

A New Algorithm of Electronic Cleansing for Weak Faecal-Tagging CT Colonography

Le Lu, Bing Jian, Dijia Wu, and Matthias Wolf

Siemens Medical Solutions USA; Siemens Corporate Research (le.lu@nih.gov)

Abstract. CT Colonography (CTC) has emerged as a mainstream clinical practice of colonic cancer screening and diagnosis. One of the most critical problems is to increase compliance with CTC examinations via minimal bowel preparation (i.e., weak faecal-tagging), which nevertheless causes much lower signal-noise-ratio than conventional preparation.

In this paper, we present a new algorithm pipeline of electronically cleansing tagging materials in CTC under reduced oral contrast dose. Our method has the following steps: 1, robust structure parsing to generate a list of volume regions of interest (ROIs) of tagging material (avoiding bone erosion); 2, effectively locating local tagging-air (AT) transitional surface regions; 3, a novel discriminative-generative algorithm to learn the higher-order image appearance model in AT using 3D Markov Random Fields (MRF); 4, accurate probability density function based voxel labeling corresponding to semantic classes. Validated on 26 weak faecal-tagging CTC cases from 3 medical sites, our method yields better visualization clarity and readability compared with the previous approach [1]. The whole system computes efficiently (e.g., < 40 seconds for CT images of $512 \times 512 \times 1000+$).

1 Introduction

Colorectal cancer is the second leading cause of cancer-related mortality in western countries. Computed Tomography Colonography (CTC) has become a feasible, non-intrusive clinical alternative of traditional optical Colonography, to offer good sensitivity and specificity for polyp detection via screening the whole colon. However, similar to traditional examination protocols, the colon is required to be fully cleansed manually or electronically. Electronic cleansing (eCleansing) of fecal residues by computed algorithms [1, 2] is very critical to avoid intensive manual bowel preparation, prior to the examination. The quality of electronic cleansing method directly impacts radiologist’s image reading performance in 3D, as well as computer-aided diagnosis. Recently, weak faecal-tagging bowel protocol [3] has gained popularity in major research clinical sites (e.g., consuming up to 40% CTC cases) since the dose reduction largely improves patient comfort and compliance. This protocol however causes much higher imaging noises, inhomogeneous intensity patterns and lower contrast between tagging materials and surrounding soft-tissue structures. Significant challenges are imposed for radiologists and previous eCleansing methods [1, 2], due to lack of precise statistical data-driven appearance model. To address this issue, our proposed algorithm is described as follows. The overall framework is summarized in Fig. 1 (**Left**).

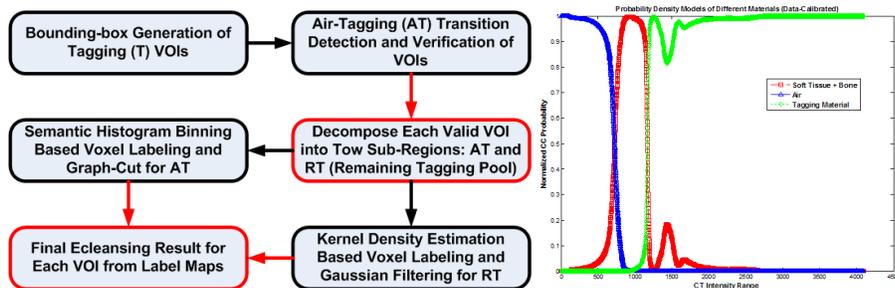


Fig. 1. (Left) The overall algorithm diagram of the proposed eCleansing approach. **(Right)** Data-calibrated class probability value plots $\in [0, 4095]$. Blue for Air class F_a ; Red for (soft-tissue + bone) class F_s ; and Green for tagging material class F_t .

2 Materials and Method

Data-driven Intensity Calibration: To accurately model the intensity variations of three anatomical classes: air, tagging and soft-tissue (including bone) in CT images, is an important prerequisite of eCleansing. In a data-driven manner, we manually label > 100000 voxels per class and perform kernel density estimation (KDE) $f_a; f_s; f_t$ using these annotated voxels. The class probability values $\in [0, 1]$ are obtained by normalization, i.e., $F_t(v) = f_t(v)/(f_a(v) + f_s(v) + f_t(v))$ where v or $v(x, y, z)$ denotes the intensity of the voxel at coordinates (x, y, z) . To achieve extremely high **efficiency** of voxel-level probability evaluation at run-time, we convert the fitted KDE models into look-up tables of normalized $F_a(v); F_s(v); F_t(v)$ since the CT intensity value is bounded in the range of $[0, 4095]$. Our intensity value is equal to its Hounsfield unit (HU) added with a fixed offset, $v = HU + 1024$.

2.1 Bounding Box Generation of Tagging ROIs

We first tackle the problem of spatially detecting and bounding each tagging material pool using a bounding box in a CT image.

Colonic air & bone compartments identification: Since the CTC tagging materials of interest must be adjacent to air in the abdominal region, an effective way of detecting these volume ROIs is to start from colonic air components and search around nearby regions. Air compartments can be effectively extracted using a generic 3D region growing algorithm and the air intensity range. However they may be from lung and other extra-colonic regions (e.g., small intestine, stomach) and we eliminate extra-colonic air findings, via robust component-level structure parsing using geometric features and rank-one SVM classification [4]. On the other hand, when searching tagging materials, we need to avoid the bone tissue voxels since they have similar Hounsfield units or CT intensity values. Anatomical bone landmarks (e.g., the centers of vertebrae) can be automatically and reliably detected using a recent cascade classifier [5] or random trees [6]. Most of bone structures in the thoracic and abdominal regions including ribs and vertebral columns can be identified by region growing from landmarks.

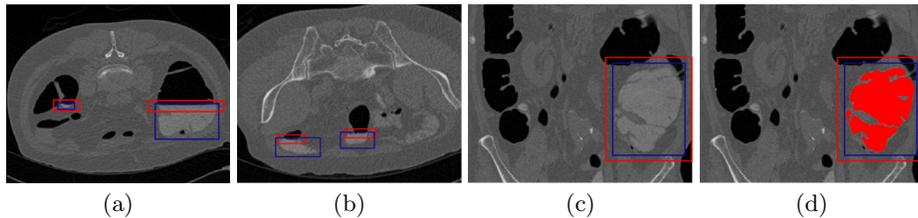


Fig. 2. Illustration of ROI detection (in **blue**) and AT region detection (in **red**) results via 3D bounding boxes. (a,b) in axial or transverse views; (c,d) in coronal views; (d) tagging voxels are annotated in **red**.

ROI generation: At each transverse CT slice, run 2D connected component (CC) labeling on the labeled colonic air compartments. From each 2D air CC, we probe along the direction towards possible tagging materials (posterior if supine otherwise anterior). If a voxel $v(x, y, z)$ being probed has not been labeled (either as bone or tagging material), its CT intensity value is used to check if it can serve as a seed point of tagging material. The above tagging voxel class probability function $F_t(v(x, y, z))$ of intensity is employed. If $F_t(v(x, y, z))$ indicates high confidence (e.g., ≥ 0.85), we employ a 3D region growing algorithm from this seed to propagate a 3D connected component of tagging materials (evaluating $F_t(\cdot)$ during propagation). For each detected 3D CC (denoted as CC_t), its corresponding 3D bounding box is computed. Then we obtain a list of 3D bounding boxes $\{CC_t\}$ as ROIs per CT volume, where each box represents a cropped sub-volume containing tagging and surrounding voxels. Fig. 2 shows the volume ROI detection results, covering tagging voxel pools with various sizes or volumes. Each ROI can be “eCleansed” in parallel (for computational efficiency) since there is no logical dependency among processing them. Using region growing to link $F_t(\cdot)$ responses is more robust than running $F_t(\cdot)$ at every voxel position, which more prone to small clusters of imaging noise. Optionally, we can reject components with small volume measurements (e.g., $< 3ml$), indicating low confidence.

2.2 Efficient ROI Verification & AT Localization

This stage contains posterior or anterior directional gradient profiling and Air-Tagging (AT) transition region detection via 3D Haar feature boosting.

Dominant air-tagging gradient search: For each ROI, we first mask out air-class voxels based on its intensity value since Hounsfield unit (HU) in CT for air is well defined. For remaining voxels at any coordinates (x, y, z) , its gradient vector $(\Delta_x, \Delta_y, \Delta_z)$ and gradient orientation (θ, ϕ) weighted by the magnitude $\gamma = \|(\Delta_x, \Delta_y, \Delta_z)\|$ are computed. In the parameter space of $[\theta, \phi]$, each voxel v_i will cast one data sample (θ_i, ϕ_i) with weight γ_i . Then the task of finding dominant air-tagging gradient direction is essentially a mode-seeking problem that can be solved effectively by mean-shift [7]. In our case, the distributions of $\{(\theta_i, \phi_i)\}$ are noisy but statistically compact. Mean-shift algorithm converges to the dominant gradient direction $(\hat{\theta}, \hat{\phi})$ in 4 ~ 6 iterations. The orientation of $(\hat{\theta}, \hat{\phi})$ coincides with Y-direction of $\geq 98.5\%$ probability.

Efficient AT detection via boosted 3D Haar features: For each volume ROI, there can be at least one or more Air-Tagging (AT) transition regions though $\cong 65\%$ ROIs have only one AT. AT regions are labeled by one bounding box per AT for 10 CTC volumes (split evenly for training and testing). Each bounding box precisely covers the AT spatial expansion in CT coronal plane from manual annotation, which is normally a narrow-band box in Y-direction including several slices or layers of air and tagging voxels. Detecting bounding boxes in run-time is approached as a generic 3D object detection problem via Haar features boosted by Adaboost or Probabilistic Boosting Tree [8]. 3D Haar features are well capable to capture intensity contrast statistics existing from air voxel layers to tagging.

Without loss of generality, we assume that $(\hat{\theta}, \hat{\phi})$ is aligned with Y-coordinates (i.e., the normal of coronal plane), to eliminate the orientation degrees-of-freedom (DOF) completely. From above gradient search, we can locate Y-locations with strong gradient magnitude sums (aggregated on each coronal plane within ROI). On each XZ plane at a hypothesized Y, we can perform a 2D connected component (CC) operation on its tagging class probability response map (by evaluating $F_t(v)$) and compute the X and Z spans of the largest tagging CC as \hat{X}, \hat{Z} . The only remaining DOF is translation in Y within ROI. In run-time, we obtain the list of final AT bounding boxes per ROI based the detection probability rank and non-maximal suppression (to remove overlapped detections). Our detector using PBT boosted Haar feature can capture the intensity contrast of air and tagging class voxels in valid AT boxes, and achieves 99.56% sensitivity with no false positives on validation set. Only 2 very small AT boxes out of 477 annotated boxes are missed in 5 training CT volumes. Examples are shown in Fig. 2 as red boxes.

2.3 Multivariate Image Model and MRF in AT

Due to partial-volume effect in the air-tagging transition regions and air-tagging-tissue T-junctions [1, 2] in AT, there are certain amounts of air-tagging mixed, or pseudo voxels having very similar (single voxel) intensity value as soft-tissue. Therefore we formulate a high-order intensity pattern to differentiate pseudo and true soft-tissue voxels through context modeling. The motivation is that soft-tissue voxels tend to have more soft-tissue neighbors while pseudo voxels having both air and tagging voxels on both sides in Y-direction (due to gravity). More precisely, for each voxel $v(x, y, z)$ at location (x, y, z) , we form a $(2N + 1)$ -dimensional intensity feature vector $V(x, y, z) = [v(x, y - N, z), \dots, v(x, y - 1, z), v(x, y, z), v(x, y + 1, z), \dots, v(x, y + N, z)]$ center at (x, y, z) where $N = 3 \sim 7$. Then $V(x, y, z)$ is classified as a supervised binary learning task. To do so, all voxels inside 36 AT boxes (Refer to Fig. 2 (d).) randomly extracted from 5 training volumes, are annotated as positive (+) soft-tissue class versus negative (-) tagging (including pseudo) class.

High-order intensity pattern modeling: Since KDE does not scale well with higher dimensionality $(2N + 1)$ in both computation and accuracy, we propose a new method Multivariate Intensity Histograms with Adaptive Semantic Binning (*MIHASEB*), via maximizing the Chernoff Information of two empirical distributions. Chernoff information [9] is defined as follows, given two discrete

distributions of $\mathcal{H}(+), \mathcal{H}(-)$.

$$\mathcal{C}(\mathcal{H}(+), \mathcal{H}(-)) = - \min_{0 \leq \lambda \leq 1} \log \left(\sum_{b=1}^B \mathcal{H}^\lambda(+, b) \mathcal{H}^{1-\lambda}(-, b) \right) \quad (1)$$

When $\lambda = \frac{1}{2}$, \mathcal{C} becomes Bhattacharyya coefficient [9] which is optimized here. Notations of histograms $\mathcal{H}(V^+)$ and $\mathcal{H}(+)$ are interchangeable (same for $\mathcal{H}(V^-), \mathcal{H}(-)$). Histogram takes an input feature vector V to return the output value, as recoded in the corresponding bin. Histogram binning is the process on how to construct histogram by finding its associated bins while maximizing the (+/-) class separability criterion of $\mathcal{C}(\mathcal{H}(+), \mathcal{H}(-))$. Maximizing $\mathcal{C}(\mathcal{H}(+), \mathcal{H}(-))$ brings the optimized classification accuracy [10, 11]. Therefore our learned models are discriminative-generative density functions. If the exhaustive histogram binning parameter search is adopted, even only 100 evenly divided intensity boundaries (CT intensity $\in [0, 4095]$) to be hypothesized and evaluated by Eq. 1, the computation complexity is already prohibitively expensive as 100^{2N+1} . Again, we draw importance samplings by leveraging the knowledge of CT intensity ranges of semantic air, soft-tissue, (weak) tagging and bone anatomies. The sampling parameters are data-calibrated similarly using normalized KDE plots as Fig. 1 (**Right**): means $\{\mu_m\} = \{350, 750, 1100, 1400\}$ (w.r.t. air, soft-tissue, weak tagging and regular tagging voxel intensity); standard deviations $\sigma_m = \sigma_m^+ = \sigma_m^- = 40$ regardless of $m = 1, 2, 3, 4$ for simplicity. Our multivariate *MIHASB* algorithm is described in Alg. 1. Semantic Sampling Function (*SSF*) is equivalent to draw a random sample s_m from $[\mu_m - \sigma_m^-, \mu_m + \sigma_m^+]$ based on uniform sampling, or from Gaussian Distribution $\mathcal{G}(\mu_m, \sigma_m)$, $\{s\} = \{s\} \cup s_m$. In this way, *SSF* implicitly encodes the intensity patterns such as how many air, soft-tissue, bone, tagging voxels existing in the relative spatial context and makes *MIHASB* binning process very efficient in $K = 2N + 1$ dimensions. The sampling derivation σ captures the CT data variations at semantic class boundaries.

In runtime, the histograms $\mathcal{H}(V^+), \mathcal{H}(V^-)$ are normalized using likelihood ratio testing [10] per bin, to form $\mathcal{P}(+)$ (i.e., the soft-tissue class \mathcal{P}_s).

$$\mathcal{P}_s(V) = \frac{\mathcal{H}(V^+)}{\mathcal{H}(V^+) + \mathcal{H}(V^-)} \quad (2)$$

Last, \mathcal{P}_s is also represented as a multi-dimensional look-up table with the estimated binning boundaries where each cell stores $\mathcal{P}_s \in [0, 1]$. Our trained histogram models $\mathcal{H}(V^+), \mathcal{H}(V^-), \mathcal{P}(+)$ have 2592 bins (when $N = 4$). Longer range context of $N = 4$ improves over $N = 3$, but performance saturates after $N > 4$.

MRF optimization: In each AT bounding box, we first evaluate every inside voxel $v(x, y, z)$ via its feature vector $V(x, y, z)$ to obtain $\mathcal{P}_s(x, y, z)$ at the location (x, y, z) . Each AT voxel is treated as a node of the 3D MRF grid and we use its negative-log form $E_{data}(x, y, z) = -\log(\mathcal{P}_s(x, y, z))$ [12] as the data term. The intensity-contrasted pairwise potential function is adopted as a spatially isotropic smoothness term (26-neighborhood is used). This energy minimization problem can be solved effectively by Max-Flow/Min-Cut algorithm [12]. The optimization output is a binary label field: $L = 1$ meaning the positive soft-tissue class (to be preserved in eCleansing); $L = 0$ indicating negative class of air, tagging and pseudo-enhanced (non soft-tissue) voxels that should be eliminated.

Algorithm 1: Adaptive Semantic Binning in Joint Multivariate Space

Input: Sets of positive samples $\{V^+\}$ and negatives $\{V^-\}$; K is the dimension of the feature vector $V \in \mathfrak{R}^{K=9}$

Output: Trained multivariate histograms $\mathcal{H}(V^+), \mathcal{H}(V^-)$

Initialize $\mathcal{H}^0(V^+), \mathcal{H}^0(V^-)$ with one bin and $\mathcal{C}_{max}^0 = 0$;

for $t = 1, \dots, T$ **do**

$\mathcal{C}_{max}^t = 0$;

for $k = 1, \dots, K$ **do**

Generate the hypothesized split set $\{s^k\}$ from semantic sampling function at Dimension k ;

for *Each* $ss \in \{s^k\}$ **do**

Update intermediate histograms $\mathcal{H}^t(V^+), \mathcal{H}^t(V^-)$ by adding the split ss into $\mathcal{H}^{t-1}(V^+)$ and $\mathcal{H}^{t-1}(V^-)$ via Cartesian data repartitioning in \mathfrak{R}^K , respectively;

Evaluate the Chernoff Information \mathcal{C} (i.e., eq. 1) given $\mathcal{H}^t(V^+), \mathcal{H}^t(V^-)$ from $\{V^+\}$ and $\{V^-\}$;

if $\mathcal{C}_{max}^t < \mathcal{C}$ **then**

$\mathcal{C}_{max}^t = \mathcal{C}$; and record the current optimal split ss ;

end

end

end

if $(\mathcal{C}_{max}^t - \mathcal{C}_{max}^{t-1}) > \epsilon$ **then**

Update and finalize $\mathcal{H}^t(V^+), \mathcal{H}^t(V^-)$ for iteration y , by adding the recorded optimal split ss (according to \mathcal{C}_{max}^t) into $\mathcal{H}^{t-1}(V^+), \mathcal{H}^{t-1}(V^-)$ via data repartitioning;

else

Return $\mathcal{H}^t(V^+), \mathcal{H}^t(V^-)$ as the final output (converged) and exit;

end

end

2.4 Probabilistic Voxel Labeling and Smoothing in RT

Recall that in Sec. 2.1, we obtain a 3D connected component of tagging materials CC_t (guided by $F_t()$) to bound each ROI. Possible tiny holes $\in CC_t$ (e.g., created by air bubbles in the inhomogeneous tagging pool) can be filled by standard hole-filling algorithm. For each remaining voxel $v \in RT = CC_t \cap (\notin AT)$, $F_s(v) = f_s(v)/(f_a(v) + f_s(v) + f_t(v))$ is used to assign the normalized soft-tissue class probability $\in [0, 1]$. The $F_s(v)$ -weighted intensity value is computed as $v_s(x, y, z) = v(x, y, z) \times F_s(x, y, z)$ for v . Finally, we apply a simple 3D isotropic Gaussian smoothing filter G (variance = $1mm$) on $\{v_s(x, y, z)\}$ to obtain denoised $\{\mathcal{G}(v_s(x, y, z))\}$ and set label $L = 0$ for voxels satisfying $\mathcal{G}(v_s) < 350$. The threshold 350 is defined by our clinical collaborator. To prepare **CTC visualization**, all voxels with label $L = 0$ are set as 0 intensity value, thus “electronically cleansed”. Voxels in the ROI but outside of CC_t remain unprocessed or “untouched”.

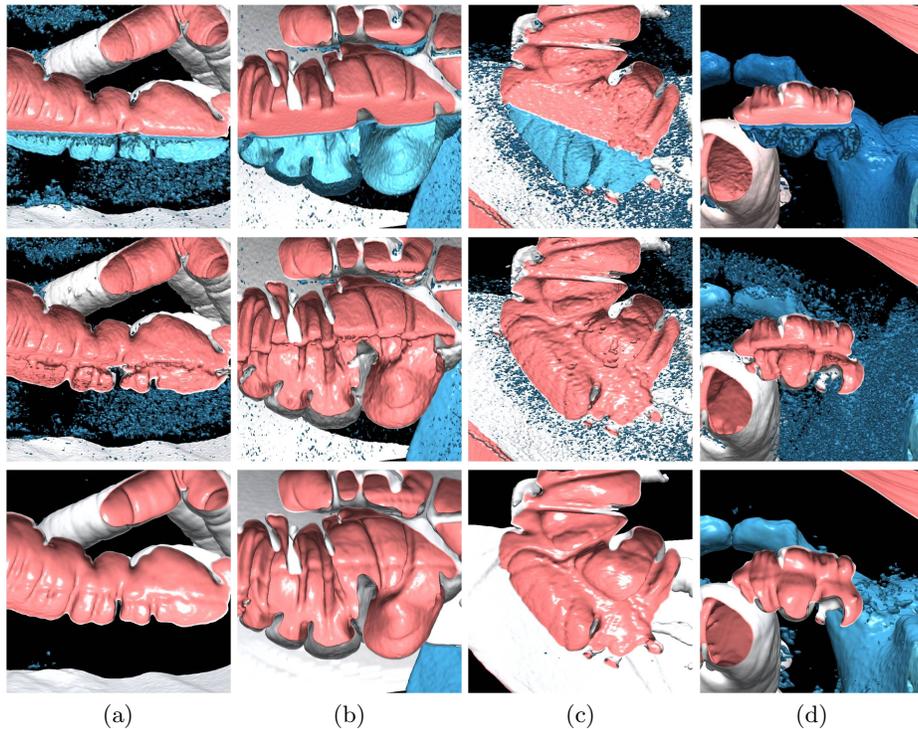


Fig. 3. Examples of E-cleansing results: original volume rendering (**Top**), generated by an implementation of GMM-EM method [1] (**Middle**) and our method (**Bottom**).

3 Experimental Results & Discussion

Data: Twenty-six weak faecal-tagging CTC scans (15 patients) from three hospitals in Europe are used. Variations of bowel preparations with different combinations of doses were employed. The axial image slice numbers vary from 728 to 1089 and the slice thickness ranges from 0.798 mm to 1.25 mm.

Accuracy: Tagging voxels of 10 out of 26 scans are annotated using interactive graph-cut [12] and manual touch-up. 10 volumes are split evenly for training and testing. Two performance criteria are used: *Sensitivity* (\mathcal{S}) on the detected and removed tagging voxels versus all labeled tagging voxels; and *Accuracy* (\mathcal{A}) on the number of removed tagging voxels versus all removed voxels. For comparison, we implement the Gaussian Mixture Model (*GMM*) based eCleansing method [1]. Our method has $\mathcal{S} = 99.2\%$ and $\mathcal{A} = 99.3\%$ while $\mathcal{S} = 93.6\%$ and $\mathcal{A} = 98.9\%$ are for our version of [1]. *The main difference is on sensitivities.* Although $\mathcal{S} = 93.6\%$ is not low, a lot of tagging voxels at three-material (i.e., air, tagging and soft-tissue) transitional areas in AT are not effectively cleansed. There are severe artifacts “left-overs” floating in the air or adjacent to colon wall which imposes significant technical challenges for clinicians to perform 3D CTC reading. The joint use of KDE and *MIHASB* shows better intensity modeling accuracy among different

voxel classes than *GMM*. Without explicit physics based three-material transition modeling [2], our data-driven approach achieves highly competitive results in less computational cost (refer “bath-tub ring” effect in Fig. 9 of [2]). We also conduct CTC 3D fly-through user study to subjectively evaluate the reader’s impact on all 26 volumes, prepared by our proposed method and [1], respectively. The visualized comparison is demonstrated in Fig. 3.

Speed: *MIHASB* algorithm is compared with *GMM* [1] and random trees (RF) via random intensity testing [6]. RF-11 and RF-15 (i.e., 11 or 15 decision trees) provide comparable or slightly better accuracy than *MIHASB* (\mathcal{S}/\mathcal{A} : 87.5%/95.1% and 87.9%/95.4% versus 87.7%/95.5% in testing). Random trees need hundreds of arithmetic operations per voxel evaluation; *MIHASB* only requires 9-13 times of integer comparison. After MRF optimization, \mathcal{S}/\mathcal{A} are improved to 96.2%/98.6%. Segmenting the voxels into mixtures [1] runs as the slowest.

In summary, our main contributions are three-fold. 1), We propose a new volume ROI based algorithm pipeline in a divide-and-conquer manner, which is substantially novel from previous work [1,2]. 2), We present statistical data-driven generative and discriminative-generative models (e.g., *MIHASB*) throughout the paper that effectively cope well with the emerging minimal bowel preparation protocols. 3), Our method has been validated under various quantitative criteria and user study, using 26 CTC studies from multiple clinical sites.

References

1. Wang, Z., Liang, Z., et al.: An Improved Electronic Colon Cleansing Method for Detection of Colonic Polyps by Virtual Colonoscopy. *IEEE Trans. Biomed. Eng.* (2006) 53:1635–46
2. Serlie, I., Vos, F., Truyen, R., Post, F., Stoker, J., van Vliet, L.: Electronic Cleansing for Computed Tomography (CT) Colonography Using a Scale-Invariant Three-Material Model. *IEEE Trans. Biomed. Eng.* (2010) 57:1306–17
3. Liedenbaum, M., Denters, M., Zijta, F., van Ravesteijn, V., Bipat, S., Vos, F., Dekker, E., Stoker, J.: Reducing the oral contrast dose in CT colonography: evaluation of faecal tagging quality and patient acceptance. *Clin. Radiol.* (2010) 30–37
4. Lu, L., Wolf, M., Liang, J., et al.: A Two-level Approach Towards Semantic Colon Segmentation: Removing Extra-colonic Findings. *MICCAI* (2009) 1:1009–1016
5. Wu, T., Jian, B., Zhou, X.: Automated identification of thoracolumbar vertebrae using orthogonal matching pursuit. *MLMI* (2011) 126–133
6. Lepetit, V., Fua, P.: Keypoint recognition using randomized trees. *IEEE Trans. Pat. Anal. Mach. Intell.* (2006) 28:1465–79
7. Comaniciu, D., Meer, P.: Mean Shift: A Robust Approach Toward Feature Space Analysis. *IEEE Trans. Pat. Anal. Mach. Intell.* (2002) 24(5):603–619
8. Tu, Z.: Probabilistic boosting-tree: Learning discriminative models for classification, recognition, and clustering. In: *ICCV*. (2005) 1589–1596
9. Cover, T.M., Thomas, J.A.: *Elements of information theory*. (1991)
10. Konishi, S., Yuille, A., Coughlan, J., Zhu, S.: Statistical edge detection: Learning and evaluating edge cues. *IEEE Trans. Pat. Anal. Mach. Intell.* (2003) 25:57–74
11. Lu, L., et al.: Hierarchical learning for tubular structure parsing in medical imaging: A study on coronary arteries using 3D CT Angiography. *ICCV* (2009)
12. Boykov, Y., Funka-Lea, G.: Graph Cuts and Efficient N-D Image Segmentation. *Int. J. Comp. Vis.* (2006) 70(2):109–131