with overlap ratio of 79.8%. Its box-to-box distance as define in equation 8 is 3.43 voxels where the annotation box size is $29.0 \times 18.0 \times 12.0$ voxels. Its orientational errors are 7.68^o, 7.77^o, 2.52^o with respect to three axes. The red box is the annotation; the green box is the detection. This picture is better visualized in color.

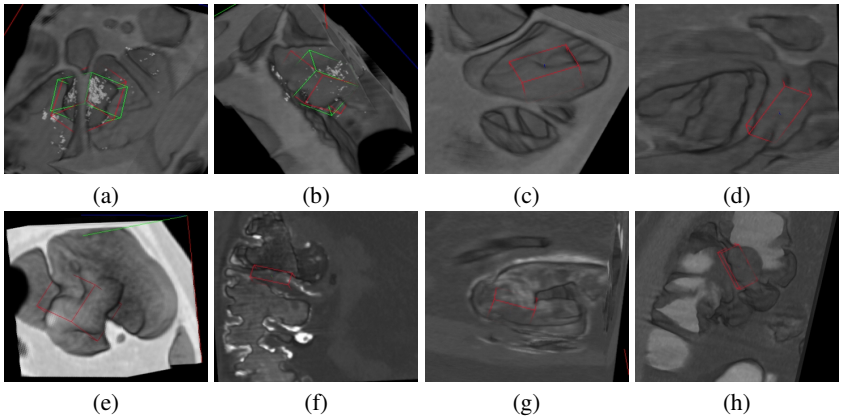


Fig. 6. (a,b) An example of ICV detection result from two viewpoints. The red box is the annotation; the green box is the detection. (c,d,e,f,g,h) Examples of ICV detection results from unseen clean colon CT volumes (c,d,e) and unseen solid (f) or liquid tagged (g,h) colon CT volumes. The red box is the final detection result where no annotation available. Note that only a CT subvolume surrounding the detected ICV box is visualized for clarity. This picture is better visualized in color.

Polyp False Positive (FP) Deduction: ICV contains many polyp-like local structures which confuse colon CAD system [11, 17, 5]. By identifying a reasonably accurate bound box for ICV, this type of ambiguous false positive polyp candidates can be removed. For this purpose, we enhanced the ICV orifice detection stage by adding the labeled polyp surface voxels into its negative training dataset. Other stages are consequentially retained in the same way. Polyp FP deduction is tested on 802 unseen CT volumes: 407 clean volumes from 10 different hospital sites acquired on Siemens and GE scanners; 395 tagged volumes, including iodine and barium preparations, from 2 sites acquired on Siemens and GE scanners. The ICV detection is implemented as post filter for our existing colon CAD system and only applied on those candidates that are labeled as "Polyp" in the preceding classification phases[4]. In clean cases, ICV detection reduced the number of false positives (fp) from 3.92 fp/patient (2.04 fp/vol.) to 3.72 fp/patient (1.92 fp/vol.) without impacting the overall sensitivity of the CAD system. It means that no true polyps were missed due to our ICV detection component integrated. In tagged cases, ICV detection reduced the number of false marks from 6.2 fp/patient (3.15 fp/vol.) to 5.78 fp/patient (2.94 fp/vol.). One polyp out of 121 polyps with a size range from 6 up to 25 mm was wrongly labeled as ICV, resulting in a sensitivity drop of 0.8%. Another version implementation of using ICV detection as a soft constraint, instead of a hard-decisioned post filter, avoids true polyp missing without sacrificing FP Deduction. In summary our ICV system achieved 5.8% and 6.7% false positive deduction rates for clean and tagged data respectively, which has significant clinical importance.

[4] Note that the use of ICV detection as post-process is dedicated to handle "difficult" polyp cases which can not be correctly classified in preceding processes.

Contextual K-Box ICV Model: To more precisely identify the 3D ICV region besides detection, a contextual K-box model is experimented. The idea is using the final ICV detection box B_1 as an anchor to explore reliable expansions. For all other high probability hypotheses $\{\hat{B}_i\}$ returned in the last step of detection, we sort them according to $Vol(\hat{B}_i - B_1 \bigcap \hat{B}_i)$ while two constraints are satisfied: $\gamma(B_1, \hat{B}_i) \geq \gamma_1$ and $\rho_R(\hat{B}_i) \geq \rho_1$. Then the box that gives the largest gain of $Vol(\hat{B}_i - B_1 \bigcap \hat{B}_i)$ is selected as the second box B_2. The two constraints guarantee that B_2 is spatially correlated with B_1 ($\gamma_1 = 0.5$) and is a highly likely ICV detection hypothesis by itself $\rho_1 = 0.8$. By taking B_1 and B_2 as a union $Box_d = B_1 \bigcup B_2$, it is straightforward to expand the model for K-box ICV model while $K > 2$. Our initial experimental results show that 2-box model improves the mean overlap ratio $\gamma(Box_a, Box_d)$ from 74.9% to 88.2% and surprisingly removes 30.2% more Polyp FPs without losing true polyps.

Previous Work on ICV Detection: Our proposed approach is the first reported, fully automatic Ileo-Cecal Valve detection system in 3D CT colonography, due to the difficulties discussed in sections 1 and 3. The closest previous work is by Summer et al. [11] that is also considered as the state-of-art technique in medical imaging community. We discuss and compare [11] and our work in two aspects. (1) For localization of ICV, Summer et al. relies on a radiologist to interactively identify the ICV by clicking on a voxel inside (approximately in the center of) the ICV. This is a requisite step for the next classification process and takes minutes for an expert to finish. On the contrary, our automatic system takes $4 \sim 10$ seconds for the whole detection procedure. (2) For classification, [11] primarily designs some heuristic rules discovered from dozens of cases by clinicians. It depends on the performance of a volume segmentor [16] which fails on $16\% \sim 38\%$ ICV cases [11]. Their overall sensitivity of ICV detection is 49% and 50% based on the testing (70 ICVs) and training datasets (34 ICVs) [11], respectively. This rule based classification method largely restricts its applicability and effectiveness on recognizing varieties of ICV samples with their low detection rates reported in [11]. Our detection rate is 98.3% for training data and 94.4% for unseen data. The superiority of our approach attributes to our effective and efficient incremental parameter learning framework optimizing object spatial configuration in a full 3D parameter space, and the discriminative feature selection algorithm (PBT + steerable features) exploring hundreds of thousands volume features.

5 Conclusion & Discussion

In this paper, we present the incremental parameter learning framework to address general 3D/2D object detection problem under high dimensional parameter spaces. The challenges are not only the computational feasibility, but also how to obtain good solutions in terms of the parameter searching complexity (essentially exponential to the dimension). The effectiveness of our method is demonstrated using an application on detecting Ileo-Cecal Valve (ICV) in 3D CT colonography with 9 DOF. To our best knowledge, ICV detection is the first fully automatic system for localizing a small (versus the whole CT volume dimension), largely deformable, unconstrainedly posed and possibly coated (by tagging material or stool in tagged volumes) 3D anatomic structure.

As a discussion, our proposed learning architecture is intuitively analogical to the famous twenty questions games, where many highly complex information extraction problems can be solved by using a flow of simpler, binary (yes/no), sequentially dependent testings (question vs. answer). We leave explorations on more sophisticated solution searching techniques [2, 18] as future work.

References

1. M. Blank, L. Gorelick, E. Shechtman, M. Irani and R, Basri, Actions as Space-Time Shapes, *ICCV*, 2005.
2. D. Geman, B. Jedynak, An Active Testing Model for Tracking Roads in Satellite Images, *IEEE Trans. Pattern Anal. Mach. Intell.*, 18(1):1-14, 1996.
3. F. Han, Z. Tu and S.C. Zhu, Range Image Segmentation by an Effective Jump-Diffusion Method. *IEEE Trans. PAMI*, 26(9), 2004.
4. C. Huang, H. Ai, Yuan Li and S. Lao. High-performance rotation invariant multiview face detection. *IEEE Trans. PAMI*, 29(4):671-686, 2007.
5. A. Jerebko, S. Lakare, P. Cathier, S. Periaswamy, L. Bogoni. Symmetric Curvature Patterns for Colonic Polyp Detection. *MICCAI* 2006.
6. M. Jones and P. Viola, Fast multi-view face detection, *CVPR*, 2003.
7. Y. Ke, R. Sukthankar, and M. Hebert, Efficient Visual Event Detection using Volumetric Features, *ICCV*, 2005.
8. L. Lu, G. Hager, Dynamic Background/Foreground Segmentation From Images and Videos using Random Patches, *NIPS*, 2006.
9. H. Rowley, S. Baluja, T. Kanade, Neural Network-Based Face Detection. *CVPR*, 1996.
10. H. Rowley, S. Baluja, T. Kanade, Rotation Invariant Neural Network-Based Face Detection, *CVPR*, 1998.
11. R. Summers, J. Yao, C. Johnson, CT Colonography with Computer-Aided Detection: Automated Recognition of Ileocecal Valve to Reduce Number of False-Positive Detections. *Radiology*, 233:266-272, (2004).
12. Z. Tu. Probabilistic boosting-tree: Learning discriminative methods for classification, recognition, and clustering. *ICCV*, 2005.
13. Z. Tu, X. S. Zhou, A. Barbu, L. Bogoni, D. Comaniciu. Probabilistic 3D polyp detection in CT images: The role of sample alignment. *CVPR*, 2006.
14. B. Wu, R. Nevatia. Cluster Boosted Tree Classifier for Multi-View, Multi-Pose Object Detection. *ICCV*, 2007.
15. P. Viola and M. Jones. Rapid object detection using a boosted cascade of simple features. *CVPR.*, pp. 511-518, 2001.
16. J. Yao, M. Miller, M. Franaszek and R. Summers, Colonic polyp segmentation in CT Colongraphy-based on fuzzy clustering and deformable models, *IEEE Trans on Medical Imaging*, 2004.
17. H. Yoshida and A. H. Dachman, CAD techniques, challenges, and controversies in computed tomographic colonography, *Abdominal Imaging, Springer*, 30(1):26-41, 2005.
18. Alan L. Yuille, James M. Coughlan, Twenty Questions, Focus of Attention, and A*: A Theoretical Comparison of Optimization Strategies, *EMMCVPR* 1997: 197-212.
19. Y. Zheng, A. Barbu, B. Georgescu, M. Scheuering and D. Comaniciu, Fast Automatic Heart Chamber Segmentation from 3D CT Data Using Marginal space Learning and Steerable Features, *ICCV*, 2007.