2006

Automatic Colonic Polyp Detection Using Multiobjective Evolutionary Techniques

Jiang Li
National Institutes of Health, jli@odu.edu

Adam Huang
National Institutes of Health

Jianhua Yao
National Institutes of Health

Ingmar Bitter
National Institutes of Health

Nicholas Petrick
Food and Drug Administration

See next page for additional authors

Original Publication Citation

Li, J., Huang, A., Yao, J., Bitter, I., Patrick, N., Summers, R. M., Pickhardt, P. J., & Cho, J. R. (2006) Automatic colonic polyp detection using multiobjective evolutionary techniques. In J.M. Reinhardt & J.P.W. Pluim (Eds.), *Medical Imaging 2006: Image Processing, Proceedings of SPIE Volume 6144* (61445E). SPIE. https://doi.org/10.1117/12.653546

Automatic Colonic Polyp Detection Using Multiobjective Evolutionary Techniques

Jiang Li[a], Adam Huang[a], Jianhua Yao[a], Ingmar Bitter[a], Nicholas Petrick[b]
Ronald M. Summers[a], Perry J. Pickhardt[c,d] and J. Richard. Choi[c,e]

[a]Diagnostic Radiology Department, Warren G. Magnuson Clinical Center, NIH
[b]NIBIB/CDRH, Joint Laboratory for the Assessment of Medical Imaging System, FDA
[c]Uniformed Services University of the Health Sciences, Bethesda, Maryland
[d]National Naval Medical Center, Bethesda, Maryland
[e]Walter Reed Army Medical Center, Washington, DC

ABSTRACT

Colonic polyps appear like elliptical protrusions on the inner wall of the colon. Curvature based features for colonic polyp detection have proved to be successful in several computer-aided diagnostic CT colonography (CTC) systems. Some simple thresholds are set for those features for creating initial polyp candidates, sophisticated classification scheme are then applied on these polyp candidates to reduce false positives. There are two objective functions, the number of missed polyps and false positive rate, that need to be minimized when setting those thresholds. These two objectives conflict and it is usually difficult to optimize them both by a gradient search. In this paper, we utilized a multiobjective evolutionary method, the Strength Pareto Evolutionary Algorithm (SPEA2),[1] to optimize those thresholds. SPEA2 incorporates the concept of Pareto dominance and applies genetic techniques to evolve individual solutions to the Pareto front. The SPEA2 algorithm was applied to colon CT images from 27 patients each having a prone and a supine scan. There are 40 colonoscopically confirmed polyps resulting in 72 positive detections in CTC reading. The results obtained by SPEA2 were compared with those obtained by our old system, where an appropriate value was set for each of those thresholds by a histogram examination method. If we keep the sensitivity the same as that of our old system, the SPEA2 algorithm reduced false positive rate by 76.4% from average false positive 55.6 to 13.3 per data set. If the false positive rate is kept the same for both systems, SPEA2 increased the sensitivity by 13.1% from 53 to 61 among 72 ground truth detections.

Keywords: Computer-Aided Detection, Pattern recognition, Statistical methods, Multiobjective Evolution, Genetic algorithm

1. INTRODUCTION

Colon cancer is the second leading cause of cancer deaths in the US.[2] It is known that colorectal cancer can be prevented if a screening procedure is done and any colon polyps found are removed. Virtual colonoscopy, also known as computed tomographic colongraphy (CTC), has been studied as a screening procedure for the past 10 years and shows promising results. Computer-aided polyp detection (CAD) CTC systems have been under investigation by a number of institutes to improve the performance of CTC. In order for the CTC CAD system to prevail, it must be both sensitive and specific. Summers et al. describe recent work on a version of computer automated polyp detection that uses geometric and volumetric features, acquired from the CT data, as the basis for polyp detection.[3,4]

In the CTC procedure, hundreds of CT scans are taken for a patient both from supine and prone views. The CTC CAD system under development by our group[5] takes the following four ordered steps to detect colon polyps.

Further author information: (Send correspondence to Ronald M. Summers.)
Ronald M. Summers: E-mail: rms@nih.gov, Tel. 1 301 402-5486
Website: http://www.cc.nih.gov/drd/summers.html
Jiang Li: E-mail: lij3@cc.nih.gov, Tel. 1 301 451-8363

Medical Imaging 2006: Image Processing, edited by Joseph M. Reinhardt, Josien P. W. Pluim,
Proc. of SPIE Vol. 6144, 61445E, (2006) · 0277-786X/06/$15 · doi: 10.1117/12.653546

After a computer reads all CT images, the CAD program first segments the colon surface by a region growing method. The CAD program then identifies initial polyp candidates along the colon surface based on curvature and geometry information. The third step is to determine the 3D segmentation of each surface detection polyp candidate in the 3D CT volume and to calculate quantitative features for the polyps. Finally, a decision of true polyp or false positive is made by a classifier based on the features that are proved clinically relevant. Previous efforts were more focused on optimization classification schemes such as neural network, decision tree or support vector machine, while little work was put on how to optimize the initial polyp detections due to lack of closed-form solutions. In the initial polyp detection, some simple thresholds are set for features such as curvature and size of lesion to cluster vertices on the colon surface as an initial detection. These thresholds were either empirically chosen or derived from mathematical modelling.[6,7] In this paper, we investigate a multiobjective optimization algorithm for setting the optimal thresholds for the initial polyp detection task.

Setting optimal thresholds for curvature-based features to create polyp candidates on the colon surface is characterized by the presence of two conflicting objectives: minimizing both the number of missed polyps and the false positive rate. Both objective functions are difficult, and perhaps impossible, to be expressed in closed form. A practical solution is to look at this parameter setting problem as a multiobjective evolutionary problem. We used the SPEA2 algorithm[1] to find the Pareto optimal set, which is a set of non-inferior or admissible solutions for the problem. This optimal set provides more freedom for a decision maker when making trade-offs between the two objectives. In the following sections, we first describe several curvature features used in the initial detection procedure. We then describe the goal and some technical details associated with the goal. Finally, the proposed method utilizing SPEA2 algorithm is outlined and results for optimization of colonic polyp detection are presented.

2. METHOD

In this section, we describe the features used in the detection procedure and how they are used, we then describe the proposed optimization algorithm.

2.1. Clustering Polyp Candidates

Colonic polyps appear as elliptical protrusions on the inner wall of the colon. Colonic polyps can be characterized by surface curvatures. 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 normal curvatures along the principal tangent directions. These principal curvatures are calculated by a kernel method, and polyps can be identified as regions with negative k_1 and k_2.[8]

After the two principal curvatures are calculated, the mean curvature H and the Gaussian curvature K are defined for each vertex on the colon surface as,

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

For a given CT scan, a region growing method was used to segment the colon and an iso-surface was created at its border. The mean curvature H and Gaussian curvature K were then calculated for each vertex on the surface using a kernel method.[8] Those vertexes will be included into a cluster if

$$x_1 < H < x_2 \text{ and } x_3 < K < x_4 \tag{2}$$

where $x_i, i = 1, \ldots, 4$ are preset thresholds to be optimized. After all vertices are examined, two additional features, mean sphericity S_m and number of vertices N, were calculated for each formed cluster,

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

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

$$x_5 < S_m < x_6 \text{ and } N > x_7 \tag{4}$$

where $x_5, \ldots, x_7$ are again thresholds to be optimized, the cluster will be counted as an polyp candidate and be delivered to the next step for further processing (polyp segmentation).[9]

2.2. Problem Formulation

Let $f_1(\overrightarrow{x}), f_2(\overrightarrow{x})$ denote the number of missed polyps and average number of false positives per data set, respectively, where $\overrightarrow{x}$ is a 7-dimensional threshold vector

$$\overrightarrow{x} = \{x_1, \ldots, x_7 | x_1, \ldots, x_6 \in R, x_7 \in I\} \tag{5}$$

Our multiobjective minimization problem (MOP) can be stated as follows,

$$\min_{\overrightarrow{x}} \overrightarrow{f}(\overrightarrow{x}) = \{f_1(\overrightarrow{x}), f_2(\overrightarrow{x})\} \tag{6}$$

subject to

$$\begin{cases} x_1 \in [-10, 0), \\ x_2 \in [-10, 0), \\ x_1 < x_2, \\ x_3 \in [0, 50], \\ x_4 \in [0, 100], \\ x_3 < x_4, \\ x_5 = 0, \\ x_6 \in (0, 2], \\ x_7 \in [6, 30]. \end{cases} \tag{7}$$

The range for the each threshold was determined experimentally such that it is wide enough to include all solutions of interest. The global optima of an MOP is the Pareto front determined by evaluating each member of the Pareto optimal solution set.[10] The Pareto optimal set consists of solutions that are not dominated by any other solutions. A solution $\overrightarrow{x}_1$ is said to dominate ($\succ$) $\overrightarrow{x}_2$ if objective vector $\overrightarrow{f}(\overrightarrow{x}_1)$ is less than or equal to $\overrightarrow{f}(\overrightarrow{x}_2)$ in all attributes, and strictly less than in at least one attribute,

$$\begin{cases} \overrightarrow{x}_1 \succ \overrightarrow{x}_2, \ iff \\ \forall i \in \{1, 2\} : f_i(\overrightarrow{x}_1) \leq f_i(\overrightarrow{x}_2) \wedge \exists j \in \{1, 2\} : f_j(\overrightarrow{x}_1) < f_j(\overrightarrow{x}_2) \end{cases} \tag{8}$$

The space formed by the objective vectors of Pareto optimal solutions is called the Pareto front. 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(s) satisfying the objectives as best as possible. In this study, we utilized the SPEA2 to obtain the Pareto optimal set for our thresholds.

2.3. Algorithm Background

The study of multiobjective optimization began in 1985.[11] Subsequently, many multiobjective optimization algorithms have been proposed in the literature.[12–15] In We are interested in the SPEA2 algorithm because of its fast convergence rate.[16] Speed is important for our task due to the expensive computation on the colon surface. In this section we describe the main parts of the optimization procedure. For details of SPEA2 and background about genetic algorithms, are given in.[1,17]

In a standard genetic algorithm, there are usually four steps in the optimization procedure: 1) randomly initialize the solution population, 2) evaluate and assign a fitness value for each individual in the population according to its performance, 3) select individuals based on their performances so that better individuals are more likely to be selected for producing the next generation and 4) 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 and in the SPEA2 algorithm, a smaller fitness value indicates a better performance. Step 2 through step 4 are repeated until a specified generation number is reached. The SPEA2 algorithm differs from the standard genetic algorithm in the following three aspects,[1]

Environmental Selection: Besides the regular population in a genetic algorithm, there is an archive maintained to contain all the nondominated solutions from the previous generation. An individual in the archive is removed only if 1) a solution has been found in the current generation that dominates it or 2) the maximum archive size is exceeded and its performance is worse than that of other solutions in the archive.

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

$$S(i) = ||\{j | j \in P_t + \bar{P}_t \wedge i \succ j\}|| \tag{9}$$

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 given to the individual i,

$$R(i) = \sum_{j \in P_t + \bar{P}_t, j \succ i} S(j) \tag{10}$$

The final fitness value is assigned by adding a density value. The density function value, $D(i)$, is estimated in objective space,

$$D(i) = \frac{1}{\delta_i^k + 2} \tag{11}$$

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,

$$F(i) = R(i) + D(i). \tag{12}$$

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

Mating Selection: In the standard genetic algorithm, the probability of an individual to be selected for producing the next generation is proportional to its performance. In SPEA2, all candidates are selected from the archive using a binary tournament selection scheme. In the binary tournament selection, we first randomly select two individuals and only the better one survives.

2.4. Algorithm Outline

We now outline the algorithm used for optimizing thresholds in the initial polyp clustering procedure.

1. **Initialization**: Set $N = 100$ (population size),
   $\bar{N} = 100$ (archive size),
   $T = 200$ (generation number),
   Randomly initialize P_t with sets of thresholds and generate an empty archive $\bar{P}_t$. All individuals are coded into 8-bit binary strings.

2. **Fitness evaluation**: In the evaluation procedure, a separate clustering program takes each individual set of thresholds in P_t and $\bar{P}_t$ and run the clustering algorithm on each colon surface, it then returns the number of missed true detections and the average false positive rate. The fitness value for each individual is then calculated using the method described above.

3. **Termination check**: If $t > T$ 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 to $\bar{N}$.

5. **Mating selection**: Select 100 individuals in $\bar{P}_t$ with replacement using the binary tournament procedure.

6. **Reproduce**: Reproduce the next generation using the standard crossover and mutation procedures. The crossover and mutation probability were set as 0.9 and 0.01 in our experiments. Store the results in P_{t+1}. Set $t = t + 1$ and go to step 2.

In the evaluation process, it is possible that some of the individuals are not feasible due to the constraints in equation (7). We used the simplest approach to handle the constraints. An individual not meeting the constraints was rejected until one feasible solution was produced. See[17, 18] for other strategies to handle constraints.

2.5. Data Acquisition

The CTC procedure was performed on 27 patients with a high suspicion of colonic polyps. All patients had at least one polyp and each polyp was verified by follow-up optical colonoscopy. These patients were chosen from a larger cohort who underwent a screening CTC procedure. Selection criteria included patients having at least 1 polyp > 5 mm. The majority of the polyps were identified on both the prone and supine views. There were 40 colonoscopically confirmed polyps (32 viewable in both scans, 3 in the supine, and 5 in the prone only) resulting in 72 positive detections in CTC reading. Fig. 1 shows one colon surface produced by the CTC CAD system.

3. RESULTS AND DISCUSSION

Fig. 2 shows the results obtained by SPEA2 and by our previous system. It is clear that the Pareto front provided a set of solutions for our initial polyp detection while our previous system only resulted in one solution which was dominated by the optimized Pareto front solutions. If we keep the sensitivity the same for both systems (the number of missed true positive detections is 19), the SPEA2 algorithm reduced false positives by 76.4% from average 55.6 to 13.3 per data set. This reduction will significantly mitigate the computation load for further processing. If the false positive is kept the same for both systems (around 56), SPEA2 increased the sensitivity by 13.1% from 53 to 61 among 72 viewable detections.

Fig. 3 shows some polyp examples produced by two different sets of thresholds for the initial polyp detection on colon surface. One threshold set is from our previous system which was determined based upon histogram examination. This old set of thresholds missed 19 true detections with an average false positive rate of 55.6 per data set. Another set of thresholds is chosen from the optimal set of solutions produced by the SPEA2 algorithm. The chosen optimized thresholds missed 11 true detections with a false positive rate of 58.1 per data set. With a similar false positive rate, the optimized thresholds detected 8 more true detections than that of the previous thresholds. We show in Fig. 3 some true detections missed by our previous threshold setting (in the first column) but picked by the optimized thresholds (in the second column). The blue dots in the figure are ground truth for the polyp on the surface, where each dot represents one vertex on the surface. The green dots represent detected ground truth vertices on the surface. Note that it is not necessarily to detect all ground truth vertices for a polyp, any detected portion of a polyp leads to a successful detection.

Fig. 4 shows some false positive detections produced by CTC CAD system with another two sets of thresholds. Again, one set is from our previous system (Fig. 4 a) and another set is chosen from the optimal sets produced by the SPEA2 algorithm (Fig. 4 b). The chosen optimized thresholds missed the same number of true detections as that of our previous setting (missed 19 true detections) but with a much lower false positive rate of 13.3 per date set. Recall that our previous thresholds setting had an average false positive rate of 55.6 per data set. The red dots in the image are false detection vertices on the colon surface. There are three false positive detections in the shown area using our previous thresholds setting, however, all the three false detections disappear if we used the optimized thresholds setting. It is clear that though the two sets of thresholds produced the same number of false negatives, the optimized thresholds produced much fewer false positives.

4. CONCLUSION

We have implemented a multiobjective evolutionary algorithm, SPEA2, for optimizing a CAD polyp detection system. This is the first introduction of such algorithms into the CTC CAD research field. The SPEA2 algorithm showed not only the potential for optimizing our CAD system but also the capability of providing a Pareto optimal set, which gives a decision maker more freedom and confidence to make trade-offs between false negative and false positive rates. If we kept the sensitivity the same as that of our old system, the SPEA2 algorithm generated a new set of thresholds that reduced the false positive rate by 76.4%. If the false positive rate is kept the same as that of our old system, we could choose another set of thresholds optimized by the SPEA2 algorithm which increased the sensitivity by 13.1%.

5. ACKNOWLEDGEMENTS

This research was supported by the Intramural Research Program of the NIH, Warren G. Magnuson Clinical Center. This study utilized the high-performance computational capabilities of the NIH Biowulf PC/Linux cluster.

REFERENCES

1. E. Zitzler, M. Laumanns, and L. Thiele, "Spea2: Improving the strength pareto evolutionary algorithm," technical report, Swiss Federal Institute of Technology, 2001.
2. A. Jemal, R. C. Tiwari, T. Murray, A. Ghafoor, A. Samuels, E. Ward, E. J. Feuer, and M. J. Thun, "Cancer statistics," *CA Cancer J Clin* **54**, pp. 8–29, 2004.
3. R. M. Summers, A. Jerebko, M. Franaszek, J. Malley, and C. Johnson, "Colonic polyps: Complementary role of computer-aided detection in ct colonography," *Radiology* **225**, pp. 391–399, 2002.
4. R. M. Summers, "Challenges for computer-aided diagnosis for ct colongraphy," *Abdominal Imaging* **27**, pp. 268–274, 2002.
5. R. M. Summers, J. Yao, P. J. Pickhardt, M. Franaszek, I. Bitter, D. Brickman, V. Krishna, and J. R. Choi, "Computed tomographic virtual colonoscopy computer-aided polyp detection in a screening population," *Gastroenterology* **129**, pp. 1832–1844, 2005.
6. H. Yoshida and J. Nappi, "Three-dimensional computer-aided diagnosis scheme for detection of colonic polyps," *IEEE Trans on Medical Imaging* **20**(12), pp. 1261–1274, 2001.
7. D. S. Paik, C. F. Beaulieu, and G. D. R. et.al, "Surface normal overlap: A computer-aided detection algorithm with application to colonic polyps and lung nodules in helical ct," *IEEE Trans on Medical Imaging* **23**(6), pp. 661–675, 2004.
8. A. Huang, R. M. Summers, and A. K. Hara, "Surface curvature estimation for automatic colonic polyp detection," *SPIE Medical Imaging* **5746**, pp. 393–402, 2005.
9. J. Yao and R. Summers, "3d colonic polyp segmentation using dynamic deformable surfaces," in *SPIE Medical Imaging*, 2004.
10. D. A. V. Veldhuizen and G. B. Lamont, "Multiobjective evolutionary algorithm test suites," in *Proceedings of the 1999 ACM Symposium on Applied Computing*, pp. 351–357, (San Antonio, Texas), 1999.
11. J. D. Schaffer, "Multiple objective optimization with vector evaluated genetic algorithms," in *Genetic Algorithms and their Applications: Proceedings of the First International Conference on Genetic Algorithms*, pp. 93–100, (Lawrence Erlbaum), 1985.
12. E. Zitzler, L. Thiele, M. Laumanns, C. M. Fonseca, and V. G. da Fonseca, "Performance assessment of multiobjective optimizers: An analysis and review," *IEEE Trans on Evolutionary Computation* **7**(2), pp. 117–132, 2003.
13. C. A. C. Coello, "An updated survey of evolutionary multiobjective optimization techniques : State of the art and future trends," in *Congress on Evolutionary Computation*, pp. 3–13, (Washington, D.C), 1999.
14. M. A. Anastasio, M. A. Kupinski, and robert M. Nishikawa, "Optimization and froc analysis of rule-based detection schemes using a multiobjective approach," *IEEE Transactions on Medical Imaging* **17**(6), pp. 1089–1093, 1998.

15. M. A. Kupinski and M. A. Anastasio, "Multiobjective genetic optimization of diagnostic classifiers with implications for generating receiver operating characteristic curves," *IEEE Transactions on Medical Imaging* **18**(8), pp. 675–685, 1999.
16. L. T. Bui, D. Essam, H. A. Abbass, and D. Green, "Performance analysis of evolutionary multi-objective optimization methods in noisy environments," in *Proceedings of The 8th Asia Pacific Symposium on Intelligent and Evolutionary Systems*, pp. 29–39, (Cairns, Australia), 2004.
17. Z. Michalewicz, *Genetic Algorithms + Data Structures = Evolution Programs*, Springer-Verlag, Berlin, Heidelberg, NY, third ed., 1999.
18. C. A. C. Coello, "A survey of constraint handling techniques used with evolutionary algorithms," technical report lania-ri-99-04, Laboratorio Nacional de Informtica Avanzada, Xalapa, Veracruz, Mexico, 1999.

Figure 1. Colon surface generated by the CTC CAD system

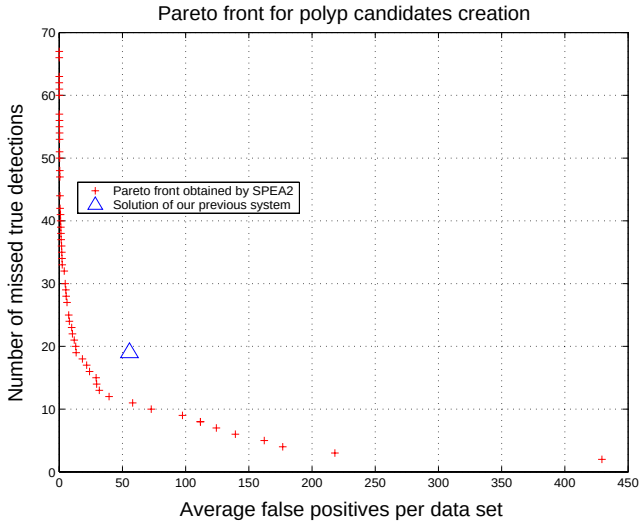
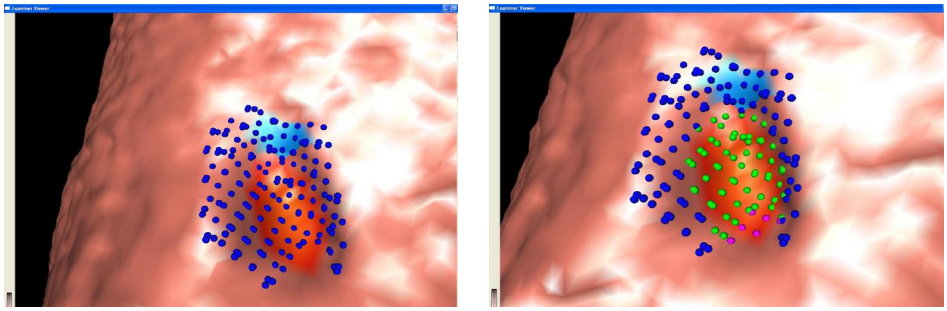
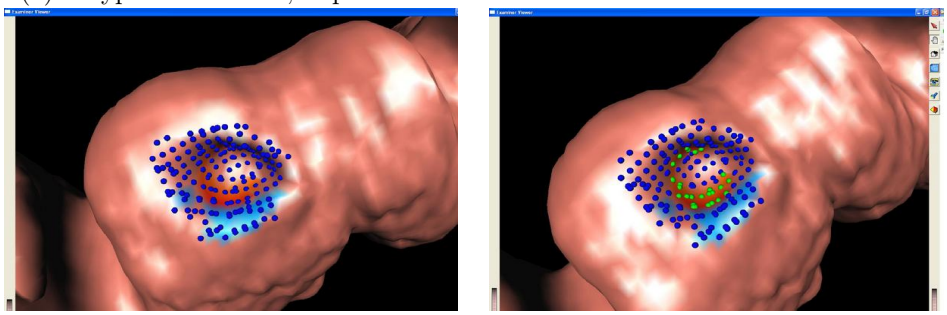


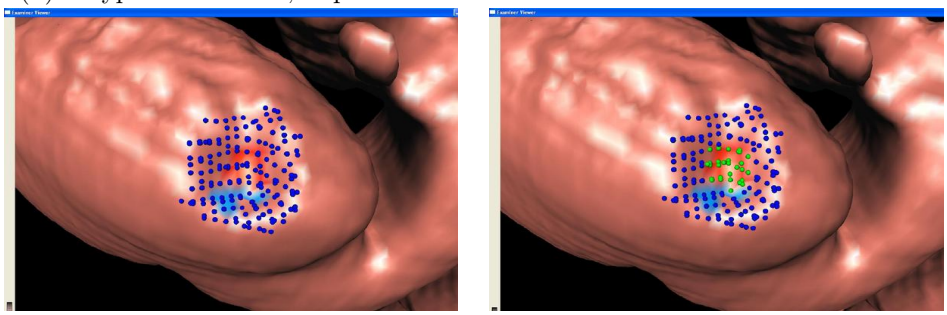
Figure 2. The solution provided by our old system and the Pareto front from the SPEA2 algorithm



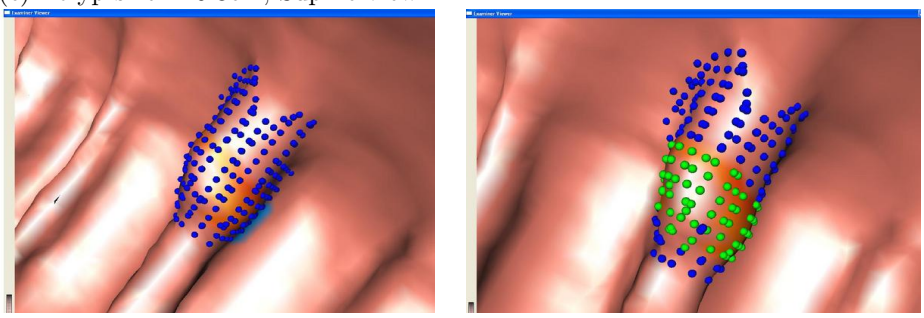
(a) Polyp size = 0.8cm, Supine view



(b) Polyp size = 0.6cm, Supine view

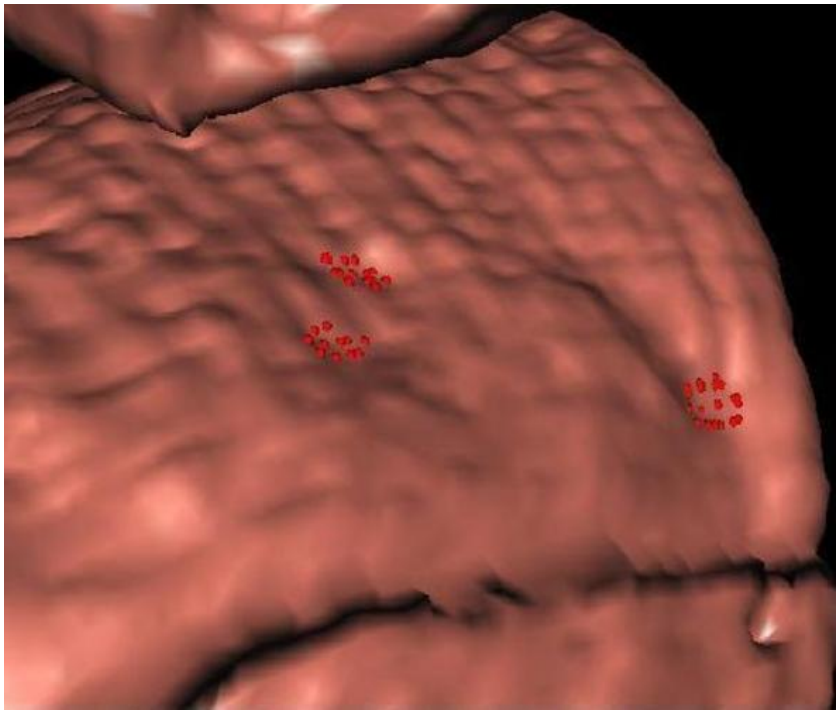


(c) Polyp size = 0.8cm, Supine view

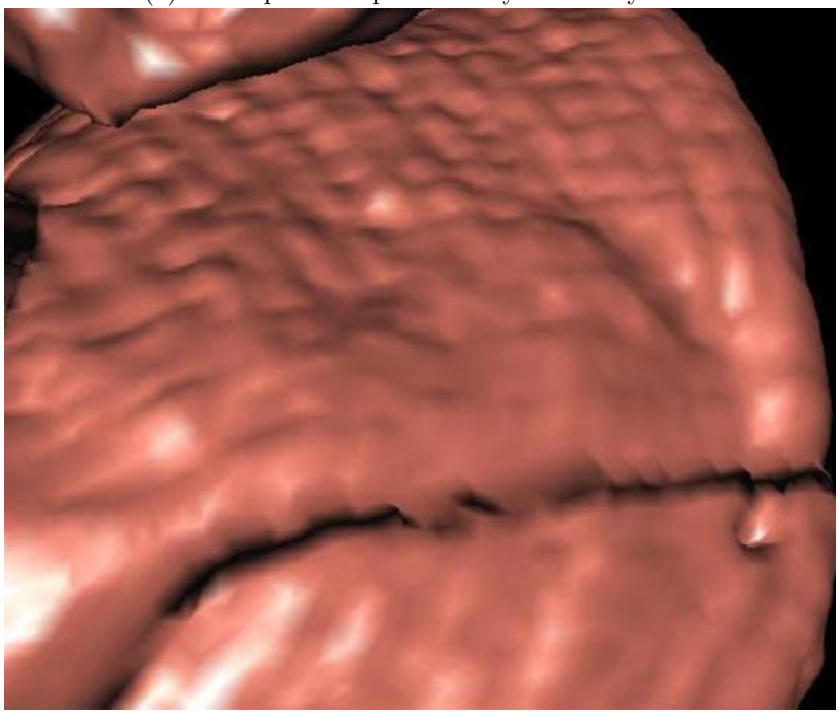


(d) Polyp size = 0.7cm, Prone view

Figure 3. Polyp detection samples comparison. The first column consists of missed true detection examples by our old system and the second column represents the same detections picked by one set of optimized thresholds, which is chosen from the Pareto optimal set. The old system missed 19 true detections while the optimized thresholds had 11 false negatives, both systems produced approximately the same amount of false positives per data set (55.6 for our old system and 58 for the optimized thresholds). The blue dots in the images are ground truth vertices on the colon surface. If the detected vertices match the ground truth, those vertices will be colored as green. Note that the blue dots appear darker than green ones if they are printed in a black and white image.



(a). False positives produced by our old system



(b). The same portion of colon produced by the optimized system

Figure 4. False positives comparison, the upper image is the output of our old system and the lower image is from one using SPEA2-optimized filter thresholds. The new method produced 13.3 false positives per data set and the same number of false negatives (19) as the old system. The red dots in the image denote the falsely detected vertices on the colon surface.