2007

Validating Pareto Optimal Operation Parameters of Polyp Detection Algorithms for CT Colonography

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

Adam Huang
National Institutes of Health

Nicholas Petrick
U.S. Food and Drug Administration

Jianhua Yao
National Institutes of Health

Ronald M. Summers
National Institutes of Health

See next page for additional authors

Original Publication Citation

Li, J., Huang, A., Petrick, N., Yao, J., & Summers, R. (2007). Validating Pareto optimal operation parameters of polyp detection algorithms for CT colonography. In M.L. Giger & N. Karssemeijer (Eds.), *Medical Imaging 2007: Computer-Aided Diagnosis, Proceedings of SPIE Vol. 6514* (65142G). SPIE of Bellingham, WA. https://doi.org/10.1117/12.707965

Validating Pareto Optimal Operation Parameters of Polyp Detection Algorithms for CT Colonography

Jiang Li[a], Adam Huang[a], Nicholas Petrick[b], Jianhua Yao[a] and Ronald M. Summers[a]

[a]Diagnostic Radiology Department, Clinical Center, NIH
[b]NIBIB/CDRH, Joint Laboratory for the Assessment of Medical Imaging System, FDA

ABSTRACT

We evaluated a Pareto front-based multiobjective evolutionary algorithm for optimizing our CT colonography (CTC) computer-aided detection (CAD) system. The system identifies colonic polyps based on curvature and volumetric based features, where a set of thresholds for these features was optimized by an evolutionary algorithm. We utilized a two-fold cross-validation (CV) method to test if the optimized thresholds can be generalized to new data sets. We performed the CV method on 133 patients; each patient had a prone and a supine scan. There were 103 colonoscopically confirmed polyps resulting in 188 positive detections in CTC reading from either the prone or the supine scan or both. In the two-fold CV, we randomly divided the 133 patients into two cohorts. Each cohort was used to obtain the Pareto front by a multiobjective genetic algorithm, where a set of optimized thresholds was applied on the test cohort to get test results. This process was repeated twice so that each cohort was used in the training and testing process once. We averaged the two training Pareto fronts as our final training Pareto front and averaged the test results from the two runs in the CV as our final test results. Our experiments demonstrated that the averaged testing results were close to the mean Pareto front determined from the training process. We conclude that the Pareto front-based algorithm appears to be generalizable to new test data.

Keywords: Computer-aided detection, Pattern recognition, Statistical methods, Multiobjective evolution, Genetic algorithm

1. INTRODUCTION

Colon cancer is the second leading cause of cancer death in the U.S.[1] A number of CT colonography (CTC) computer-aided detection (CAD) systems for the early detection of colonic polyps are under development.[2–7] Successful detection of polyps will increase the possibility of appropriate intervention for these polyps. One of the challenges for polyp detection is to accurately identify polyp candidates on the colon surface. We have developed a Pareto front-based multiobjective evolutionary algorithm for the initial polyp detection, where excellent results were achieved from our training data sets.[8] The purpose of this paper is to evaluate if the optimized results produced by the Pareto front-based algorithm using training data sets can be generalized to unseen data.

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[2] takes the following four ordered steps to detect colon polyps. 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 candidates. 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 optimizing 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 on the colon surface due to lack of closed-form solutions. In the initial polyp detection, some simple thresholds are set for

Send correspondence to Ronald M. Summers: E-mail: rms@nih.gov, Tel. 1 301 402-5486
Website: http://www.cc.nih.gov/drd/summers.html
Jiang Li is now with Electrical and Computer Engineering Department and VMASC, Old Dominion University.
E-mail: JLi@odu.edu, Tel. 1 757 683-6748

Medical Imaging 2007: Computer-Aided Diagnosis, edited by Maryellen L. Giger, Nico Karssemeijer,
Proc. of SPIE Vol. 6514, 65142G, (2007) · 1605-7422/07/$18 · doi: 10.1117/12.707965

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 modeling.[5,6] In our previous paper,[8] we presented a multiobjective optimization algorithm for setting the optimal thresholds and showed good results on training data sets. In this paper, we present validation results on testing data sets for optimized thresholds.

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 forms. A practical solution is to look at this parameter setting problem as a multiobjective evolutionary problem. We used the improved strength Pareto evolutionary algorithm (SPEA2)[9] 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 and a sequential classifier to create initial polyp candidates based on the features. We then present the problem formation for optimizing the sequential classifier. Finally, validation results of the optimized sequential classifier are given.

2. METHOD

In this section, we first describe an initial polyp clustering method which uses a sequential classifier. We then briey describe how to optimize and validate the 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.[10]

We calculate the mean curvature H and Gaussian curvature K for each vertex on the surface using a kernel method.[10] To form polyp candidates, we first check vertices on the surface if the following criteria are satisfied,

$$x_1 < H < x_2 \text{ and } x_3 < K < x_4 \tag{1}$$

where $x_i, i = 1, \ldots, 4$ are preset thresholds to be optimized. Vertices that meet the above conditions and share an edge 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,

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

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

$$x_5 < S_m < x_6 \text{ and } N > x_7 \tag{3}$$

where $x_5, \ldots, x_7$ are again thresholds, the polyp candidate is kept and delivered to the next step.

The next step is to examine the polyp candidate by checking how many vertices in the candidate and the vertices' neighbors, which might be outside of the candidate, satisfy the conditions,

$$x_8 < H < x_9, x_{10} < K < x_{11} \text{ and } x_{12} < S < x_{13} \tag{4}$$

where again, $x_8, \ldots, x_{13}$ are preset thresholds, and

$$S = 2 \cdot \frac{k_2 - k_1}{k_2 + k_1} \tag{5}$$

is sphericity for a single vertex. In this step, the polyp candidate can grow or shrink depending upon if the set of thresholds $x_8 \dots x_{13}$ are "wider" or "narrower" than those in equation (1) and (3). If $x_8 \dots x_{13}$ are "wider" than those in equation (1) and (3), the classifier will let more vertices be included in the polyp candidate such that the polyp candidate will grow. On the other hand, if a set of "narrower" thresholds are set, only part of the vertices in the polyp candidate can satisfy the criteria in equation (4). Finally, if N_b, the number of vertices satisfying (4), is greater than x_{14}, where x_{14} is another threshold, the candidate is delivered to the polyp segmentation procedure.[11]

2.2. Optimization Thresholds

The task of our optimization problem is to find a set of thresholds for the sequential classifier such that the CTC CAD system produces the minimal number of false negatives and false positive rate in the initial detection procedure. However, those two objectives are conflicting, and minimizing one of them usually leads to a increase of another. In practice, we must make trade-offs between the two objectives to meet our requirements. Pareto front based multiobjective optimization algorithms provide a set of non-dominated solutions to the problem at hand, which allows trade-offs based on our needs. Therefore, it is natural to form our task as a multiobjective optimization problem.

Let $f_1(\overrightarrow{x}$ and $f_2(\overrightarrow{x})$ denote the number of missed polyps and average number of false positives per data set, respectively, where $\overrightarrow{x}$ is a 14-dimensional threshold vector

$$\overrightarrow{x} = \{x_1, \dots, x_{14} | x_1, \dots, x_6 \in R, x_8 \dots x_{13} \in R, x_7 \in I, x_{14} \in I\} \tag{6}$$

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

$$\min_{\overrightarrow{x}} \overrightarrow{f}(\overrightarrow{x}) = \{f_1(\overrightarrow{x}), f_2(\overrightarrow{x})\} \tag{7}$$

subject to

$$\begin{cases} x_1 \in [-10, 0), x_2 \in [-10, 0), x_1 < x_2, \\ x_3 \in [0, 50], x_4 \in [0, 100], x_3 < x_4, \\ x_5 = 0, x_6 \in (0, 2], x_7 \in [6, 30], \\ x_8 \in [-10, 0), x_9 \in [-10, 0), x_8 < x_9, \\ x_{10} \in [0, 50], x_{11} \in [0, 100], x_{10} < x_{11}, \\ x_{12} = 0, x_{13} \in (0, 2], x_{12} < x_{13}, \\ x_{14} \in [6, 30]. \end{cases} \tag{8}$$

The range for each threshold was determined experimentally such that it is wide enough to include all solutions of interest. This pre-selecting of the threshold range is not a necessary step, but it allows reducing time and computational burden. The global optima of an MOP is called Pareto optimal set, which consists of solutions (thresholds) that are not dominated by any other solutions.[12] 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,

$$\begin{cases} \overrightarrow{x}_1 \succ \overrightarrow{x}_2, \ iff \\ \forall i \in \{1, 2\} : f_i(\overrightarrow{x}_1) \leq f_i(\overrightarrow{x}_2) \wedge \exists j \in \{1, 2\} : f_j(\overrightarrow{x}_1) < f_j(\overrightarrow{x}_2) \end{cases} \tag{9}$$

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.

The Pareto optimal solution are often obtained by multiobjective optimization algorithms. The multiobjective optimization technique was first investigated in 1985,[13] and since then, many multiobjective optimization algorithms have been reported.[14–17] We are particularly interested in the SPEA2 algorithm,[9,18] because of its fast convergence rate.[19] Further details on the SPEA2 algorithm for automatic polyp detection task can be found in.[8]

2.3. Validating Optimal Solutions

Our previous results showed that the sensitivity and specificity of our CTC CAD system for the initial polyp detection can be increased greatly on training data sets.[8] To validate the generalization capability of Pareto optimal sets obtained by SPEA2, we used a two-fold cross-validation (CV) method to see if testing results are close to training results. We performed CV on 133 patients, each patient had a prone and a supine CT scan. There were 103 colonoscopically confirmed polyps resulting in 188 positive detections in CTC reading. Some of the polyps were visible on both prone and supine views. We randomly divided the 133 patients' CT scans into two parts. Each part was held out as a test set. The remaining set was used by SPEA2 to get the Pareto front, and the obtained Pareto optimal set was used on the held-out part to provide test results. The training and test procedures were repeated two times so that each part was used as a test set only once. Finally, we average the training Pareto fronts and the test results to validate if the mean Pareto front is close to the average test results.

3. RESULTS AND DISCUSSION

We present our experiments and results of the validation in this section. Discussion on the results are also provided.

3.1. Data Acquisition

The CTC procedure was performed on 133 patients with a high suspicion of colonic polyps. All polyps were verified by follow-up optical colonoscopy. These patients were chosen from a larger cohort who underwent a CTC screening procedure. The majority of the polyps were identified on both the prone and supine views. There were 103 colonoscopically confirmed polyps resulting 188 positive detections.

3.2. Results

Fig. 1 shows some polyp detections on the colon surface. Fig. 2 shows our previous results on a subset of the data set used in this paper, where we obtained the optimized thresholds on 27 patients. Fig.2 compares the results obtained by SPEA2 to that by heuristic, where the horizontal axis represents false positive rate per data set and the vertical axis denotes the number of false negatives. 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. The remaining question is that if the optimized thresholds are able to be generalized to new data set.

Fig. 3 shows the two-fold CV results on the 133 patients, where we plot average Pareto front on training data and mean results on test data in the two-fold CV. Again, the horizontal axis in Fig. 3 denotes the false positive rate per data set. In the two-fold CV, we randomly divided the 133 patients into two parts, and it is not guaranteed that each part will have the same number of polyps. For a fair comparison between the training and testing results in the CV, we used the fraction of the false negatives as denoted in the vertical axis in Fig. 3.

It is observed that the test results are close to the training Pareto front. Note that a lower curve is better because it reduced both false negatives and false positive rate. We should mention that the test results may or may not contain Pareto dominated solutions. Therefore, the test results do not necessarily compose a Pareto front.

3.3. Discussion

Multiobjective optimization algorithms have been applied to generate upper bounds of receiver operating characteristic (ROC) or free-response ROC (FROC) curves.[16,17] However, the investigators efforts were put only on training data sets. It is not clear if the ROC or FROC curves generated by the multiobjective optimization technique are still valid for testing data sets. To our knowledge, this is the first attempt to validate Pareto optimal solutions in the setting of a CTC CAD system. From our experiments, the multiobjective optimization technique appears to be a valid tool to adjust parameters in the CTC CAD system. We are currently working to validate a CTC CAD system that uses one of the Pareto optimal sets for the sequential classifier in the initial polyp detection on the colon surface. We plan to examine if the optimized sequential classifier can improve the whole system, in term of the FROC curve on independent testing data.

4. CONCLUSION

We have implemented a multiobjective evolutionary algorithm, SPEA2, for optimizing a CAD polyp detection system. We demonstrated that the algorithm is able to generalized to new data for our CTC CAD system. The Pareto front-based optimization algorithm appears to be a robust and powerful tool for optimizing CTC CAD systems.

5. ACKNOWLEDGEMENTS

We 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 Beowolf PC/Linux cluster.

REFERENCES

1. 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.
2. 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.
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 colonography," *Abdominal Imaging* **27**, pp. 268–274, 2002.
5. 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.
6. 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.
7. Z. Wang, Z. Liang, L. Li, X. Li, B. Li, J. Anderson, and D. Harrington, "Reduction of false positives by internal features for polyp detection in CT-based virtual colonoscopy," *Medical Physics* **32**(12), pp. 3602–3616, 2005.
8. J. Li, A. Huang, J. Yao, I. Bitter, N. Petrick, R. M. Summers, P. J. Pickhardt, and J. R. Choi, "Automatic colonic polyp detection using multiobjective evolutionary techniques," in *Proceeding of SPIE*, 2006.
9. E. Zitzler, M. Laumanns, and L. Thiele, "SPEA2: Improving the strength pareto evolutionary algorithm," technical report, Swiss Federal Institute of Technology, 2001.
10. 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.
11. J. Yao and R. Summers, "3D colonic polyp segmentation using dynamic deformable surfaces," in *SPIE Medical Imaging*, 2004.
12. 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.
13. 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.
14. 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.
15. 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.
16. M. A. Anastasio, M. A. Kupinski, and robert M. Nishikawa, "Optimization and FROC analysis of rule-based detection schemes using a multiobjective approach," *IEEE Trans. on Medical Imaging* **17**(6), pp. 1089–1093, 1998.

17. M. A. Kupinski and M. A. Anastasio, "Multiobjective genetic optimization of diagnostic classifiers with implications for generating receiver operating characteristic curves," *IEEE Trans. on Medical Imaging* **18**(8), pp. 675–685, 1999.
18. Z. Michalewicz, *Genetic Algorithms + Data Structures = Evolution Programs*, Springer-Verlag, Berlin, Heidelberg, NY, third ed., 1999.
19. 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.

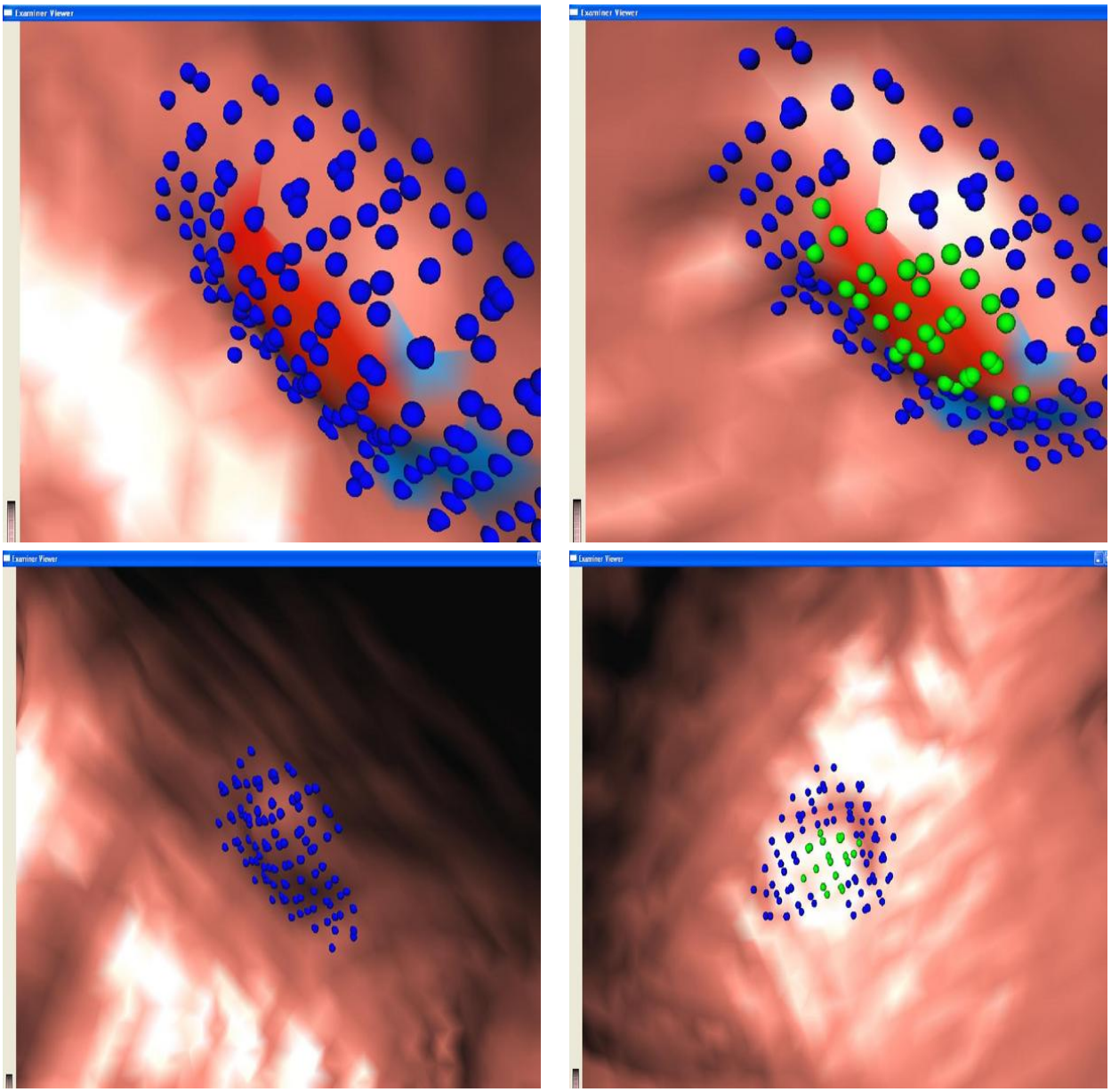

Figure 1. Polyp detection samples comparison. The first column consists of missed true detection examples by our old system and the second column represents the same detections picked by our system using one set of optimized thresholds, which is chosen from the training Pareto optimal set. The blue dots represent ground truth vertices for polyps, and the green dots denote detected ground truth vertices.

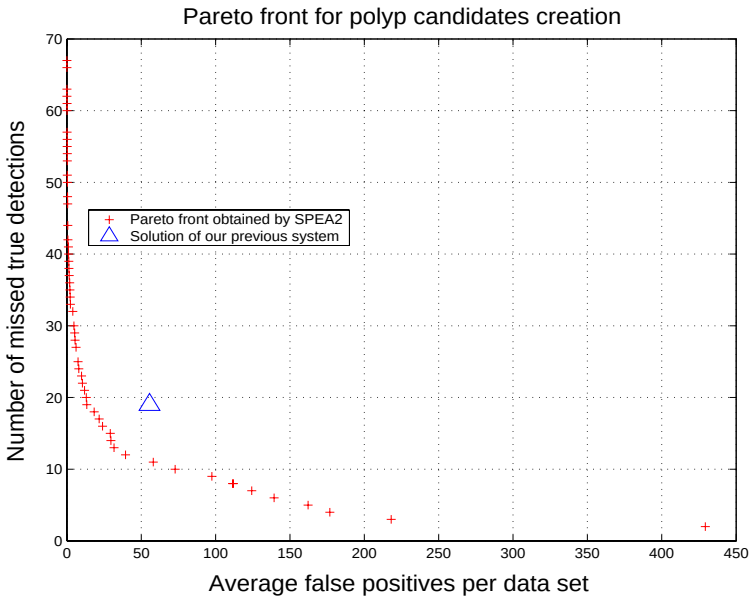

Figure 2. The solution provided by our old system and the Pareto front from the SPEA2 algorithm (reprint from our previous work in SPIE 2006)

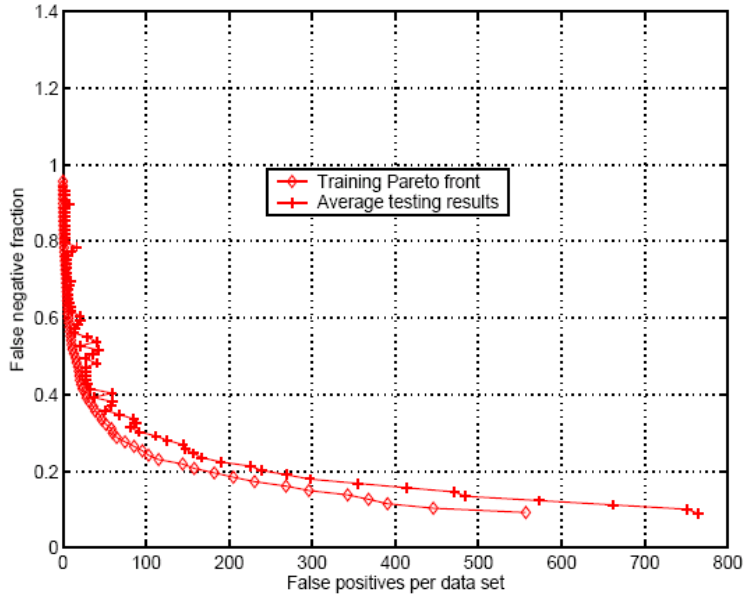

Figure 3. Two-fold cross-validation results on 133 patients