2009

Optimizing Computer-Aided Colonic Polyp Detection for CT Colonography by Evolving the Pareto Front

Jiang Li
Old Dominion University

Adam Huang

Jack Tao

Jiamin Liu

Robert L. Van Uitert

See next page for additional authors

Original Publication Citation

Li, J., Huang, A., Yao, J., Liu, J. M., Van Uitert, R. L., Petrick, N., & Summers, R. M. (2009). Optimizing computer-aided colonic polyp detection for CT colonography by evolving the pareto front. *Medical Physics, 36*(1), 201-212. doi:10.1118/1.3040177

Optimizing computer-aided colonic polyp detection for CT colonography by evolving the Pareto front[a)]

Jiang Li
Radiology and Imaging Sciences, Clinical Center, National Institutes of Health, Bethesda, Maryland 20892-1182 and Department of Electrical and Computer Engineering, VMASC, Old Dominion University, Norfolk, Virginia 23529

Adam Huang, Jack Yao, Jiamin Liu, and Robert L. Van Uitert
Radiology and Imaging Sciences, Clinical Center, National Institutes of Health, Bethesda, Maryland 20892-1182

Nicholas Petrick
Division of Imaging and Applied Math, LAMIS Image Analysis Laboratory, FDA/CDRH/OSEL, 10903 New Hampshire Avenue WO62-4116, Silver Spring, Maryland 20993-0002

Ronald M. Summers[b)]
Radiology and Imaging Sciences, Clinical Center, National Institutes of Health, Bethesda, Maryland 20892-1182

(Received 1 August 2008; revised 27 October 2008; accepted for publication 12 November 2008; published 22 December 2008)

A multiobjective genetic algorithm is designed to optimize a computer-aided detection (CAD) system for identifying colonic polyps. Colonic polyps appear as elliptical protrusions on the inner surface of the colon. Curvature-based features for colonic polyp detection have proved to be successful in several CT colonography (CTC) CAD systems. Our CTC CAD program uses a sequential classifier to form initial polyp detections on the colon surface. The classifier utilizes a set of thresholds on curvature-based features to cluster suspicious colon surface regions into polyp candidates. The thresholds were previously chosen experimentally by using feature histograms. The chosen thresholds were effective for detecting polyps sized 10 mm or larger in diameter. However, many medium-sized polyps, 6–9 mm in diameter, were missed in the initial detection procedure. In this paper, the task of finding optimal thresholds as a multiobjective optimization problem was formulated, and a genetic algorithm to solve it was utilized by evolving the Pareto front of the Pareto optimal set. The new CTC CAD system was tested on 792 patients. The sensitivities of the optimized system improved significantly, from 61.68% to 74.71% with an increase of 13.03% (95% CI [6.57%, 19.5%], $p=7.78\times10^{-5}$) for the size category of 6–9 mm polyps, from 65.02% to 77.4% with an increase of 12.38% (95% CI [6.23%, 18.53%], $p=7.95\times10^{-5}$) for polyps 6 mm or larger, and from 82.2% to 90.58% with an increase of 8.38% (95%CI [0.75%, 16%], $p=0.03$) for polyps 8 mm or larger at comparable false positive rates. The sensitivities of the optimized system are nearly equivalent to those of expert radiologists. © *2009 American Association of Physicists in Medicine*. [DOI: 10.1118/1.3040177]

Key words: computer-aided detection, pattern recognition, statistical methods, multiobjective evolution, genetic algorithm

I. INTRODUCTION

Colon cancer is the second leading cause of cancer death in the U.S.[1] It is known that colorectal cancer can be prevented if a screening procedure is performed and any polyps that are found are removed. Optical colonoscopy is currently considered to be the reference standard tool for polyp detection. In a colonoscopy, a doctor inserts a long flexible scope into a patient's colon. Computed tomographic colonography (CTC), in which radiologists examine a detailed colon picture created from CT images, has been studied as an alternative noninvasive screening procedure for the past 10 years and shows promising results as a colorectal screening tool.[2] Computer-aided polyp detection (CAD) has been suggested as one way of aiding the radiologist in reading these exams.[3–7]

True polyps generally stand out as small growths in the colon lining which can be detected by an appropriately designed computer algorithm utilizing shape and 3D texture information derived from CT images. Methods in existing CTC CAD systems usually consist of several steps for detecting polyps. An initial step identifies polyp candidates on the colon surface and a refined classification step reduces false positives based on additional information extracted from 3D volume data. Sometimes, additional false positive reduction steps are used to further increase specificity.[8–10] Current CTC CAD systems have high sensitivities and specificities for detecting polyps 1 cm or larger in diameter, but tend to have low sensitivities and specificities when detecting smaller 6–9 mm polyps.[3–6]

There are a set of parameters associated with each of the

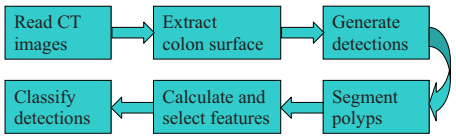


FIG. 1. A block diagram representation of our CTC CAD system.

detection steps. To achieve a better performance for CTC CAD systems, those parameters need to be optimized. A number of methods[3–6,11–14] have been proposed for optimizing the parameters in the second classification step. Few attempts have been made for optimizing those parameters involved in the initial polyp detection on the colon surface, partially due to the lack of a mathematical formalization of the problem and the potentially prohibitive computational burden associated with the optimization task. Indeed, we found that many of the 6–9 mm polyps undetected by our CAD software were missed in the initial polyp detection step. The purpose of this article is to formalize the initial detection step in the framework of the multiobjective evolutionary computation[15] and to optimize the associated parameters utilizing a genetic algorithm and parallel computing techniques.

II. BACKGROUND

Our previously developed CTC CAD system identifies polyps based on geometric features of the colon surface and volumetric properties of the candidate polyps.[16] A brief outline of our system is provided in Fig. 1. For a set of CT scan images, the CTC CAD system first segments the colon using a region growing algorithm.[17] The colon surface is then extracted using an isosurface technique.[18] Geometric features such as curvature are calculated for each vertex on the colon surface and input to a sequential classifier, which uses a set of predefined thresholds to filter vertices on the colon surface. After the filtering process, connected vertices that survive the filtering are clustered together as polyp candidates.[19] This process is referred to as the initial polyp detection procedure. A knowledge-based polyp segmentation algorithm is then applied to the 3D volume data, starting from the initial polyp candidates, to segment the polyp detections.[20] Next, more than 100 quantitative features are calculated for each segmented polyp candidate. A feature selection procedure based on classifier accuracy is used to reduce the number of features to less than 20. Finally, the selected features are presented to a support vector machine (SVM) committee classifier to classify the polyp candidate as a true or false detection.[14,21]

A large population study showed that our CTC CAD system can successfully detect polyps 1 cm or larger in diameter, with a sensitivity of 89.3% at the specificity of 2.1 false positives per patient.[16] However, sensitivities and specificities for detecting polyps smaller than 1 cm are much lower, i.e., the sensitivity is only 61.3% for detecting polyps 6 mm or larger at the specificity of 7.9 false positives per patient.[16] There is a consensus that polyps sized 1 cm and larger are important, which is the threshold above which colonic polyps are at significantly greater risk of progressing to malignancy. Polyps smaller than 6 mm are difficult to detect reliably and are generally thought to be of little clinical importance.[22] However, 6–9 mm polyps are in a size range that is also likely to be important for clinical management.[23] While most 6–9 mm polyps are believed not to grow and are low in risk, if the polyp is found to grow it could then be removed to reduce its possibility of progressing to malignancy. Some researchers advocate that such polyps may be observed by CTC rather than undergo immediate resection in optical colonoscopy. Consequently, there is a need to improve the sensitivity of CTC CAD systems for detecting these 6–9 mm polyps.

We found that most of the undetected 6–9 mm polyps by our CTC CAD system were removed by the sequential classifier in the initial polyp detection procedure. The predefined thresholds in the sequential classifier were either empirically chosen or derived using mathematical modeling.[4,5] However, setting optimal thresholds for the sequential classifier is not an easy task. The thresholds are paired to define acceptable ranges on the feature values. Wider ranges, which admit more vertices, form more polyp candidates but also produce a higher number of false positives because of the many false positive candidates. On the other hand, narrower ranges, which admit fewer vertices, reduce false positives by filtering out more vertices on the colon surface but increase false negatives. The two objectives, a low number of false positives and false negatives, are conflicting and usually cannot be minimized independently: a smaller number for one objective is usually accompanied by a larger number for the other. The goal of this article is to obtain a set of thresholds that jointly minimize the two objectives.

Traditional optimization algorithms dealing with one objective (or cost function) may be used to optimize these thresholds, but require that the two objectives be combined together into a single overall cost function. In practical applications, it is often difficult to know the specific weights that optimally combine the two objectives, though we all might agree that reducing false negatives is more important than reducing false positives. Inappropriate weighting of the two objectives in early polyp detection stages may prevent the overall optimization goal, i.e., achieving a good free response receiver operating characteristic curve (FROC) for our CTC CAD system. Any suboptimal weighting scheme leads to a loss of information.

An alternative solution is to formulate the task as a multiobjective problem (MOP).[15] In the MOP, one tries to keep all possible best solutions without weighting the two objectives in the optimization procedure. Therefore, the possible information loss associated with the weighting is prevented. At the end of the optimization, a set of noninferior or nondominated solutions known as a Pareto set is obtained. The Pareto optimal set is composed of a Pareto front in the objective space,[24–27] and provides freedom for decision makers to make a final determination in the trade-offs between the two objectives after the optimization.

In this article we formalize the initial polyp candidate detection procedure on the colon surface as an MOP prob-

lem, i.e., we optimize the thresholds used by the sequential classifier such that the number of false negatives and false positives produced in the initial detection procedure are jointly minimized. We utilize a multiobjective evolutionary method, the Strength Pareto Evolutionary Algorithm (SPEA2),[28] to solve the MOP problem. The study of multiobjective optimization based on evolutionary methods began in 1985.[15] Subsequently, many multiobjective optimization algorithms have been proposed in the literature.[25–29] We are interested in the SPEA2 algorithm because of its good performances and fast convergence rate.[24] Convergence speed is important for our task due to the expensive computation on colon surface.

Note that using the SPEA2 algorithm to optimize the thresholds based on a large set of patient data is computationally prohibitive for a single computer. The computational barriers of the optimization come from the fact that we need to traverse millions of vertices to evaluate the goodness of one set of thresholds on each colon surface. Furthermore, the optimization procedure is performed on many colon surfaces and is often repeated many times till the optimization converges. We overcome the computational challenge by using parallel computing techniques. Evaluation of one set of thresholds is surface independent, that is, the performance of the thresholds on one surface does not depend on that of other surfaces. We thus can distribute the evaluation process to different computing nodes on a Beowulf cluster and collected the results after each node completes its assigned job. The Beowulf cluster we utilized in this article is a parallel virtual supercomputer located at the National Institutes of Health (http://www.lobos.nih.gov/).

Our previous work on a small data set showed that the optimized thresholds can obtain a much better training result than that of heuristically chosen thresholds.[19] A later work showed that the optimized thresholds are generalizable to new data.[30] In this article, we performed the optimization process on a very large data set of 1186 patients. We randomly divided these patients into training ($n=394$) and testing ($n=792$) sets in a 1:2 ratio, and performed the optimization process on some of the training data to find the best thresholds. The resulting thresholds were then applied to the testing data.

In the following sections, we first describe the sequential classifier and the features used in the initial detection procedure. We then describe the optimization goal and the technical details associated with achieving this goal. Next, the proposed method utilizing the SPEA2 algorithm is outlined and results for the initial detection are presented. Finally, we validate the optimized CTC CAD system on a large testing data set using FROC analysis, and compare this performance with our original nonoptimized approach.

III. METHOD

In this section, we describe a set of features used in the sequential classifier which forms the initial polyp detections on the colon surface.[31] We then describe an algorithm to optimize the feature thresholds in our initial rule-based candidate detection stage.[19]

III.A. Clustering polyp candidates

Colonic polyps can be characterized by surface curvatures.[3] Surface curvatures are local geometric properties which quantitatively describe how the surface curves or bends locally. This surface shape can be characterized by two principal curvatures which are the maximum k_1 and minimum k_2 principal curvatures along the principal tangent directions, and polyps can be identified as regions with negative k_1 and k_2.[3,32]

Based on the two principal curvatures, the mean curvature H and the Gaussian curvature K are defined as

$$\begin{cases} H = \dfrac{k_1 + k_2}{2}, \\ K = k_1 \cdot k_2. \end{cases} \tag{1}$$

We calculate the mean curvature H and Gaussian curvature K for each vertex on the colon surface using a kernel method.[32,33] To form polyp candidates, we first check vertices on the surface if the following criteria are satisfied:

$$\omega_1 < H < \omega_2 \quad \text{and } \omega_3 < K < \omega_4, \tag{2}$$

where ω_i, $i=1,\ldots,4$ are preset thresholds to be optimized. Vertices that meet the above conditions and are connected to each other are clustered together as initial polyp candidates. After all vertices are examined, two additional features, mean sphericity S_m and number of vertices N, are calculated for each formed polyp candidate,[3]

$$S_m = 2 \left| \frac{1/N \sum k_2 - 1/N \sum k_1}{1/N \sum k_2 + 1/N \sum k_1} \right|, \tag{3}$$

where the summations are over the formed polyp candidate. The sphericity denotes how round a surface is and ranges from 0 (sphere) to 2 (ridge). Any intermediate value represents an ellipsoid. If

$$\omega_5 < S_m < \omega_6 \quad \text{and } N \geqslant \omega_7, \tag{4}$$

where $\omega_5,\ldots,\omega_7$ are again thresholds, the polyp candidate is kept and delivered to the next step.

The next step is to examine each of the polyp candidates by checking how many vertices in the candidate and the vertices' neighbors having negative principal curvatures k_1 and k_2, which might be outside of the candidate, satisfy the conditions

$$H < \omega_8 \quad \text{and } \omega_9 < S < \omega_{10}, \tag{5}$$

where again, $\omega_8,\ldots,\omega_{10}$ are preset thresholds, and

$$S = 2 \cdot \frac{k_2 - k_1}{k_2 + k_1} \tag{6}$$

is sphericity for a single vertex. In this step, the polyp candidate can grow or shrink if the set of thresholds $\omega_8,\ldots,\omega_{10}$ is "wider" or "narrower" than the corresponding thresholds

in Eqs. (2) and (4). If $\omega_8, \ldots, \omega_{10}$ are wider than those in Eqs. (2) and (4), the classifier will include some unclustered neighbors into the polyp candidate such that the polyp candidate grows. On the other hand, if a set of narrower thresholds is selected, only part of the vertices in the polyp candidate satisfies the criteria in Eq. (5) and is kept. Finally, if N_b, the number of vertices satisfying Eq. (5), is greater than or equal to ω_{11}, where ω_{11} is another threshold, the candidate is delivered to the polyp segmentation procedure.[34] These rules result in a total of 11 thresholds that need to be optimized.

Note that the "growing" or "shrinking" of the initial polyp candidate depends upon the values of thresholds, and the thresholds are optimized based on the available training data using a genetic algorithm described later in this article. Using this approach, the polyp candidate generation is adapted to training data.

III.B. Problem formulation

The task of our optimization problem is to find a set of thresholds for the sequential classifier such that our CTC CAD system produces the minimal number of false negatives and false positives in the initial detection procedure. However, those two objectives are conflicting, and minimizing one of them usually leads to an increase in the other. Traditional algorithms usually weight the conflicting objectives during optimization and obtain a single solution for the problem. In practice, it is difficult to find the optimal weights for the objectives resulting in a solution that is likely suboptimal. Alternatively, recently developed Pareto front-based multiobjective optimization algorithms do not weight the objectives but instead provide a set of nondominated solutions to the problem.[28] The nondominated solutions consist of all possible "best" solutions in terms of the conflicting objectives. Based on the framework of multiobjective optimization, we can form our task as follows.

Let $f_1(\vec{\omega}), f_2(\vec{\omega})$ denote the number of missed polyps (false negatives) and average number of false positives per data set, respectively, where $\vec{\omega}$ is an 11-dimensional threshold vector

$$\vec{\omega} = \{\omega_1, \ldots, \omega_{11} | \omega_1, \ldots, \omega_6 \in R, \omega_8, \ldots, \omega_{10} \in R, \omega_7 \in I, \omega_{11} \in I\}. \tag{7}$$

Our multiobjective minimization problem (MOP) can be stated as

$$\min_{\vec{\omega}} \vec{f}(\vec{\omega}) = \{f_1(\vec{\omega}), f_2(\vec{\omega})\}, \tag{8}$$

subject to

$$\begin{cases} \omega_1 \in [-10,0), \omega_2 \in [-10,0), \omega_1 < \omega_2, \\ \omega_3 \in [-60,60), \omega_4 \in [0,100], \omega_3 < \omega_4, \\ \omega_5 = 0, \omega_6 \in (0,2], \omega_7 \in [6,30], \\ \omega_8 \in [-10,0), \\ \omega_9 = 0, \omega_{10} \in (0,2], \omega_9 < \omega_{10}, \\ \omega_{11} \in [6,30]. \end{cases} \tag{9}$$

The range for each threshold was determined experimentally such that it is wide enough to include all solutions of interest, but limited in ranges to make the optimization more manageable. The global optima of an MOP is called the Pareto optimal set, which consists of solutions (thresholds) that are not dominated by any other solutions.[35]

A solution $\vec{\omega}_1$ is said to dominate $(\succ)\vec{\omega}_2$ if the objective vector $\vec{f}(\vec{\omega}_1)$ is less than or equal to $\vec{f}(\vec{\omega}_2)$ in all attributes, and strictly less than $\vec{f}(\vec{\omega}_2)$ in at least one of these attributes,

$$\begin{cases} \vec{\omega}_1 \succ \vec{\omega}_2, iff \\ \forall i \in \{1,2\}: f_i(\vec{\omega}_1) \leqslant f_i(\vec{\omega}_2) \wedge \exists j \in \{1,2\}: f_j(\vec{\omega}_1) < f_j(\vec{\omega}_2) \end{cases}. \tag{10}$$

The space formed by the objective vectors of the Pareto optimal solutions is called the Pareto front.[28] It is clear that any final design solution should preferably be a member of the Pareto optimal set. Therefore, identifying a set of Pareto optimal solutions is key for a decision maker's selection of a "compromise" solution. In this study, we utilized the SPEA2 algorithm to obtain the Pareto optimal set for our thresholds. We recently showed that the optimized sequential classifier had a very good generalization capability on unseen testing data.[30]

III.C. Problem solving

The Pareto optimal solution is often obtained by multiobjective optimization algorithms. The study of multiobjective optimization based on evolutionary methods began in 1985.[15] Subsequently, many multiobjective optimization algorithms have been proposed in the literature.[25–29] We are interested in the SPEA2 algorithm because of its good performances and fast convergence rate.[24] Before starting to describe the algorithm, we introduce several terminologies often used in evolutionary computation.

- Population: A set of solutions to be optimized
- Population size: Number of solutions in the population
- Fitness: A value denoting the goodness of a solution in the population, the smaller the better in this article
- Individual: A solution in the population
- Crossover: A genetic operator used to combine two individuals to produce two new individuals
- Mutation: A genetic operator used to alter an individual to form a new individual
- Generation: One iteration of the algorithm

The SPEA2 algorithm usually consists of four steps in the optimization procedure: (1) Initialization: Randomly initialize the solution population; (2) Fitness evaluation: Evaluate and assign a fitness value for each individual in the popula-

tion according to its performance; (3) Environmental and mating selection: Select individuals based on their performances so that better individuals are more likely to be selected for producing the next generation, and (4) Reproduction: Use crossover and mutation to produce the next generation from the selected individuals. In step 2, either a high or a low fitness value can be used to represent a better performance. In this paper, a smaller fitness value indicates a better performance. Step 2 through step 4 are repeated till a specified generation number is reached. We briefly describe the SPEA2 algorithm in the Appendix; more background about this algorithm can be found in Refs. 28 and 36.

III.D. Algorithm outline

We now outline the optimizing procedure for the thresholds in the initial polyp detection on the colon surface.

(1) Initialization:
Set $N=100$ (population size),
$\bar{N}=100$ (archive size),
$T=300$ (generation number),
Randomly initialize P_t with sets of thresholds and generate an empty archive $\bar{P}_t$. Each threshold is coded into an 8-bit binary string. One individual solution, which consists of 11 thresholds, is thus 88 bits long.
(2) Fitness evaluation: We distribute colon surfaces to multiple CPU nodes in the Beowulf cluster for evaluation of each individual set of thresholds in P_t and $\bar{P}_t$. Each CPU runs the clustering program that computes the number of missed polyps and the number of false positives on the assigned colon surfaces. After jobs at all CPU nodes are completed, results are collected and the total number of missed polyps and the average number of false positives are calculated. The fitness value is then calculated using the method described in the Appendix.
(3) Termination check: If t, the current generation number, is greater than T or other specified condition is satisfied, return nondominated individuals in $\bar{P}_t$ as the final result.
(4) Environmental selection: Copy all nondominated individuals in P_t and $\bar{P}_t$ to $\bar{P}_{t+1}$. If the size of $\bar{P}_{t+1}$ exceeds $\bar{N}$, truncate $\bar{P}_{t+1}$ by deleting the worst solutions (highest fitness values) in $\bar{P}_{t+1}$. If the size of $\bar{P}_{t+1}$ is less than $\bar{N}$, copy the dominated solutions in P_t having smaller fitness values (better solutions) into $\bar{P}_t$ such that the size of $\bar{P}_{t+1}$ equals $\bar{N}$.
(5) Mating selection: Select 100 individuals in $\bar{P}_{t+1}$ with replacement using the binary tournament procedure.
(6) Reproduction: Reproduce the next generation using the standard crossover and mutation procedures.[36] The crossover and mutation probability were set to 0.9 and 0.01, respectively, in our experiments. Store the results in P_{t+1}. Set $t=t+1$ and go to step 2.

The optimization algorithm returns the Pareto optimal solutions for the given problem after $T=300$ generations. In the reproduction process, it is possible that some of the reproduced individuals are not feasible due to the constraints in Eq. (9). For example, the individuals are outside the ranges. We used the simplest approach to handle the constraints. Individuals not meeting the constraints were rejected and new individuals were regenerated till a feasible solution was produced. See Refs. 36 and 37 for other constraint handling strategies.

IV. EXPERIMENTS

IV.A. Data acquisition

The patient population consisted of 1253 asymptomatic adults between 40 and 79 years of age at three medical centers (institutions 1–3), of whom 1233 underwent complete same day virtual and optical colonoscopy. Twenty of the 1253 patients were excluded because of incomplete optical colonoscopy, inadequate preparation, or failure of the CT colonographic system. Full CT colonography data were available for 1186 of these patients. The use of this patient data for CAD development and assessment was part of the original institutional review board-approved project and included in the patient consent forms that were signed by all patients participating in the original study;[2] continuing IRB review was subsequently waived. Recently we published the performance of our CAD system using a portion of this data set.[16]

Patients underwent a 24 h colonic preparation that consisted of oral administration of 90 mL sodium phosphate, 10 mg bisacodyl, 500 mL barium (2.1% by weight), and 120 mL diatrizoate meglumine and diatrizoate sodium given in divided doses. Each patient was scanned in the supine and prone positions using a high-resolution scanner (GE Healthcare Technologies, Waukesha, WI). CT scanning parameters included 1.25- to 2.5 mm section collimation, 15 mm/s table speed, 1 mm reconstruction interval, 100 mAs, and 120 kVp. Optical colonoscopy was performed the same day by one of 17 experienced colonoscopists. Colonoscopists used a calibrated guidewire to measure polyp size, recorded whether the polyp was located on a haustral fold, and gave a subjective assessment of polyp shape (sessile, pedunculated, or flat).

IV.B. Ground truth recording and matching

The ground truths for polyps were based on manual determination of the three-dimensional borders of polyps using Viatronix software (VIATRONIX V3D COLON, research version 1.3.0.0; Viatronix, Stony Brook, NY). Each polyp greater than 6 mm found at optical colonoscopy was located on the prone and supine virtual colonoscopy examinations and a marker was placed in the center of each polyp manually. The border of each polyp was then traced in each CT slice and the location of each voxel inside the polyp was recorded. The ground truth recording procedure was described previously.[16] We compared the CAD polyp detection on colon surface with the ground truth in a blinded fashion. If any part of detection matched any part of a manual tracing of a polyp,

TABLE I. Patient population in the database.

	Train (n=394)[a]	Test (n=792)[a]
No. of men (%)	227 (57.6)	473 (59.7%)
No. of women (%)	167 (42.4)	319 (40.3%)
No. at institution 1 (%)	122 (31.0)	283 (35.7)
No. at institution 2 (%)	123 (31.2)	190 (24.0)
No. at institution 3 (%)	149 (37.8)	319 (40.3)
Age, y (mean ± SD)	58.0 ± 7.4	57.7 ± 7.1

[a]n denotes the number of patients.

the detection was considered a true positive; otherwise, the detection was considered a false positive.

IV.C. Data partition

We randomly divided the patient cohort into a training and a testing set using the same partitions as in our previous article.[16] We choose to train on one-third and test on the remaining two-thirds of the data because this partitioning of the data yields conservative estimates of sensitivity and specificity. Random selection was the only rule used in the partitioning and led to the polyp distributions in the training and test sets. Table I shows characteristics of patients in the training and test sets. Table II summarizes polyp information in the training and test data sets.

IV.D. Multiobjective optimization

The goal of the multiobjective optimization is to obtain an optimal set of thresholds for generating initial polyp detections on colon surface. Constructing colon surface is a computationally expensive procedure in the CTC CAD program, and the surfaces do not depend on the Pareto point chosen. To avoid repeating this time-consuming process, we ran our CTC CAD program one time for each patient data used in the optimization and saved the segmented colon surfaces to a file. In the SPEA2 algorithm, the saved colon surfaces were then retrieved and used in the optimization procedure. There were 79 patients in the training data set who had polyps and 315 normal patients. We selected all 79 patients with polyps and 55 randomly chosen normal patients in the training data set for the multiobjective optimization. We thus had 134 patients selected for threshold optimization.

TABLE II. Polyp ground truth information.

	Train (79 patients having polyps)	Test (173 patients having polyps)
No. of adenomas (%)		
6–7 mm	24 (40.0)	67 (55.4)
8–9 mm	17 (28.3)	24 (19.8)
⩾10 mm	19 (31.7)	28 (23.1)
No. of carcinomas (%)	0 (0.0)	2 (1.7)

Numbers are polyps which are retrospectively identifiable.

IV.E. Parallel computing

In the fitness evaluation procedure, we distributed the 268 colon surfaces for the selected 134 patients into multiple compute nodes. The evaluation process is independent for each colon surface, which allowed us to take advantage of the parallel computing technique. Our Beowulf cluster consisted of 1250 compute nodes (2500 CPUs) running the GNU/LINUX operating system. To evaluate a set of thresholds, we assigned four colon surfaces to each compute node (each CPU handles two surfaces). Each CPU computed the number of false negatives and false positives on the assigned surfaces for the set of thresholds. After all nodes accomplished the assigned jobs, we collected the results from each node and calculated the total number of false negatives and the average false positive rate. There were 67 compute nodes involved in each round of the fitness evaluation procedure on the 268 colon surfaces.

We ran the multiobjective genetic algorithm 300 generations and obtained the Pareto optimal set, which contains a set of nondominated solutions that compose the Pareto front in the objective space. After around 250 generations, the Pareto front became stable. A Pareto front is stable if no point on the front changes its location during one generation to the next. We found that 300 generations were adequate to approximate the asymptotic solution for the SPEA2 algorithm.

The time required to run each case varied from 5 to 20 s depending upon the complexity of the surface. It took approximately 7 h to run all cases through the 300 generations on the Beowulf cluster. For those who do not have access to such a large computational infrastructure, the whole optimization will take about 9.8 days for a single PC having four CPUs. With the rapid development of multiple core machines, this type of computational power soon will be much more commonplace even in home PCs.

IV.F. Operation point selection on the Pareto front

While the Pareto optimal set contains many good solutions, in practice we need to select one particular point on the Pareto front with a good trade-off between the number of false negatives and the false positive rates. The chosen point should have a high sensitivity and a low false positive rate for the overall CTC CAD system. Due to its expensive computation, we only evaluated three points on the resulting Pareto front by performing FROC analysis on the training data set. For each chosen point on the Pareto front, we ran the CTC CAD system using the corresponding set of thresholds in the sequential classifier to form initial polyp detections. The feature and SVM committee selection were also performed to produce an FROC curve for the chosen Pareto front point. The final chosen operational point on the Pareto front corresponded to the point achieving the best FROC curve. All these experiments were performed on the training data set (394 patients).

IV.G. Data analysis

We computed sensitivity using only those polyps found at segmentally unblinded optical colonoscopy and visible on retrospective review of the CT colonography images. The technique for segmental unblinding of virtual colonoscopy results at optical colonoscopy has been previously described.[2] Some of the polyps, especially those 6 or 7 mm in diameter, could not be found on the supine and/or prone views. Therefore, it is not possible to train on them or to test whether CAD can detect them.

We compared the optimized CTC CAD system with our previous reported system in terms of FROC analysis.[16]. FROC analysis produces curves that graphically show the sensitivity of CAD for detecting polyps versus false positives rate for different settings of a tunable parameter in the classifier. For comparison purposes, we present FROC curves for different adenoma polyp size categories and for training and testing as reported in Ref. 16. Because we are particularly interested in detecting the medium-sized polyps in this article, we also present a performance comparison between the prior system and the optimized system on the medium-sized polyps, 6–9 mm in diameter. Please See Fig. 1.

While FROC curves show the spectrum of CAD sensitivities across a range of false positive rates, for clinical use a CAD system is typically operated at a fixed operation point on the FROC curve. We selected one operation point for each of the polyp size categories based on its training FROC curves, and report the sensitivity and false positive rate at that operation point on the corresponding testing FROC curve. The operation points were chosen on relatively flat parts of the FROC curves where there were diminishing gains in sensitivity as the false positive rates increased. For comparison, if the point chosen above for the optimized system was far away from the point that was chosen for the prior system, we either changed the point of the optimized system to the one that is closest to the prior system's operation point, or vice versa. The operating points were chosen somewhat arbitrarily, but represented reasonable trade-offs between sensitivity and false positive rates.[16]

To perform statistical analysis on results obtained from our prior system and those from the optimized system, a bootstrapping technique was used on the testing data to compute standard deviations over a range of operating points.[38] The bootstrapping was conducted by determining FROC curves for each of 100 random samples of 792 test patients with replacement (duplicates allowed). With the bootstrapped test results, we used a Gaussian fitted to the bootstrapped mean and variance of the mean to estimate confidence intervals and p values for the sensitivity differences between the two systems. We considered statistical significance to be $p<0.05$.

V. RESULTS

V.A. Pareto front on the selected patients

The Pareto front obtained by SPEA2 on the selected 134 patients is shown in Fig. 2. The three operation points on the Pareto front are also shown. The middle point was chosen such that it had a false positive number close to that of our prior system, and there was a gap to the next point on the front. The other two points relatively evenly split the front but were chosen arbitrarily. The Pareto front indicates that a smaller false negative number is usually associated with a higher false positive rate, which concurs with our intuition.

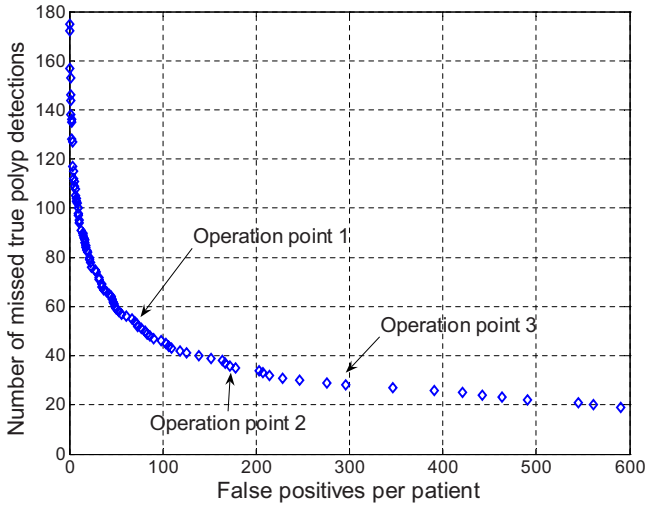


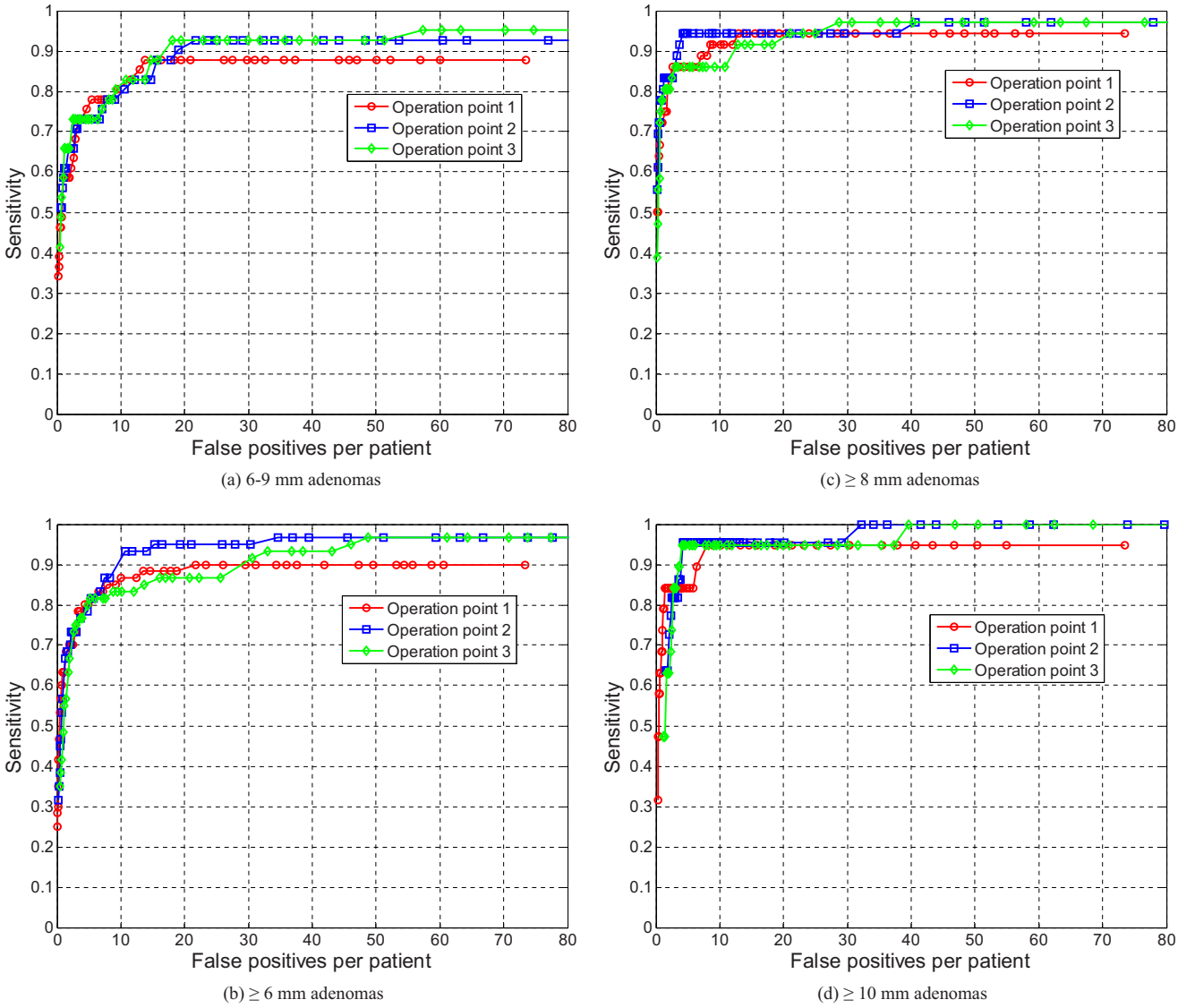
FIG. 2. The Pareto front from the SPEA2 algorithm on 134 patients selected from the training data set. The horizontal axis represents false positive rates per patient and the vertical axis denotes the number of missed true detections. Three potential operation points are also shown. The operation point 1 has 51 missed true detections with 77.6 false positives per patient, while point 2 has 32 missed true detections with 179 false positives per patient, and point 3 has 28 missed true detections with 296 false positives per patient.

V.B. Operation points evaluation on the Pareto front

The FROC curves for the training data using the three points on the Pareto front are shown in Fig. 3. Operation point 2 gave the best FROC curves (in the range of 0 to 20 false positives per patient) for size categories of $\geqslant$6 mm and $\geqslant$8 mm adenoma polyps. For the size category of 6–9 mm the three operation points performed similarly. Operation point 3 was the best for size category of $\geqslant$10 mm. Based on these observations, we chose operation point 2 as the new optimized configuration for the CTC CAD system.

V.C. FROC analysis

FROC curves (Fig. 4) and testing results (Table III) show that the optimized CTC CAD system had a better performance than that of our prior system for detecting adenomas in all size categories except the 10 mm or larger category. The optimized system was especially good at detecting the "medium-size" polyps, 6–9 mm in diameter. For example, at the chosen classifier operation points, the sensitivity improvement was 13.03% [6.57%, 19.5%] ($p<10^{-4}$). The increases in sensitivity for detecting polyps 6 mm or larger and 8 mm or larger at the chosen operation points were 12.8% [6.23%, 18.53%] ($p<10^{-4}$) and 8.38% [0.75%, 16%] (p



(a) 6-9 mm adenomas

(c) ≥ 8 mm adenomas

(b) ≥ 6 mm adenomas

(d) ≥ 10 mm adenomas

FIG. 3. FROC curves on patients in training data set for each of the three chosen Pareto front operation points.

=0.03), respectively. The optimized system has a similar sensitivity (91.28%) compared to that of the prior system (91.09%) for detecting polyps sized 10 mm or larger with a difference of 0.19% [−12.89%, 13.27%] (p=0.98).

V.D. Examples of detected polyps

Figure 5 shows two true polyps detected by our CTC CAD system but not detected by the prior system. The optimized system detected four and 11 more medium-sized polyps than the prior CTC CAD system in the training and testing data sets, respectively.

V.E. Optimized parameters

The optimized parameters corresponding to the operation point 2 on the Pareto front in Fig. 2 with the best FROC results have the following values:

$$\omega_1 = -10, \quad \omega_2 = -0.47, \quad \omega_3 = -35.94,$$

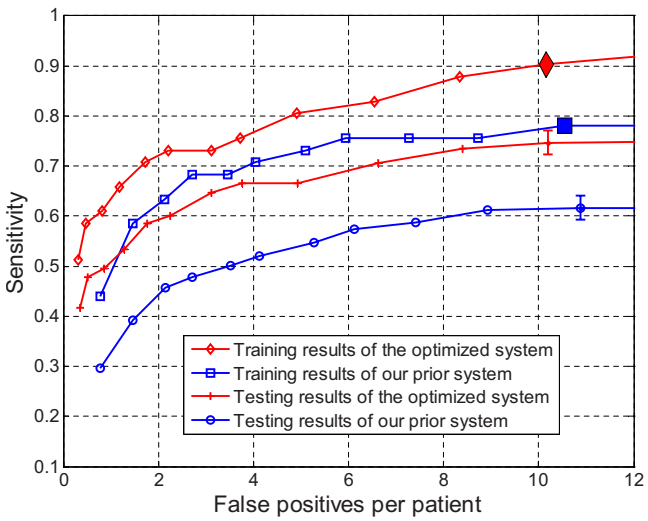
$$\omega_4 = 18.75, \quad \omega_5 = 0, \quad \omega_6 = 1.30, \quad \omega_7 = 19,$$

$$\omega_8 = -0.625, \quad \omega_9 = 0, \quad \omega_{10} = 1.02, \quad \omega_{11} = 11.$$

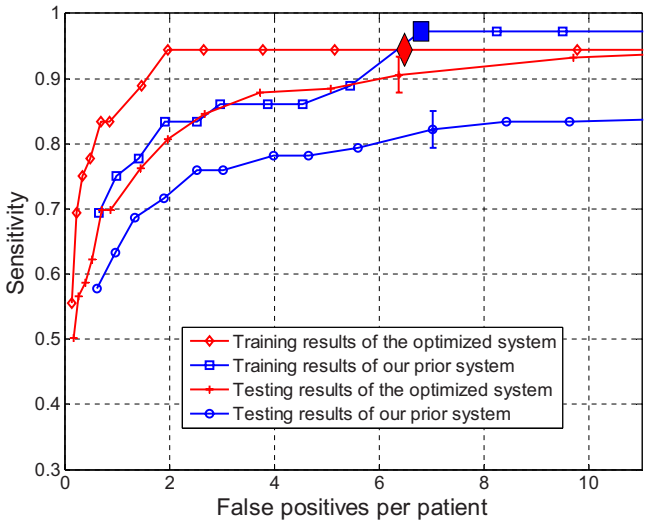
Recall that ω_7 is the number of vertices in the cluster in the first step of the initial clustering procedure and ω_{11} is the number of vertices in the second step of the initial clustering procedure. Because the value of ω_7 is bigger than that of ω_{11}, the initial vertex cluster shrinks during the clustering procedure.

VI. DISCUSSION

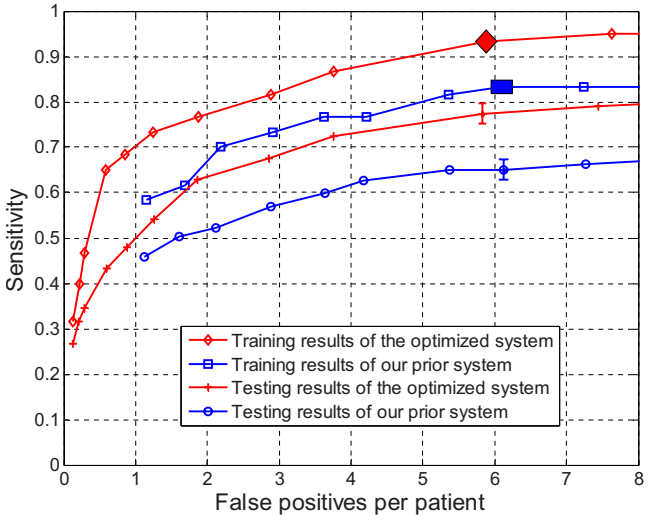
The optimized CAD system has a significantly higher sensitivity for detecting medium-sized polyps. At a similar false positive rate to our earlier CAD system, the optimized system improved sensitivity by about 13% for polyps sized 6–9 mm in diameter, yielding a sensitivity of 74% for those retrospectively identifiable polyps. This sensitivity is close to that of the radiologists who made the original interpretations



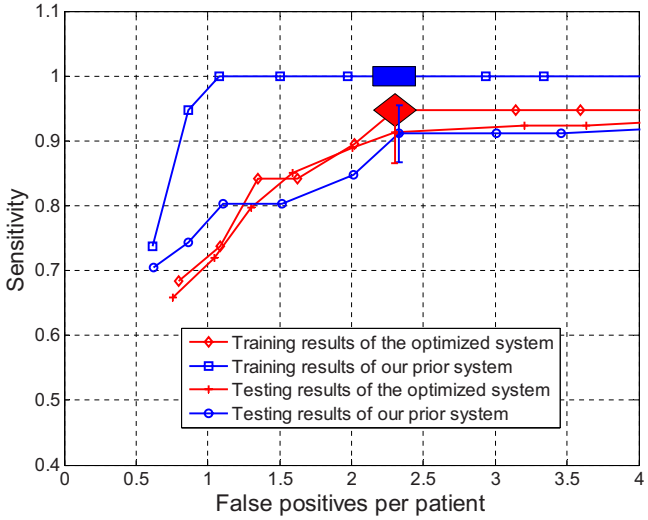
(a) 6-9 mm adenomas



(c) ≥ 8 mm adenomas



(b) ≥ 6 mm adenomas



(d) ≥ 10 mm adenomas

FIG. 4. Final FROC curves for the two systems on patients in the training and testing data sets. A filled square or diamond denotes the chosen classifier operation point based on the training FROC curve. Error bars on the testing results represent two standard deviations and are obtained at the same classifier operating points.

TABLE III. Performance comparisons for different polyp size categories at the chosen classifier operation.

Size Categories	System	Sensitivity ± STD	FPs ± STD[a]	Sensitivity improvement [CI], p
6–9 nm	Old	61.68 ± 2.37%	1.09 ± 0.26	
6–9 mm	Optimized	74.71 ± 2.39%	10.2 ± 0.22	13.03% [6.57%, 19.5%], $p=7.78\times10^{-5}$
⩾6 mm	Old	65.02% ± 2.24%	6.12 ± 1.39	
⩾6 mm	Optimized	77.40% ± 2.23%	5.83 ± 0.09	12.38% [6.23%, 18.53%] $p=7.95\times10^{-5}$
⩾8 nm	Old	82.20% ± 2.83%	7.02 ± 0.15	
⩾8 mm	Optimized	90.58% ± 2.72%	6.37 ± 0.13	8.38% [0.75%, 16%], $p=0.03$
⩾10 mm	Old	91.09% ± 4.35%	2.33 ± 0.05	
⩾10 mm	Optimized	91.28 ± 4.72%	2.30 ± 0.06	0.19% [−12.89%, 13.27%], $p=0.98$

[a]Abbreviations: "FPs" represents false positive rate, "CI" means 95% confidence interval, and "STD" denotes standard deviation.

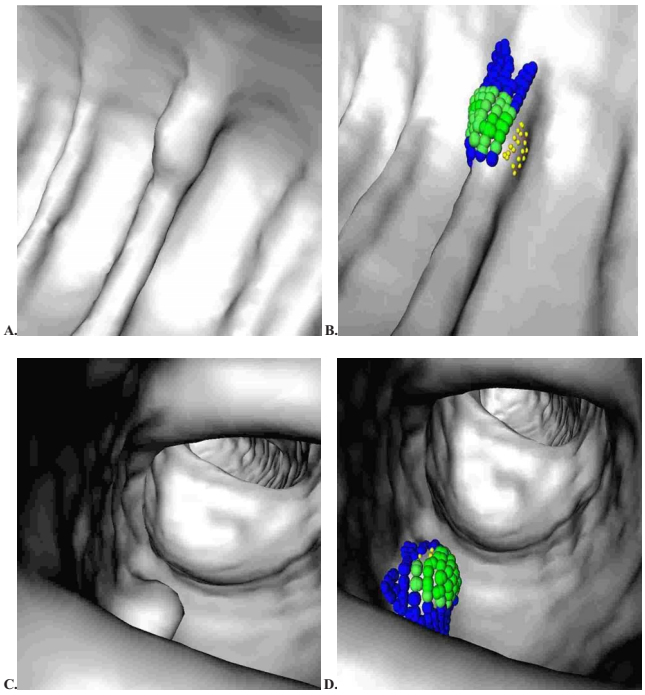


FIG. 5. Medium-sized polyps in the training data set detected by Pareto front optimized CTC CAD system but not by our prior system. Surface rendered 3D endoluminal images for two polyps (A), (C) without and (B), (D) with CTC CAD detections. (A), (B) 6 mm pedunculated adenoma in the sigmoid colon on supine CTC (72 y.o. female). (C), (D) 7 mm sessile adenoma in the descending colon on prone CTC (57 y.o. male). The dark gray dots in (B, D) are ground truth vertices on the colon surface. Detected vertices that match the ground truth are shown in light gray. Vertices clustered in the true detection but not marked as ground truth are shown as small, light gray dots.

without CAD (Refs. 39 and 40) and to that found in a meta-analysis of 33 CTC published clinical trials on 6393 patients,[40] in which the sensitivity of CTC was found to be 70% [CI, 55% to 84%] for polyps 6–9 mm in size. Radiologists computed sensitivities based on all polyps found by optical colonoscopy in the meta-analysis. Note that our sensitivities do not include 13.12% of the optical colonoscopy-confirmed polyps that were not retrospectively identifiable in our data set (11.17% for 6–7 mm polyps, 1.46% for 8–9 mm polyps, and 0.49% for polyps 10 mm or larger).

Performances for other polyp size categories except 1 cm or larger are also improved significantly. Sensitivity was not improved for detecting larger polyps, probably because our prior CTC CAD is already good enough for detecting those larger polyps. A large population study has shown that our system is comparable to a radiologist for detecting larger polyps.[16] Further optimizations like the one proposed in this article will not significantly improve its ability to identify larger polyps.

We do not report performance on hyperplastic polyps. There are 19 and 43 hyperplastic polyps sized 6 mm or larger in the training and the testing sets, respectively. While hyperplastic polyps may appear indistinguishable from adenomas on CT colonography, they have no malignant potential and consequently it is less important to detect them.

Our new method was based on a multiobjective genetic algorithm that minimizes the number of false positives and false negatives simultaneously by means of multiple nondominated solutions known as a Pareto front. The Pareto front showed points that are so-called nondominated solutions. For any false positive point along the x-axis in Fig. 2, there was no set of parameters that led the CAD system to have a lower false negative rate at that false positive setting. This technique greatly simplifies the difficult problem of selecting the many different thresholds.

In prior studies, multiobjective optimization algorithms had been applied to generate upper bounds of receiver operating characteristic (ROC) or the free-response ROC (FROC) curves on training data sets.[26,27] However, it was not clear if the ROC or FROC curves generated by the multiobjective optimization technique were still valid for testing data sets. To our knowledge, our study is the first attempt to optimize a CTC CAD system using the multiobjective algorithm and to test the optimized system on a large data set. Our results showed that the multiobjective optimization technique was a powerful tool to improve the detection accuracy, especially for "medium-sized" polyps.

We created FROC curves for three operation points on the Pareto front using the training data set. Operation point 2 on the Pareto front in Fig. 2 was chosen for the system based on its performance in terms of the training FROC curve. Though any one of these points along the Pareto front could be the settings for our system, we must make a trade-off between the computational complexity of generating a training FROC curve for one point on the Pareto front and the denser sampling on the Pareto front. Note that generating the training FROC curve needs to run the CTC CAD program on the training data, and the whole pipeline of feature selection and classifier training, which is computationally expensive.

Parameter values corresponding to the operation point 2 in Fig. 2 shows that the best initial clustering strategy for our data set might be a shrinking process, i.e., we first admit many vertices into an initial polyp candidate on the colon surface using "wider" thresholds; we then eliminate some of the vertices in the polyp candidate by applying "narrower" thresholds. In fact, among the 100 Pareto sets produced by the SPEA2 algorithm, there are only three Pareto sets that have a growing process in the initial clustering procedure. A growing clustering procedure means that we first admit few vertices into a polyp candidate using narrower thresholds and we then accept more neighbor vertices by applying wider thresholds. Those three Pareto sets missed 78, 83, and 183 true polyp detections, which of course are not good choices of operation points (too many false negatives). This fact suggests that a shrinking clustering strategy might be a better initial polyp candidate formation method on the colon surface for our data set.

Yoshida *et al.*[4] used a similar two-step strategy to generate initial polyp candidates on the segmented colon wall. Two geometric features utilized in their initial detection process are the volumetric shape index and curvedness. In the first step, narrower thresholds were set for the two features and vertices on the colon wall passing the threshold test were extracted as *seed regions*. Thresholds for the two features

were then relaxed so that more neighbor vertices were included into the seed regions, and the seed regions grew. However, it is not clear if the growing strategy is still a better method than a shrinking scheme for their data if they applied a similar optimization algorithm to the thresholds.

We did not compare the optimized system with other CTC CAD systems. CTC CAD results on different data sets are available.[4,6] However, those are difficult to compare with since either the data set was different or the polyp size range was divided in a different way.

Our optimized initial polyp detection algorithm is less complex than the prior sequential classifier. The number of threshold pairs was reduced from 15 to seven. The thresholds define feature value ranges on the mean and Gaussian curvature of the polyp candidate, the average sphericity of clusters, and the number of vertices in the clusters.

VII. CONCLUSION

We showed that the Pareto front can significantly improve the sensitivities of our CTC CAD system for detecting polyps of 6–9 mm, 6 mm or larger, 8 mm or larger size categories while still keeping the high sensitivity for detecting 10 mm or larger polyps. The Pareto front is a potentially powerful technique for optimizing CTC CAD systems.

ACKNOWLEDGMENTS

The authors thank Perry J. Pickhardt, M.D., J. Richard Choi, M.D., and William Schindler, D.O., for providing the CTC data. This research was supported by the Intramural Research Program of the NIH, Clinical Center. This study utilized the high-performance computational capabilities of the NIH Biowulf PC/Linux cluster. No FDA endorsement of any product or company mentioned in this manuscript should be inferred.

APPENDIX: REVIEW OF THE SPEA2 ALGORITHM (REF. 28)

Like other genetic algorithms, the SPEA2 algorithm usually contains four steps in the optimization procedure:

Initialization: A set of initial solutions is randomly generated in this step. Each solution is usually coded into a binary string. Besides the regular population P_t in the algorithm, there is an archive, $\bar{P}_t$, maintained that contains all the nondominated solutions from the previous generation. In other words, $\bar{P}_t$ keeps best solutions found in the previous generation. The size of $\bar{P}_t$ is kept the same as that of P_t. An individual in $\bar{P}_t$ is removed only if (1) a solution has been found in the current generation that dominates it or (2) the size of $\bar{P}_t$ exceeds P_t. If the size of $\bar{P}_t$ is less than P_t, other better dominated solutions are added into $\bar{P}_t$ to keep the archive size the same as that of the population.

Fitness evaluation: The fitness evaluations for the individuals are based on both the population and the archive, and a good individual is assigned a smaller fitness value in this article. First, each individual i in the P_t and $\bar{P}_t$ is assigned a strength value $S(i)$, the number of solutions it dominates,

$$S(i) = \|\{j | j \in P_t + \bar{P}_t \wedge i \succ j\}\|, \tag{A1}$$

where $\|\cdot\|$ represents the cardinality of a set, $+$ stands for multiset union, and the symbol $\succ$ corresponds to the Pareto dominance relation. Based on the value of $S(i)$, a raw fitness value, $R(i)$, is then given to the individual i,

$$R(i) = \sum_{j \in P_t + \bar{P}_t, j \succ i} S(j), \tag{A2}$$

which is the number of solutions that dominate i. The final fitness value is assigned by adding a density value. The density function value, $D(i)$, is estimated in objective space,

$$D(i) = \frac{1}{\delta_i^k + 2}, \tag{A3}$$

where δ_i^k denotes the kth nearest distance for the ith individual among P_t and $\bar{P}_t$ in objective space. k is usually set as $\sqrt{N+\bar{N}}$, where N is the population size and $\bar{N}$ the archive size. Finally, the fitness value for the ith individual is calculated as

$$F(i) = R(i) + D(i). \tag{A4}$$

From the definition above, a better solution will be assigned a smaller fitness value.

Environmental and mating selection: All nondominated individuals in $\bar{P}_t$ and P_t are kept in $\bar{P}_{t+1}$, and candidates in $\bar{P}_{t+1}$ are then selected to produce the next generation. In the standard genetic algorithm, the probability of an individual to be selected is proportional to its performance. A better individual is more likely to be selected. In SPEA2, all candidates are selected using a binary tournament selection scheme. In the binary tournament selection, it first randomly selects two individuals and only the better one survives. Ties are broken randomly.

Reproduction: The standard crossover and mutation methods are used to produce the next generation from the individuals selected in the environment selection procedure;[36] the produced next generation is then stored in $\bar{P}_{t+1}$.

All steps except the initialization are repeated till the specified generation number is reached.

[a]Partially presented at the 2007 European Congress of Radiology
[b]Author to whom correspondence should be addressed. Electronic mail: rms@nih.gov; Telephone: (301) 402-5486; Fax: (301) 451-5721

[1]A. Jemal, R. Siegel, E. Ward, T. Murray, J. Xu, and M. J. Thun, "Cancer statistics, 2007," Ca-Cancer J. Clin. **57**, 43–66 (2007).
[2]P. J. Pickhardt *et al.*, "Computed tomographic virtual colonoscopy to screen for colorectal neoplasia in asymptomatic adults," N. Engl. J. Med. **349**, 2191–2200 (2003).
[3]R. M. Summers *et al.*, "Automated polyp detector for CT colonography: feasibility study," Radiology **216**, 284–290 (2000).
[4]H. Yoshida and J. Nappi, "Three-dimensional computer-aided diagnosis scheme for detection of colonic polyps," IEEE Trans. Med. Imaging **20**, 1261–1274 (2001).
[5]D. S. Paik *et al.*, "Surface normal overlap: A computer-aided detection algorithm with application to colonic polyps and lung nodules in helical CT," IEEE Trans. Med. Imaging **23**, 661–675 (2004).

[6]Z. Wang *et al.*, "Reduction of false positives by internal features for polyp detection in CT-based virtual colonoscopy," Med. Phys. **32**, 3602–3616 (2005).
[7]H. Yoshida and A. H. Dachman, "CAD techniques, challenges, and controversies in computed tomographic colonography," Abdom. Imaging **30**, 26–41 (2005).
[8]K. Suzuki, H. Yoshida, J. Nappi, and A. H. Dachman, "Massive training Artificial Neural Network (MTANN) for reduction of false positives in computer-aided detection of polyps: Suppression of rectal tubes," Med. Phys. **33**, 3814–3824 (2006).
[9]K. Suzuki, H. Yoshida, J. Nappi, S. G. Armato III, and A. H. Dachman, "Mixture of expert 3D massive-training ANNs for reduction of multiple types of false positives in CAD for detection of polyps in CT colonography," Med. Phys. **35**, 694–703 (2008).
[10]J. Li *et al.*, "Wavelet method for CT colonography computer-aided polyp detection," Med. Phys. **35**, 3527–3538 (2008).
[11]A. K. Jerebko, R. M. Summers, J. D. Malley, M. Franaszek, and C. D. Johnson, "Computer-assisted detection of colonic polyps with CT colonography using neural networks and binary classification trees," Med. Phys. **30**, 52–60 (2003).
[12]A. K. Jerebko, J. D. Malley, M. Franaszek, and R. M. Summers, "Multiple neural network classification scheme for detection of colonic polyps in CT colonography data sets," Acad. Radiol. **10**, 154–160 (2003).
[13]H. Yoshida, Y. Masutani, P. MacEneaney, D. T. Rubin, and A. H. Dachman, "Computerized detection of colonic polyps at CT colonography on the basis of volumetric features: Pilot study," Radiology **222**, 327–336 (2002).
[14]A. K. Jerebko, J. D. Malley, M. Franaszek, and R. M. Summers, "Support vector machines committee classification method for computer-aided polyp detection in CT colonography," Acad. Radiol. **12**, 479–486 (2005).
[15]D. J. Shaffer, "Multiple objective optimization with vector evaluated genetic algorithms," in *Genetic Algorithms and Their Applications: Proceedings of the First International Conference of Genetic Algorithms* (Lawrence Erlbaum, Hillsdale, NJ, 1985), pp. 93–100.
[16]R. M. Summers *et al.*, "Computed tomographic virtual colonoscopy computer-aided polyp detection in a screening population," Gastroenterology **129**, 1832–1844 (2005).
[17]P. K. Sahoo, S. Soltani, K. C. Wong, and Y. C. Chen, "A survey of thresholding techniques," Comput. Vis. Graph. Image Process. **41**, 233–260 (1988).
[18]W. E. Lorensen and H. E. Cline, "Marching cubes: A high resolution 3D surface construction algorithm," in *Proceedings of the 14th Annual Conference on Computer Graphics and Interactive Techniques* (ACM Press, Anaheim, CA, 1987), pp. 163–169.
[19]J. Li, "Automatic colonic polyp detection using multi-objective evolutionary techniques," in *Image Processing*, Proc. SPIE **6144**, 1742–1750 (2006).
[20]J. Yao, M. Miller, M. Franaszek, and R. M. Summers, "Colonic polyp segmentation in CT colonography-based on fuzzy clustering and deformable models," IEEE Trans. Med. Imaging **23**, 1344–1352 (2004).
[21]J. Li, J. Yao, R. M. Summers, N. Petrick, M. T. Manry, and A. K. Hara, "An efficient feature selection algorithm for computer-aided polyp detection," Int. J. Artif. Intell. Tools **15**, 893–915 (2006).
[22]M. E. Zalis *et al.*, "CT colonography reporting and data system: a consensus proposal," Radiology **236**, 3–9 (2005).
[23]L. Butterly, M. Chase, H. Pohl, and G. S. Fiarman, "Prevalence of clinically important histology in small adenomas," Clin. Gastroenterol Hepatol **4**, 343–348 (2006).
[24]L. T. Bui, D. Esam, H. A. Abbass, and D. Green, "Performance analysis of evolutionary multi-objective optimization methods in noisy environments," in *Proceeding of the 8th Asian Pacific Symposium on Intelligent and Evolutionary Systems*, Cairns, Australia, 2004, pp. 29–39.
[25]C. A. C. Coello, "An updated survey of evolutionary multiobjective optimization techniques: State of the art and future trends," in *Congress on Evolutionary Computation*, Washington, D.C., 1999, pp. 3–13.
[26]M. A. Anastasio, M. A. Kupinski, and R. M. Nishikawa, "Optimization and FROC analysis of rule-based detection schemes using a multiobjective approach," IEEE Trans. Med. Imaging **17**, 1089–1093 (1998).
[27]M. A. Kupinski and M. A. Anastasio, "Multiobjective genetic optimization of diagnostic classifiers with implications for generating receiver operating characteristic curves," IEEE Trans. Med. Imaging **18**, 675–685 (1999).
[28]E. Zitzler, M. Laumanns, and L. Thiele, "SPEA2: Improving the strength Pareto evolutionary algorithm," in *Swiss Federal Institute of Technology*, 2001.
[29]E. Zitzler, L. Thiele, M. Laumanns, C. M. Fonseca, and V. G. Fonseca, "Performance assessment of multiobjective optimizers: An analysis and review," Evol. Comput. **7**, 117–132 (2003).
[30]J. Li *et al.*, "Validating Pareto optimal operation parameters of polyp detection algorithms for CT colonography," Proc. SPIE **6514**, 65142G (2007).
[31]A. Huang, J. Li, R. M. Summers, N. Petrick, and A. K. Hara, "Using Pareto fronts to evaluate polyp detection algorithms for CT colonography," Proc. SPIE **6514**, 651407 (2007).
[32]A. Huang and R. M. Summers, "Surface curvature estimation for automatic colonic polyp detector," Proc. SPIE **5746**, 393–402 (2005).
[33]R. M. Summers, L. M. Pusanik, and J. D. Malley, "Automatic detection of endobronchial lesions with virtual bronchoscopy: Comparison of two methods," Proc. SPIE **3338**, 327–335 (1998).
[34]J. Yao and R. M. Summers, "3D colonic polyp segmentation using dynamic deformable surfaces," Proc. SPIE **5369**, 280–289 (2004).
[35]D. A. van Veldhuizen and G. B. Lamont, "Multiobjective Evolutionary Algorithm Test Suites," in *Proceedings of the 1999 ACM Symposium on Applied Computing*, San Antonio, TX, 1999, pp. 351–357.
[36]Z. Michalewicz, *Genetic Algorithm+Data Structures=Evolution Programs* (Springer-Verlag, Berlin, 1996).
[37]C. A. C. Coello, *A Survey of Constraint Handling Techniques used with Evolutionary Algorithms* (Laboratorio Nacional de Informatica Avanzada, Xalapa, Veracruz, Mexico, 1999).
[38]B. Efron and R. Tibshirani, *An Introduction to the Bootstrap* (Chapman & Hall, New York, 1993).
[39]S. Halligan *et al.*, "CT colonography in the detection of colorectal polyps and cancer: Systematic review, meta-analysis, and proposed minimum data set for study level reporting," Radiology **237**, 893–904 (2005).
[40]B. P. Mulhall, G. R. Veerappan, and J. L. Jackson, "Meta-analysis: computed tomographic colonography," Ann. Intern Med. **142**, 635–650 (2005).