Virtual tagging for laxative-free CT colonography: Pilot evaluation

Janne Näppi^{a)} and Hiroyuki Yoshida

Massachusetts General Hospital and Harvard Medical School, 25 New Chardon Street, Suite 400C,
Boston, Massachusetts 02114

(Received 30 September 2008; revised 13 March 2009; accepted for publication 16 March 2009;
published 22 April 2009)

Laxative-free computed tomographic colonography (*lfCTC*) could significantly improve patient adherence to colorectal screening. However, the interpretation of *lfCTC* data is complicated by the presence of poorly tagged feces and partial-volume artifacts that imitate colorectal lesions. The authors developed a method for virtual tagging of such artifacts. A probabilistic model of colonic wall was developed, and virtual tagging was performed on artifacts that were identified by the model. The method was evaluated with 46 clinical *lfCTC* cases that were prepared with dietary fecal tagging only. Visual examples show that the method can label partial-volume artifacts, poorly tagged feces, nonadhering completely untagged feces, and artifacts such as rectal tubes. The effect of virtual tagging was evaluated by comparing the detection accuracy of a fully automated polyp detection scheme without and with the method. With virtual tagging, the per-lesion detection sensitivity was 100% for lesions ≥ 10 mm ($n=4$) with 3.8 false positives per patient (per two CT scan volumes) and 90% for lesions ≥ 6 mm ($n=10$) with 5.4 false positives per patient on average. The improvement in detection performance by virtual tagging was statistically significant ($p=0.03$; JAFROC and JAFROC-1). © 2009 American Association of Physicists in Medicine.
[DOI: [10.1118/1.3113893](https://doi.org/10.1118/1.3113893)]

Key words: virtual colonoscopy, bowel preparation, partial-volume effect, fecal tagging, computer-aided detection

I. INTRODUCTION

Although colorectal cancer is the second leading cause of cancer deaths in the United States, it would be largely preventable if its precursor colorectal lesions were removed.¹ However, less than 40% of age-eligible U.S. adults participate in full colorectal examinations.² Patient surveys indicate that the rigorous laxative bowel cleansing that is currently required by full colorectal screening is the single most important reason for patients to avoid colorectal examinations.³ Laxative cleansing interferes with patients' daily activities by causing diarrhea, and there are potential side effects such as nausea, abdominal discomfort, vomiting, and/or flatulence.⁴

Studies have indicated that *computed tomographic colonography* (CTC) could be used to implement an effective laxative-free colorectal screening scheme.^{4–6} Laxative bowel cleansing and diarrhea may be avoided in CTC if residual bowel contents are opacified by use of a tagging agent that does not induce cathartic effect.⁶ However, the visualization and image processing of such *laxative-free CTC* (*lfCTC*) data are complicated by the presence of poorly tagged feces and partial-volume artifacts (Fig. 1). These tend to imitate the appearance of colorectal lesions, and therefore conventional image processing methods that have been developed for laxative fluid-tagging CTC tend to display numerous artifacts and *false-positive* (FP) detections in *lfCTC*.^{7,8} As can be seen from Fig. 1, artifacts in *lfCTC* data can imitate not only the shape but also the radiodensity of soft-tissue structures. Therefore, attempts to segment tagged regions automatically by a high CT number threshold value as in fluid-tagging CTC tend to misinterpret much of the

tagged materials as soft tissue in *lfCTC*. On the other hand, attempts to segment tagged regions more precisely with a low CT number threshold can segment parts of pseudo-enhanced native soft tissue, including polyps, as tagged region.^{9,10}

In this study, we hypothesized that the identification of poorly tagged residual materials and partial-volume artifacts could be improved in *lfCTC* by virtual tagging of such artifacts. We note that the observed radiodensity of tagged materials tends to be diluted by the *partial-volume effect* (PVE) that occurs within the material interface between lumen air and nonair materials, and that such diluted tagged materials appear similar to soft tissue [Figs. 1(b) and 1(c)]. However, we also observe that soft-tissue lesions that might appear at the same location as those diluted tagged materials would have even lower observed radiodensity. Therefore, by comparison of CT numbers and their gradient within the PVE region, we can differentiate poorly tagged materials from actual soft tissue. To explore this hypothesis, we developed a probabilistic colon surface model for identifying native colon surface regions. Poorly tagged regions that are not identified by the model as part of colon surface are tagged virtually in the CTC data to complement physical tagging. The virtual tagging method is described in Sec. II B.

The study design is shown in Fig. 2. To provide conservative and unbiased results, we used independent development and evaluation data sets. Because of the relatively small number of available *lfCTC* cases, we assigned all available *lfCTC* cases as the evaluation data set. The development data set was an independent set of CTC cases that

FIG. 1. Examples of artifacts that imitate polyps in laxative-free CTC data. (a) Untagged feces (arrow) can imitate large and small lesions. (b) Partial-volume effect lowers CT numbers of tagged materials which can make the outcome look similar to small polyps (arrow). (c) Partial-volume effects can imitate small polyps. Axial view (larger image) indicates a polyp (white arrow), but coronal view (smaller image) indicates that the lesion is a partial-volume artifact (arrow). (d) In 3D view, multiple solid feces that imitate polypoid shapes distract readers from focusing on true polyps.

was prepared with reduced volume of laxatives and barium tagging. The study materials are described in Sec. II A.

The evaluation methods are described in Sec. II C. In addition to visual examples, we evaluated the effect of virtual tagging on the polyp detection accuracy of a fully automated detection scheme that was trained with independent CTC data. The results are described in Sec. III. The discussion and conclusions are described in Secs. IV and V, respectively.

II. MATERIALS AND METHODS

II.A. Materials

The study was approved by the institutional review boards. Because the number of available *l*/CTC cases was small, we used reduced-laxative CTC cases as a development data set. The reduced-laxative cases present a variety of PVE conditions that are similar to those of *l*/CTC cases. All available true *l*/CTC cases were used as the evaluation data set.

The development data set included 82 clinical CTC cases (164 supine and prone CT scans). The patients were recruited randomly as part of daily colorectal screening practice. The bowel preparation was performed with a reduced volume of laxatives (16 mg of magnesium citrate and four tablets of bisacodyl) and a reduced volume of tagging (50 ml of barium).¹¹ For the development set, the CT acquisition was performed in supine and prone positions by use of four types

FIG. 3. Examples of the laxative-free evaluation cases. [(a) and (b)] Examples of barium tagging. [(c) and (d)] Examples of iodine tagging. (e) A 12 mm adenoma (barium tagging). (f) Another 12 mm adenoma (iodine tagging). [(g) and (h)] A 6 mm adenoma (barium tagging) seen in axial and 3D views, respectively.

of CT scanners with 1.25–5.0 mm collimations, 0.6–3.0 mm reconstruction intervals, 29–100 mA currents, and 120–140 kVp voltages.

The evaluation data set included 46 *l*/CTC cases (92 supine and prone CT scans). The patients were recruited randomly at two hospitals as part of daily colorectal screening practice. There were two groups as described below. No bowel cleansing was performed prior to the CTC. Instead, 30 patients were prepared with dietary tagging by barium, and 16 patients were prepared with dietary tagging by nonionic iodine. Nonionic iodine is less hypertonic and has higher safety profile than the ionic iodine used previously in CTC.¹² Figures 3(a)–3(d) show examples of the evaluation data set. All available *l*/CTC cases were used regardless of their diagnostic quality.

In the barium tagging group, 30 patients [10 males, 20 females; age range: 38–82 years (mean: 58 years)] were advised to drink apple-flavored barium suspension (with concentrations of 2.1%, 4%, and 40% weight/volume⁴) with meals for 1 or 2 days prior to the CTC.⁴ 1 day before the CTC, the patients followed a dedicated low-residue diet.⁴ For this barium evaluation group, the CT acquisition was performed in supine and prone positions (60 CT scans) by use of two types of CT scanners (LightSpeed Ultra, GE Medical Systems, Milwaukee, WI; Sensation 64, Siemens Medical Solutions, Malvern, PA) with 1.0–2.5 mm collimations, 0.7–1.3 mm reconstruction intervals, 28–84 mA currents, and 120–140 kVp voltages.

In the iodine tagging group, 16 patients [11 males, 5 females; age range: 52–67 years (mean: 58 years)] were advised to drink 10 ml of iohexol diluted in ≥ 150 ml of beverages with meals for 2 days prior to the CTC. The patients were advised to follow a low-residue diet based on a list of food to eat and to avoid. For this iodine evaluation group, the CT acquisition was performed in supine and prone positions (32 CT scans) by use of three types of CT scanners (LightSpeed Plus, LightSpeed Ultra and LightSpeed 16, GE Medi-

FIG. 2. Overview of the study design.

FIG. 4. (a) Illustration of the partial-volume effect region (between arrows) at the colon surface. The white solid line represents the true material surface. The dotted white lines indicate the extent of the partial-volume effect. (b) A region of interest. (c) High values of the gradient magnitude of CT numbers indicate partial-volume region in the region of interest. (d) Lesions that are not connected directly to bowel wall can be excluded immediately as untagged feces (arrow). (e) A pseudolesion (stool) that is connected to the bowel wall through an artificial soft-tissue interface (arrows) due to PVE. (f) The first step of the virtual tagging method (light grey color) indicates the artificial soft-tissue interfaces.

cal Systems, Milwaukee, WI) with 2.5 mm collimation, 1.25–2.50 mm reconstruction intervals, 70–110 mA currents, and 120–140 kVp voltages.

Conventional optical colonoscopy was performed within 1 week after the CTC. The colonoscopy findings were correlated with the CTC data. There were four biopsy-proven adenomas/carcinomas ≥ 10 mm in three patients, and six biopsy-proven adenomas 6–9 mm in three other patients. These included two pedunculated and two sessile or flat lesions ≥ 10 mm and three pedunculated and three sessile or flat lesions 6–9 mm. All lesions but one touched or were covered by tagged materials. The CTC cases that contained proven lesions had been scanned by three of the four CT scanner models of this study. Figures 3(e)–3(h) show three examples of the lesions.

II.B. Virtual tagging

II.B.1. Probabilistic surface model

Computed tomography (CT) represents a target volume as a discrete three-dimensional (3D) voxel grid. Each voxel has a CT number that represents x-ray attenuation at the relative spatial location of the voxel in the target volume. The CT number can be modeled as

$$v = \sum_{i=1}^n v_i r_i, \quad (1)$$

where n is the number of materials that share the voxel, v_i is the radiodensity of target material i , and r_i is the partial-volume fraction of the material i at the spatial location of the voxel, with $\sum_{i=1}^n r_i = 1$.¹³ The PVE occurs when different materials that have different radiodensities share the physical space of the same voxel [Fig. 4(a)]. In such regions, the CT

number v is a weighted average of the different radiodensities covered by the voxel.

Application of positive-contrast tagging can cause local pseudoenhancement of adjacent materials.¹⁰ Pseudoenhancement complicates the identification of tagged materials based on CT numbers. For example, a CT number of 200 HU (Hounsfield unit) can indicate either tagged material or pseudoenhanced soft-tissue structure such as a polyp. For this study, the CTC cases were corrected for the pseudoenhancement effect by use of an image-based adaptive density correction method that we developed previously.¹⁰ Therefore, we assume that Eq. (1) is a faithful model of CT numbers regardless of the use of tagging agent.

The PVE can dilute CT numbers significantly at material interfaces that involve air. For example, suppose that a voxel is shared by 20% of air (with a radiodensity of -1000 HUs) and by 80% of tagged material with a radiodensity of 400 HU. According to Eq. (1), the voxel would have an observed CT number of 120 HU that is very close to the radiodensity of soft tissue. However, suppose that the voxel was shared by 20% of air and by 80% of soft tissue that had a 50 HU radiodensity. In this case, the observed CT number would be -160 HU. Thus, although the PVE can lower CT numbers of tagged materials significantly, the CT numbers of soft-tissue materials would be lowered more at the same location.

We hypothesized that the PVE region of the colon surface can be modeled by use of two features: The CT number and the magnitude of the gradient of CT numbers. The gradient magnitude is a particularly effective indicator of the PVE region: Low values indicate homogeneous materials, whereas high values indicate PVE between different materials [Figs. 4(b) and 4(c)].

The colon surface model was implemented as a Bayesian scheme. Let C represent a binary category variable with two outcomes: S (surface) and N (nonsurface). Let V and G denote the statistical feature variables of the observed CT number and its gradient magnitude, respectively. According to Bayes' theorem,¹⁴ we can write

$$P(C|V, G) = \frac{P(C)P(V, G|C)}{P(V, G)}. \quad (2)$$

The denominator can be omitted because it does not depend on C . Therefore, Eq. (2) can be simplified to

$$P(C|V, G) \propto P(C)P(V, G|C). \quad (3)$$

We estimated the prior and conditional probabilities of Eq. (3) by use of the 82 reduced-laxative development CTC cases (Sec. II A) as follows. The features were sampled within a thick region encompassing the colonic surface. Methods for automated extraction of such a region have been described previously.^{15,16} Although different extraction algorithms might yield different extracted thick regions around the colonic surface, this should have marginal effect on the sampling step because the samples are collected at PVE regions, i.e., voxels that are outside the partial-volume colon surface, for example, within lumen air or within soft tissue, are excluded based upon their low gradient value.

For sampling prior probabilities, it is necessary to determine tagged noncolonic surfaces and native colonic surfaces in the CTC data. To identify tagged nonsurface regions, we thresholded the CT data at a CT number value of 150 HU, expanded the thresholded region with three layers of voxels, and intersected the resulting binary region with the extracted thick surface region. The other part of the thick surface region was considered to represent native colonic surface. The 3D gradient magnitudes were calculated from CT numbers by use of anisotropic $3 \times 3 \times 3$ Sobel kernels.¹⁷

To determine if a voxel represents native colonic surface, we note that the category $L \in \{S, N\}$ that best explains an observed CT number v and gradient magnitude g at a voxel can be derived from Eq. (3) by use of the maximum *a posteriori* estimate¹⁸

$$\hat{C} = \operatorname{argmax}_L P(C=L)P(V=v, G=g|C=L). \quad (4)$$

Therefore, a voxel can be considered to represent the colonic surface if

$$P(V=v, G=g|C=S) \geq P(V=v, G=g|C=N). \quad (5)$$

Combinations of CT numbers v and gradient magnitudes g that did not appear among the sampled combinations were considered to represent tagged or artifactual materials ($C=N$). Because the development data set provided a large number of samples of untagged soft-tissue surfaces and because soft-tissue compositions are not expected to vary drastically between cases, it is more likely that new unseen (v, g) combinations would represent new cases of tagging densities, tagging artifacts, or metallic objects rather than unseen cases of untagged soft-tissue surfaces. Examples of the resulting distributions will be shown in Sec. III A.

II.B.2. Application of virtual tagging

The application of the virtual tagging method consists of two steps. First, we calculate the CT number v and the gradient magnitude g at each voxel. The Bayesian scheme is used to determine the voxels that do not represent native colonic surface. Such voxels are tagged virtually by mapping their CT numbers according to

$$v' = \max\{v, T_v\}, \quad (6)$$

where v is the observed CT number and $T_v=300$ HU is a predefined smallest CT number of a virtually tagged voxel.

Although completely untagged feces can be indistinguishable from colorectal lesions, they can be identified as pseudolesions when they do not adhere directly to the colonic wall [Fig. 4(d)]. However, automated identification of this condition is nontrivial, because tagged materials have an artificial soft-tissue PVE interface that provides an artificial soft-tissue connection to the colonic wall [Fig. 4(e)]. Nevertheless, we note that the first step of the virtual tagging method identifies such PVE interfaces as tagged material [Fig. 4(f)]. Therefore, in the second step of the method, we perform virtual tagging also on regions that, after the first step, do not adhere directly to the colonic wall through soft-tissue-like CT numbers. To determine such regions, we

threshold voxels that have CT numbers of <-800 HU (air voxels) or >150 HU (physically or virtually tagged voxels) and determine the largest connected component in the thresholded region.^{17,19} This largest component represents the abdominal soft-tissue region, and therefore it includes the complete region of colonic surface. All other connected components, such as the untagged stool in Fig. 4(f), are considered as residual materials that can be virtually tagged.

II.C. Evaluation methods

The effect of virtual tagging was evaluated by use of the 46 *lf*CTC evaluation cases (Sec. II A). Visual examples were provided by reviewing of the original and virtually tagged CTC data in two-dimensional (2D) views. The effect of virtual tagging was assessed by evaluation of the detection accuracy of a fully automated polyp detection scheme, where polyps are detected by an analysis of CT numbers within a thick volumetric region encompassing the colonic surface.^{8,20} Before the detection step, the scheme performs a fast pseudocleansing of tagged materials from CTC data by mapping of CT numbers >100 HU toward the level of air.⁸ Therefore, a separate electronic cleansing step for subtracting tagged materials is not needed. The scheme was trained to detect colorectal lesions with laxative CTC cases that were completely independent from those of this study.²¹

A detection was considered as a *true positive* (TP) if

$$\sqrt{(x-X)^2 + (y-Y)^2 + (z-Z)^2} \leq D/2, \quad (7)$$

where D is the size of a colonoscopy-confirmed lesion located at (X, Y, Z) , and (x, y, z) is the center of a computer-extracted region²² of the detection. All other detections were considered as FP detections. Here, it should be noted that all distance-based units, including the parameters of the detection scheme, are normalized to resolution-independent millimeter units based on the DICOM header information of each input CT scan volume.

Because an automated detection scheme that displays a large number of FP detections is not likely to be useful in routine clinical practice,^{23,24} we limited the evaluation to the range of 0–6 FP detections per patient (per two CT scan volumes) on average. Any lesions that would have been detected at a higher FP rate were considered as false negatives.

The detection accuracy was characterized by use of the location-specific *free-response receiver operating characteristic* (FROC) analysis that yields the detection sensitivity as a function of the number of FP detections per patient. The FROC curves were generated by interpreting the polyp-likelihood values that the automated detection scheme calculates for each detection as a decision variable.²⁵ To estimate detection accuracy for lesions ≥ 10 and 6–9 mm, we determined the polyp-likelihood values that yield maximum sensitivity for these types of lesions, where the TP detections were identified according to Eq. (7). These polyp-likelihood values were then used as the polyp-likelihood thresholds for calculating detection accuracy, where the average numbers of FP detections per patient were calculated by dividing the number of resulting FP detections at these polyp-likelihood

FIG. 5. Distribution of sampled features among (a) untagged and (b) tagged voxels.

threshold levels by the number of patients. All FP detections were counted as different FP detections, whereas multiple TP detections of the same true lesion (if this occurred) were counted as one TP detection. The same detection parameter values were used for all experiments.

To estimate statistical significance, paired Wilcoxon signed-rank test was used to evaluate the FP reduction in automated detection without and with virtual tagging. To evaluate the improvement in detection performance, we used the location-specific *jackknife* FROC (JAFROC) and JAFROC-1 methods.^{26–28} These methods estimate whether the difference between two alternative FROC (AFROC) curves is statistically significant. Jackknifing is used to delete each case, one at a time, and the area under the curve is recalculated and characterized by a figure-of-merit (FOM) value in $[0,1]$. The FOM rewards an observer for good decisions (true positives and true negatives) and penalizes for bad decisions (false negatives and false positives).²⁸ In JAFROC, the FOM is the nonparametric estimate of the area under AFROC curve except that only normal cases are used for calculating FP fraction, whereas in JAFROC-1, the highest-rated FP detections on abnormal cases are also included in the FOM calculation. The pseudovalues are analyzed for statistical significance by an analysis of variance.²⁹

Finally, because of the wide use of conventional ROC analysis, we also evaluated the result in terms of the patient-based ROC method where each case was represented by the highest polyp-likelihood of detections. It should be noted that this ROC method is not really applicable to our experiment, in which the purpose was to evaluate the accuracy of detecting specific lesions within observed data rather than to determine whether a patient is normal or abnormal.²⁸ It should also be noted that, because the location-specific data are neglected in the ROC method, the application of ROC introduces substantial statistical power penalty over location-specific methods,²⁶ thereby resulting in a higher p value than with JAFROC or JAFROC-1 methods.

We also analyzed the sources of FP detections before and after the application of virtual tagging. The automated detection scheme was used to analyze the virtually tagged and not virtually tagged CTC data, and the reported detections were reviewed by use of a graphical CTC workstation with standard 2D and 3D views.

FIG. 6. Visual examples of virtual tagging. In each example, the images on the left are original views, those on the right are virtually tagged views, and those at the top are shown in a lung CT and those at the bottom in a soft-tissue CT display window. (a) Large chunk of tagged feces. (b) Multiple pieces of tagged feces. (c) Poorly tagged feces/stool. (d) Untagged feces (arrow). (e) Rectal tube. (f) A 6 mm polyp (arrow).

III. RESULTS

III.A. Probabilistic model

Figure 5 illustrates the sampled distributions of the CT numbers and gradient magnitudes. Here, white color indicates the combinations of CT number and gradient magnitude values that occurred in highest frequency, and black color indicates the combinations that occurred with lowest frequency.

The untagged samples yield a bimodal distribution where most CT numbers represent either air (-1000 HU) or soft tissue (0 HU). Such feature combinations tend to appear at the ends of the PVE material transitions of colonic wall in CTC data. Because of the limited image contrast between soft tissue and air, most gradient magnitude values within the PVE transition are lower than 500 HU.

TABLE I. Detection accuracy of automated per-lesion detection without and with virtual tagging. Abbreviations: VT=virtual tagging; FPs=false positives per patient (per two CT scan volumes) on average.

	≥ 10 mm sensitivity	FPs	6–9 mm sensitivity	FPs	≥ 6 mm sensitivity	FPs
Without VT	100% (4/4)	4.8	50% (3/6)	4.8	70% (7/10)	4.8
With VT	100% (4/4)	3.8	83% (5/6)	5.4	90% (9/10)	5.4

The CT numbers of tagged samples are concentrated on values >0 HU. Most gradient magnitude values within the PVE transition are higher than 500 HU because of the high image contrast between tagging and air.

III.B. Visual examples

Visual review of the cases indicated that virtual tagging complements physical tagging by labeling poorly tagged materials and partial-volume interfaces of tagged materials in CTC data. Also, completely untagged feces are tagged virtually when not adhering to the colonic wall.

Figure 6 shows 2D visualizations of the outcome of virtual tagging. In each example, the combination of physical tagging and virtual tagging provides the most complete and precise delineation of residual materials. The images on the left represent original views, the images on the right represent virtually tagged views, the images at the top are shown in a lung CT display window (-200 ± 800 HU), and the images at the bottom are shown in a soft-tissue CT display window (0 ± 300 HU).

Figures 6(a) and 6(b) show how virtual tagging improves the delineation of the partial-volume surface of tagging over original views. Figure 6(c) shows how virtual tagging enhances poorly tagged stool: The stool looks similar to a soft-tissue lesion with subtle central enhancement reminiscent of an adenoma or cancer, but it has been virtually tagged be-

cause the combination of CT number and gradient magnitude values is different from that of soft tissue. Figure 6(d) demonstrates the identification of untagged feces that do not adhere to colonic wall. Figure 6(e) demonstrates implicit identification of a rectal tube by the method: The rectal tube is labeled by virtual tagging, because its surface characteristics are different from those of soft tissue. Figure 6(f) shows a polyp: Virtual tagging enhances tagged regions outside the polyp without enhancing the visually perceived region of the polyp.

III.C. Effect on automated detection

III.C.1. Detection accuracy

Table I shows the per-lesion detection accuracy of automated detection without and with virtual tagging for lesions ≥ 10 mm ($n=4$), 6–9 mm ($n=6$), and ≥ 6 mm ($n=10$). The corresponding numerical threshold values of the decision variable are 0.025491, 0.01762, and 0.01762, respectively. The per-patient detection sensitivity was 100% with virtual tagging in all these categories.

Figure 7 shows the FROC curves for the detection of lesions ≥ 6 mm, in which approximate error bars have been provided for visual summary. With virtual tagging, the detection sensitivity was 90% with 5.4 FPs per patient on average. The FROC curve has a discrete rather than smooth appearance because the curve has been constructed by connecting a small number of observed operating points corresponding to the available lesions. The actual assessment of detection accuracy, which was performed by use of the JAFROC and JAFROC-1 methods, indicated that the application of virtual tagging yielded statistically significant improvement for automated detection (Table II). As expected, the per-patient ROC analysis did not indicate significant improvement.

For lesions ≥ 10 mm, the detection sensitivity was 100% regardless of the use of virtual tagging, but the use of virtual tagging yielded statistically significant reduction in the number of FP detections ($p < 0.00001$). At a 70% detection sen-

Detection performance for lesions 6 mm or larger

FIG. 7. Free-response receiver operating characteristic curves of automated detection for lesions ≥ 6 mm without (dotted line) and with (solid line) virtual tagging.

TABLE II. Analysis of the evaluation result for the detection of lesions ≥ 6 mm. The numerator and denominator degrees of freedom were 1 and 45, respectively. Abbreviations: Δ FOM=difference in figure of merit; Δ FOM/CI=95% confidence interval for Δ FOM.

Method	<i>p</i> value	<i>F</i> statistics	Δ FOM	Δ FOM/CI
JAFROC-1	0.040	4.48	0.150	(0.010, 0.290)
JAFROC	0.034	4.81	0.161	(0.013, 0.309)
ROC	0.311	1.05	0.113	(-0.101, 0.334)

FIG. 8. Distribution of the three largest sources of false-positive detections without (left) and with (right) the application of virtual tagging. The images at bottom show two examples of major sources of false-positive detections (arrows): Untagged feces adhering to colonic wall and inhomogeneously tagged feces.

sitivity for lesions ≥ 6 mm, again, the use of virtual tagging yielded significant reduction in the number of FP detections ($p < 0.00001$).

III.C.2. Sources of FP detections

We analyzed the sources of FP detections at the FP rate of 5.4 per patient on average where, with virtual tagging, the detection sensitivity was 90% for lesions ≥ 6 mm. Figure 8 depicts the three largest sources of FP detections without (left) and with (right) virtual tagging. The figure shows that regardless of the use of virtual tagging, most FP detections (46%–55%) are caused by completely untagged feces adhering to the colonic wall. Without virtual tagging, inhomogeneously tagged feces were the second major source of FP detections (20%), and completely untagged feces not adhering to the colonic wall were the third major source (11%).

As expected, the application of virtual tagging eliminated the presence of untagged pseudolesions not adhering to the colonic wall, and it also reduced the fraction of FP detections due to poorly tagged feces. With virtual tagging, prominent folds were the second leading source of FP detections (10%), and inhomogeneously tagged feces were the third most common source of FP detections (8%). An example image of tagged feces adhering to colonic wall and an example of inhomogeneously tagged feces are shown at the bottom of Fig. 8.

IV. DISCUSSION

Several large clinical multicenter trials have shown that laxative fluid-tagging CTC has comparable detection accuracy to optical colonoscopy for large clinically significant lesions when interpreted with state-of-the-art reading methods.^{30–33} However, as we pointed out in the Introduction, laxative CTC and colonoscopy involve rigorous physical laxative purgation cleansing that is seen as the single most important reason for low colorectal screening rates. Unless the CTC examination is made easier to tolerate for patients, it may not provide a lasting solution for the colorectal screening of large populations.

The *lf*CTC would present an almost ideal colorectal examination for the screening of large populations.^{4,6} However, while several visualization and detection tools have been de-

veloped for laxative fluid-tagging CTC, the technical challenges of *lf*CTC have remained largely unaddressed.³⁴ This is partly because practically all CTC studies to date have involved bowel preparations where residual bowel materials appear as fluid due to the catharsis induced by laxative cleansing and/or by cathartic dose of a contrast agent.^{6,33,35} Furthermore, while large CTC databases have been made available for researchers, such databases have represented rigorous forms of laxative bowel preparation. A common misperception is that conventional visualization and detection methods that yield high performance in conventional fluid-tagging CTC would also yield high performance in *lf*CTC. However, several studies have demonstrated that conventional methods that have been developed for fluid-tagging CTC tend to fail in *lf*CTC, and that new approaches are needed.^{8–10,20,36–38} Unless suitable visualization and detection tools are developed for *lf*CTC, the clinical potential of *lf*CTC might never be realized. Our study represents one of the very first efforts toward developing such interpretation tools.

The virtual tagging method extends our previous work where we identified several new technical challenges that have been introduced to CTC by the applications of positive-contrast tagging and reduced bowel preparation.²⁰ Previously, we developed a method for pseudoenhancement correction of CT numbers for reliable delineation of tagged materials.¹⁰ The virtual tagging method extends this work by providing improved identification of partial-volume artifacts and poorly tagged feces that have observed radiodensities reminiscent of soft-tissue materials and thus cannot be segmented directly by global thresholding of CT numbers. In particular, such artifacts complicate the detection of flat lesions and polyps as small as 6–9 mm that tend to be represented by less than ten voxels in a mixture of soft tissue, air, and/or tagging. In addition to poorly tagged materials, the virtual tagging method is able to label image artifacts such as rectal tubes or metallic artifacts, because these artifacts tend to have radiodensities that are unlike those of normal soft-tissue structures of colonic surface.

To the best of our knowledge, this is the first study to present adequate detection accuracy in automated polyp detection for *lf*CTC. Conventional methods tend to assume that residual bowel materials appear only as fluid, and therefore

they can misinterpret solid feces and partial-volume artifacts as colorectal lesions even when tagged, thereby displaying a prohibitively large number of FP detections for clinical application of *l*/CTC.⁷ Previously, our detection scheme was shown to yield high detection accuracy in laxative CTC,^{20,25} reduced-laxative CTC,²⁰ and in populations comprising of multiple types of bowel preparations.⁸ In this study, we demonstrated that while the detection scheme already has high detection sensitivity for large polyps in *l*/CTC, application of the virtual tagging method can yield significant improvement in the detection accuracy.

The analysis of the sources of FP detections indicated that, with virtual tagging, untagged feces (55%) and prominent folds (10%) are the two leading sources of FP detections in automated detection for *l*/CTC. It is of interest to note the similarity between this and our previous result where the two largest sources of FP detections in laxative CTC were found to be prominent folds (45%) and untagged feces (25%).³⁹ However, the analysis of FP detections indicates that virtual tagging cannot fully compensate for the shortcomings of bowel preparation, and therefore it is desirable to optimize the *l*/CTC technique for minimizing the presence of large pieces of untagged feces that adhere to colonic wall and may be confused with clinically significant colorectal lesions.

This pilot study had a few limitations. First, the number of laxative-free cases (46) was relatively small. This is because we considered only true laxative-free cases where no catharsis had been induced by laxatives or by a tagging agent. There are hardly any examples of such cases, because previous clinical studies have focused on bowel preparations with cathartic components. Nevertheless, although the number of cases was small, the evaluation data of this study were independent from the development data, the data were collected at two sites with various scan parameters, all cases were considered regardless of their diagnostic quality, and we were able to demonstrate statistically significant improvement in the performance of automated detection. A second study limitation was that the number of abnormal cases (six) was small. This was a practical limitation due to the relatively low incidence of clinically significant polyps (13%) which reflects that of an asymptomatic screening population.³¹ Most clinical and validation studies that have used JAFROC or ROC analyses have involved a larger number of abnormal cases, and a larger study would be needed to reach a more definitive conclusion. A third limitation was that there were no *l*/CTC cases in the development or training data sets. If *l*/CTC cases were used also in the development data, the detection performance could be improved due to accounting for the unique characteristics of *l*/CTC cases. Finally, we did not consider human observer performance or application to other image processing applications such as electronic cleansing. Further studies are required to establish the full potential of virtual tagging in clinical applications, including a confirmation of the results of this pilot study in a large screening population.

V. CONCLUSIONS

We developed a method for virtual tagging of partial-volume artifacts and poorly tagged feces in laxative-free CTC data. Pilot evaluation with 46 clinical laxative-free cases (92 CT scan volumes) indicated that application of the method can improve the identification of poorly tagged feces, partial-volume artifacts, completely untagged feces, and artifacts such as rectal tubes. In automated polyp detection, significant improvement in detection performance was observed with the method. The results indicate that virtual tagging is a potentially useful method for complementing physical tagging in the interpretation of laxative-free CTC data.

ACKNOWLEDGMENTS

The author thank Dr. Philippe Lefere and Dr. Stefaan Gryspeerdt (Stedelijk Ziekenhuis, Roeselare, Belgium) and Dr. Michael Zalis (Massachusetts General Hospital and Harvard Medical School, Boston, MA) for providing the CTC cases for this study. This study was supported in part by Grant No. CA095279 from the U.S. Public Health Service and by research scholar Grant No. RSG-05-088-01-CCE from the American Cancer Society.

^aElectronic mail: jnappi@partners.org

¹S. J. Winawer *et al.*, "Prevention of colorectal cancer by colonoscopic polypectomy. The National Polyp Study Workgroup," *N. Engl. J. Med.* **329**, 1977–1981 (1993).

²H. I. Meissner, N. Breen, C. N. Klabunde, and S. W. Vernon, "Patterns of colorectal cancer screening uptake among men and women in the United States," *Cancer Epidemiol. Biomarkers Prev.* **15**, 389–394 (2006).

³T. Beebe, C. D. Johnson, S. M. Stoner, K. J. Anderson, and P. J. Limburg, "Assessing attitudes toward laxative preparation in colorectal cancer screening and effects on future testing: Potential receptivity to computed tomographic colonography," *Mayo Clin. Proc.* **82**, 666–671 (2007).

⁴P. Lefere, S. Gryspeerdt, M. Baekelandt, and B. Van Holsbeek, "Laxative-free CT colonography," *Am. J. Roentgenol.* **183**, 945–948 (2004).

⁵M. R. Callstrom *et al.*, "CT colonography without cathartic preparation: Feasibility study," *Radiology* **219**, 693–698 (2001).

⁶C. D. Johnson, A. Manduca, J. G. Fletcher, R. L. MacCarty, M. J. Carston, W. S. Harmsen, and J. N. Mandrek, "Noncathartic CT colonography with stool tagging: Performance with and without electronic stool subtraction," *Am. J. Roentgenol.* **190**, 361–366 (2008).

⁷H. Yoshida and J. Näppi, "CAD in CT colonography without and with oral contrast agents: Progress and challenges," *Comput. Med. Imaging Graph.* **31**, 267–284 (2007).

⁸J. Näppi and H. Yoshida, *Proceedings of the MICCAI 2008 Workshop on Computational and Visualization Challenges in the New Era of Virtual Colonoscopy*, edited by H. Yoshida (MICCAI 2008 Virtual Colonoscopy Workshop, New York, 2008), pp. 127–134.

⁹K. T. Johnson, M. J. Carston, R. J. Wentz, A. Manduca, S. M. Anderson, and C. D. Johnson, "Development of a cathartic-free colorectal cancer screening test using virtual colonoscopy: A feasibility study," *Am. J. Roentgenol.* **188**, W29–W36 (2007).

¹⁰J. Näppi and H. Yoshida, "Adaptive correction of the pseudo-enhancement of CT attenuation for fecal-tagging CT colonography," *Med. Image Anal.* **12**, 413–426 (2008).

¹¹P. Lefere, S. Gryspeerdt, J. Marrannes, M. Baekelandt, and B. Van Holsbeek, "CT colonography after fecal tagging with reduced cathartic cleansing and a reduced volume of barium," *Am. J. Roentgenol.* **184**, 1836–1842 (2005).

¹²M. E. Zalis, J. J. Perumpillichira, C. Magee, G. Kohlberg, and P. F. Hahn, "Tagging-based, electronic cleansed CT colonography: Evaluation of patient comfort and image readability," *Radiology* **239**, 149–159 (2006).

¹³M. Prokop and M. Galanski, *Spiral and Multislice Computed Tomography of the Body* (Thieme, Ludwigsburg, 2003).

- ¹⁴T. Bayes, "An essay toward solving a problem in the doctrine of chances," *Philos. Trans.* **53**, 370–418 (1763).
- ¹⁵Y. Masutani, H. Yoshida, P. MacEneaney, and A. H. Dachman, "Automated segmentation of colonic walls for computerized detection of polyps in CT colonography," *J. Comput. Assist. Tomogr.* **25**, 629–638 (2001).
- ¹⁶H. Frimmel, J. Näppi, and H. Yoshida, "Centerline-based colon segmentation for CT colonography," *Med. Phys.* **32**, 2665–2672 (2005).
- ¹⁷J. C. Russ, *The Image Processing Handbook* (CRC, Boca Raton, 1994).
- ¹⁸H. W. Sorenson, *Parameter Estimation: Principles and Problems* (Marcel Dekker, New York, 1980).
- ¹⁹R. Gonzalez and R. Woods, *Digital Image Processing* (Addison-Wesley, Reading, 1993).
- ²⁰J. Näppi and H. Yoshida, "Fully automated three-dimensional detection of polyps in fecal-tagging CT colonography," *Acad. Radiol.* **25**, 287–300 (2007).
- ²¹H. Yoshida, J. Näppi, K. Nagata, C. Richard, and D. C. Rockey, "Comparison of fully automated CAD with unaided human reading in CT colonography," *Proceedings of the Eight International Symposium of Virtual Colonoscopy*, Boston, October 15–17, 2007, pp. 96–97.
- ²²J. Näppi and H. Yoshida, "Feature-guided analysis for reduction of false positives in CAD of polyps for CT colonography," *Med. Phys.* **30**, 1592–1601 (2003).
- ²³J. J. Fenton, S. H. Taplin, P. A. Carney, L. Abraham, E. A. Sickles, C. D'Orsi, E. A. Berns, G. Cutter, R. E. Hendrick, W. E. Barlow, and J. G. Elmore, "Influence of computer-aided detection on performance of screening mammography," *N. Engl. J. Med.* **356**, 1399–1409 (2007).
- ²⁴S. Halligan, D. G. Altman, S. Mallett, S. A. Taylor, D. Burling, M. Roddie, L. Honeyfield, J. McQuillan, H. Amin, and J. Dehmeshki, "Computed tomographic colonography: Assessment of radiologist performance with and without computer-aided detection," *Gastroenterology* **131**, 1690–1699 (2006).
- ²⁵H. Yoshida and J. Näppi, "Three-dimensional computer-aided diagnosis scheme for detection of colonic polyps," *IEEE Trans. Med. Imaging* **20**, 1261–1274 (2001).
- ²⁶D. P. Chakraborty, "Validation and statistical power comparison of methods for analyzing free-response observer performance studies," *Acad. Radiol.* **15**, 1554–1566 (2008).
- ²⁷D. P. Chakraborty, "Analysis of location specific observer performance data: Validated extension of the jackknife free-response (JAFROC) method," *Acad. Radiol.* **13**, 1187–1193 (2006).
- ²⁸D. P. Chakraborty and K. S. Berbaum, "Observer studies involving detection and localization: Modeling, analysis, and validation," *Med. Phys.* **31**, 2313–2330 (2004).
- ²⁹D. D. Dorfman, K. S. Berbaum, and C. E. Metz, "ROC characteristic rating analysis: Generalization to the population of readers and patients with the jackknife method," *Invest. Radiol.* **27**, 723–731 (1992).
- ³⁰P. J. Pickhardt, J. R. Choi, I. Hwang, J. A. Butler, M. L. Puckett, H. A. Hildebrandt, R. K. Wong, P. A. Nugent, P. A. Mysliwiec, and W. R. Schindler, "Computed tomographic virtual colonoscopy to screen for colorectal neoplasia in asymptomatic adults," *N. Engl. J. Med.* **349**, 2191–2200 (2003).
- ³¹D. H. Kim, P. J. Pickhardt, A. J. Taylor, W. K. Leung, T. C. Winter, J. L. Hinshaw, D. V. Gopal, M. Reichelderfer, R. H. Hsu, and P. R. Pfau, "CT colonography versus colonoscopy for the detection of advanced neoplasia," *N. Engl. J. Med.* **357**, 1403–1412 (2007).
- ³²C. D. Johnson *et al.*, "Accuracy of CT colonography for detection of large adenomas and cancers," *N. Engl. J. Med.* **359**, 1207–1217 (2008).
- ³³B. Levin *et al.*, "Screening and surveillance for the early detection of colorectal cancer and adenomatous polyps, 2008: A joint guideline from the American Cancer Society, the US Multi-Society Task Force on Colorectal Cancer, and the American College of Radiology," *Gastroenterology* **134**, 1570–1595 (2008).
- ³⁴M. Macari and E. J. Bini, "CT colonography: Where have we been and where are we going?," *Radiology* **237**, 819–833 (2005).
- ³⁵P. Lefere and S. Gryspeerdt, *Virtual Colonoscopy: A Practical Guide* (Springer, New York, 2006).
- ³⁶M. G. Linguraru, R. L. Van Uitert, S. Zhao, J. Liu, J. G. Fletcher, C. D. Johnson, and R. M. Summers, *Proceedings of the MICCAI 2008 Workshop on Computational and Visualization Challenges in the New Era of Virtual Colonoscopy*, edited by H. Yoshida (MICCAI 2008 Virtual Colonoscopy Workshop, New York, 2008), pp. 85–90.
- ³⁷W. Cai, M. Zalis, J. Näppi, and H. Yoshida, "Structure-analysis method for electronic cleansing in CT colonography," *Med. Phys.* **35**, 3259–3277 (2008).
- ³⁸M. E. Zalis, H. Yoshida, J. Näppi, C. Magee, and P. Hahn, "Evaluation of false-positive detections in combined computer-aided polyp detection and minimal preparation/digital subtraction CT colonography (CTC)," *Proceedings of the RSNA, Chicago*, November 28–December 3, 2004, p. 578.
- ³⁹A. H. Dachman, J. Näppi, H. Frimmel, and H. Yoshida, "Sources of false positives in computerized detection of polyps in CT colonography," *Radiology* **225**, 303 (2002).