Published in final edited form as:
*Int J Comput Intell Appl*. 2010 January 1; 9(1): 1–15. doi:10.1142/S1469026810002744.

# Combining Statistical and Geometric Features for Colonic Polyp Detection in CTC Based on Multiple Kernel Learning

**Shijun Wang**[1], **Jianhua Yao**[1], **Nicholas Petrick**[2], and **Ronald M. Summers**[1,*]

[1]Imaging Biomarkers and Computer-Aided Diagnosis Laboratory, Radiology and Imaging Sciences, National Institutes of Health Clinical Center, Building 10 Room 1C368X MSC 1182, Bethesda, MD 20892-1182

[2]NIBIB/CDRH Laboratory for the Assessment of Medical Imaging Systems, Food and Drug Administration

## Abstract

Colon cancer is the second leading cause of cancer-related deaths in the United States. Computed tomographic colonography (CTC) combined with a computer aided detection system provides a feasible approach for improving colonic polyps detection and increasing the use of CTC for colon cancer screening. To distinguish true polyps from false positives, various features extracted from polyp candidates have been proposed. Most of these traditional features try to capture the shape information of polyp candidates or neighborhood knowledge about the surrounding structures (fold, colon wall, etc.). In this paper, we propose a new set of shape descriptors for polyp candidates based on statistical curvature information. These features called histograms of curvature features are rotation, translation and scale invariant and can be treated as complementing existing feature set. Then in order to make full use of the traditional geometric features (defined as group A) and the new statistical features (group B) which are highly heterogeneous, we employed a multiple kernel learning method based on semi-definite programming to learn an optimized classification kernel from the two groups of features. We conducted leave-one-patient-out test on a CTC dataset which contained scans from 66 patients. Experimental results show that a support vector machine (SVM) based on the combined feature set and the semi-definite optimization kernel achieved higher FROC performance compared to SVMs using the two groups of features separately. At a false positive per scan rate of 5, the sensitivity of the SVM using the combined features improved from 0.77 (Group A) and 0.73 (Group B) to 0.83 ($p \leq 0.01$).

### Keywords

Computed Tomographic Colonography; colonic polyp detection; multiple kernel learning; histograms of curvature features; semi-definite programming

## 1. Introduction

Colon cancer is the second leading cause of cancer-related deaths in the United States. Computed tomographic colonography (CTC), also known as virtual colonoscopy (VC) when a fly through viewing mode is used, provides an alternative to optical colonoscopy in screening

*Corresponding Author and Reprint Requests: Ronald M. Summers, M.D., Ph.D., Senior Investigator and Staff Radiologist, Chief, Clinical Image Processing Service, Chief, Imaging Biomarkers and Computer-Aided Diagnosis Laboratory, Radiology and Imaging Sciences, National Institutes of Health Clinical Center, Building 10 Room 1C368X MSC 1182, Bethesda, MD 20892-1182, Phone (office): (301) 402-5486, Phone (lab): (301) 435-0055, Fax: (301) 451-5721, rms@nih.gov, http://www.cc.nih.gov/about/SeniorStaff/ronald_summers.html.

patients for colonic polyps. CTC x-ray exposure making it less invasive than traditional optical colonoscopy. When CTC is performed in conjunction with computer aided detection (CAD) software, it can be highly accurate for detecting colonic polyps [1–7].

CAD systems for detection of polyps on CTC have been under investigation over the past decades. Summers et al. [2–4] employed many geometrical features to find polyp candidates. Most of these features are related with statistics of various curvatures. Yoshida et al. [8] applied the shape index and curvedness measures to describe polyp candidates. These two measures can be derived from principal curvatures of an inner colon surface and show better performance for polyp detection compared with Gaussian and mean curvatures. Paik et al. [6] developed a novel detection method called surface normal overlap which can capture the shape of polyp candidates. In order to make use of the internal texture information of polyp candidates, Wang et al. [7] proposed a polyp detection method which employs geometrical, morphological and textural features inside polyp candidates.

To discern true polyps from false positives, one needs to capture the intrinsic difference between the two groups. In our previous CAD system [2–4], after 3-dimensional colon surface segmentation and polyp candidate segmentation, we extracted 18 discriminative features from each polyp candidate. These features were designed to mimic the experience of radiologists in distinguishing polyps. Examples of these features include the average of mean curvatures on the neck of polyp candidate, the average of Gaussian curvatures of the lesion, the distance between the two most separated vertices on the surface cluster and wall thickness to name a few. To some extent, they incorporate something akin to the human knowledge used in detecting polyps.

In recent years, feature extraction based on statistical information of basic image descriptors has shown promising results in image processing and computer vision research. The basic image descriptors include edge descriptor, gradient descriptor, etc. Representative methods include Scale-Invariant Feature Transform (SIFT) [9], Shape Context [10] and Histograms of Oriented Gradient (HOG) descriptor [11], etc. The SIFT descriptors focus on the local appearance of the object at particularly interesting points [9]. By using orientation assignments at key points and matching of orientation histograms, the SIFT features are robust to changes of illumination and viewpoint and to the presence of noise and occlusion. Shape Context [10] describes the shape of an object by using log-polar bins to capture the relative location distribution of other edge points in relation to the central edge point. The key idea of Shape Context is that, when randomly choosing one key point, the distribution of the relative positions of other key points to the chosen point is a robust, compact and highly discriminative descriptor that can handle non-rigid transformation. Histogram of Oriented Gradient descriptor counts occurrences of gradient orientation in localized portions of an image [11]. For the detection of colonic polyps, CT images have lower resolutions compared with that of optical images. Consequently, SIFT and Shape Context do not appear to be the most suitable descriptors because they rely on key or edge points. Usually in CT images there are no distinct local structures which can serve as markers. HOG does not need to locate key points but it is sensitive to image rotation.

Inspired by the idea of HOG, in this paper we propose a set of new shape descriptors called histograms of curvature features (HCFs). HCFs are calculated based on the distributions of curvature features of polyp candidates: we first calculate curvature features for each voxel inside a polyp candidate; then we build a histogram for each curvature feature and each polyp candidate. HCFs are statistical descriptors that may be able to capture some of the intrinsic properties of true polyps, for example, shape and texture information. We utilize curvature-related descriptors as basic image descriptors, i.e. shape index, curvedness, Gaussian and mean curvatures, etc. The advantage of these curvature related descriptors is that they are rotation,

translation and scale invariant. HCFs can be treated a more complete statistical view of a polyp candidate based on the set of curvature-related descriptors.

Now we have two groups of features developed to describe polyp candidates. One incorporates the human experience about distinguishing true polyps from false positives and the other contains more complete statistical information about the features. They are complementary to some extent and can be treated as different depiction of the polyp candidates. Our hypothesis for this study is that the combination of these two groups of features will lead to better overall classification performance. The simplest technique to combine the feature sets is to concatenate them together but this may merge the information in an undesirable way because a large number of irrelevant features may overwhelm informative features. Our approach to solving this issue is to use a multiple kernel learning approach which was developed in recent years to solve real applications which involve multiple, heterogeneous data sources [12–14]. This so-called "multiple kernel learning" problem usually can be solved by considering the convex combinations of $K$ kernels, i.e.,

$$K\left(x_i, x_j\right) = \sum_{k=1}^{K} \beta_k K_k\left(x_i, x_j\right),$$

with $\beta_k \geq 0$ and $\sum_{k=1}^{K} \beta_k = 1$, where each kernel $K_k$ uses a group of features from one information source and $x_i$, $x_j$ are samples. To make full use of both the traditional and HCF features, in this paper we treat the problem as learning from multiple information sources and apply a multiple kernel learning method to develop an optimized kernel for classification from the two groups of features. The paper is organized as follows: in Sec. 2, we introduce recent advances in CTC CAD systems. in Sec. 3, we propose a new group of features - the histograms of curvature features. A well-known multiple kernel learning method based on semi-definite programming is briefly introduced in Sec. 4. Then we show experimental results in Sec. 5. We compare the performance of our method with existing CTC CAD systems in Sec. 6. Sec. 7 concludes with a short summary.

## 2. Recent advances in CTC CAD

In the past few years, CTC CAD has seen a number of advances. CTC CAD's are complicated system which involve multiple stages including: image segmentation, initial polyp candidate detection, feature extraction, classification, etc. In this section we will introduce several exemplar technologies from the most recent literature. The performance of these CTC CAD systems are listed in Table 3. Because of the rapid development in this research area, we may not be able to illustrate all the recent advances.

Geometric analysis plays an important role in the CTC polyp detection. In reference [15], Zhu et al. proposed the idea of volumetric mucosa for a thick colon wall representation. Then they employed a level set-based adaptive convolution method for first- and second- order spatial derivative calculation which is more accurate than traditional geometric analysis methods based on single layer mucosa. The success of the proposed method shows that the importance of initial image processing stages, such as the colon wall segmentation, on overall CTC CAD system performance.

Pattern recognition methods based on massive-training artificial neural networks (MTANNs) - are another new approach in CTC CAD research [16]. MTANN is a supervised image-processing/pattern recognition technique based on neural networks. It extracts features in an implicit way by interweaving feature extraction and classification in human perception.

MTANN methods are thought to be better approximation of the detection process used by actual radiologists which may lead to additional improvements for CTC polyp detection.

Inspired by radiologists' reading behavior when they read CTC data in endoluminal flythrough mode, Li et al. proposed a wavelet analysis based method for false positive reduction in CTC CAD [17]. This approach first develops a 2D snapshots of CTC CAD detections from the rendered 3D space. The camera is placed along the colon centerline and the view angle is optimized to differentiate false positives and true polyps. An enhanced wavelet analysis method is then used to extract features from the 2D snapshots. Experimental results show that the proposed method can significantly reduce false positives while maintaining a high polyp detection sensitivity.

Classifier design is a critical procedure for a practical CTC CAD system. In reference [18],van Ravesteijn designed a CAD system using logistic regression. The proposed linear logistic classifier can cope with small training set and imbalance between true and false detections. The proposed method has high sensitivity and low false positives per scan with only three features employed: the protrusion of the colon wall, the mean internal intensity and an auxiliary feature to determine whether the detections are on the rectal enema tube.

## 3. Histograms of curvature features

Shape analysis plays an important role in the detection of colonic polyps. Colonic polyps appear as bulbous protrusions that adhere either to the inner wall of the colon or to colonic folds that have elongated, ridge like structures. Fig. 1 shows a typical adenomatous polyp from prone and supine scans which locates in the conjunction of two folds on the sigmoid colon. How to extract distinct features from polyp candidates is a key factor in the detection of colonic polyps. Perhaps the most widely used features to describe polyp candidates are Gaussian and mean curvatures because they capture the semispherical property of polyps.

Intuitively, curvature measures the extent to which a geometric object deviates from a flat surface [19]. For a two-dimensional surface embedded in $R^3$, the intersection of the surface with a plane containing the normal vector and one of the tangent vectors at a point on the surface is a plane curve and has a curvature called normal curvature. The maximum and minimum values of the normal curvature at a point are called the principal curvatures, $k_1$ and $k_2$. The directions of the corresponding tangent vectors are called principal directions. The Gaussian curvature is defined as the product of the principal curvatures: $k_{Gaussian} = k_1 k_2$. A surface is locally convex when Gaussian curvature is positive; it is locally saddle when the Gaussian curvature is negative. The mean curvature is one-half of the sum of the principal curvatures: $k_{mean} = \frac{k_1 + k_2}{2}$.

Besides Gaussian and mean curvatures, shape index (SI) and curvedness (CV) can also describe the shape of a polyp candidate [8]. At a given voxel $p$, the SI and CV features can be defined as follows:

$$SI(p) = \frac{1}{2} - \frac{1}{\pi} \arctan \frac{k_1(p) + k_2(p)}{k_1(p) - k_2(p)}, \quad (1)$$

$$CV(p) = \sqrt{\frac{k_1^2(p) + k_2^2(p)}{2}}, \quad (2)$$

where $k_1(p) > k_2(p)$ are the two principal curvatures. To characterize polyp candidates, H. Yoshida et al. used mean SI, mean directional gradient concentration (DGC) and the variance of the CT value as features in the classification procedure [8]. Given the extracted region R of a polyp candidate, the mean value of the SI is defined as: $SI(R) = \frac{1}{|R|}\sum_{p\in R} SI(p)$, where $|R|$ is the number of voxels in region R.

Using the mean value of features for all points on a segmented detection is a good way to distinguish true polyps and false positives under certain circumstances. But at the same time, we also lose some information during this process. If we treat the calculation of the mean value as a dimensionality reduction process, then during the mapping from the high dimensional space to the one dimensional space (the mean space), we will inevitably lose certain information from the original high dimensional data. This introduces a new question: how to make full use of statistical features calculated on each voxel.

To make full use of curvature information and capture internal texture information of polyp candidates, inspired by the idea of HOG, we utilize histograms of curvature features to represent polyp candidates. Fig. 2 shows comparisons of the shape index and curvedness histograms obtained from a true polyp and a false positive. The plots in the figure show that the full histograms for true and false positive candidates can be very different even when a single summary measure might suggest that the two are similar. In Table 1, we list six curvature-related features contained in HCFs. For each feature, we choose a range and divide it into 198 equally-spaced bins. Voxels whose feature values are smaller than the lower limit or greater than the upper limit are counted in two additional bins. The lower and upper limits are selected based on the distributions of features. In addition, previous research shows that the CT value is also an informative feature [2–4]. We also compute the histogram of CT values for each polyp candidate. We concatenate the seven histograms and get a feature vector with 1400 dimensions for each polyp candidate.

## 4. Multi-kernel learning by semi-definite programming

By applying traditional supervised and unsupervised learning methods in the feature space, kernel methods provide powerful tools for data analysis and have been found to be successful in a number of real applications. Support vector machines (SVMs) are a set of kernel based supervised learning methods used for classification and regression [20]. SVMs try to minimize the empirical classification error and maximize the geometric margin simultaneously on the training set which leads to high generalization ability on the new samples.

For a two-class classification problem, given training samples $\{(x_1, y_1),\ldots,(x_n, y_n)\}$, $y_i \in \{-1, +1\}$, the optimization problem for learning a linear classifier in the feature space is defined as (hard margin):

$$\min_{w,b} \langle w, w\rangle, \tag{3}$$

$$\text{subject to } y_i(\langle w, \Phi(x_i)\rangle + b) \geq 1,\ i=1,\ldots,n, \tag{4}$$

where $\Phi$ is the mapping from original space to feature space. It is a quadratic programming (QP) optimization problem and it is convex. The optimal $(w_*, b_*)$ is a maximal margin classifier with geometric margin $\gamma = 1/\|w_*\|^2$ if it exists.

When the data are not linearly separable in the feature space, the above problem can be relaxed by introducing relaxation variables. For example, the 2-norm soft margin SVM is defined as:

$$\min_{w,b,\xi} \langle w, w\rangle + C\sum_{i=1}^{n}\xi_i^2, \tag{5}$$

$$\text{subject to } y_i\left(\langle w, \Phi(x_i)\rangle + b\right) \geq 1-\xi_i,\ \xi_i \geq 0,\ i=1,\ldots,n \tag{6}$$

where C controls overall impact of the constraints and $\xi_i$, $i = 1,\ldots,n$, are the relaxation variables for each sample.

In the basic optimization problems of SVMs introduced above, there is only one mapping function Φ which maps input vector to feature space. In real applications, we often have multiple information sources to describe the same object. For example, in the colonic polyp detection problem, histograms of curvature features can be treated as another descriptor to describe a polyp candidate besides traditional features which incorporate human knowledge in describing polyps. So multi-kernel learning is one natural solution to solve our optimization problem.

G. Lanckriet et al. first formulated the kernel matrix learning from multiple information sources as a semidefinite programming problem [13]. In the semidefinite programming problem, the objective function is a linear objective and the constraints are a linear matrix inequality and affine equality constraints defined as:

$$\min_x c^T x, \text{ subject to,} \tag{7}$$

$$F(x) = F_0^j + x_1F_1^j + \ldots + x_qF_q^j \prec = 0,\ j=1,\ldots,L \text{ and } Ax=b, \tag{8}$$

where $x \in R^q$ is the vector of decision variables, $c \in R^q$ is the objective vector, and matrices $F_i^j$ are given and symmetric, positive semidefinite (e.g. kernels calculated from different information sources).

Under the transduction setting, given training data $S_{n_{tr}} = \{(x_1, y_1),\ldots,(x_{n_{tr}}, y_{n_{tr}})\}$ and the test set $T_{n_t} = \{x_{n_{tr}+1},\ldots,x_{n_{tr}+n_t}\}$, we can construct a kernel matrix $K = \begin{pmatrix} K_{tr} & K_{tr,t} \\ K_{tr,t}^T & K_t \end{pmatrix}$ where $K_{ij} = \langle \Phi(x_i), \Phi(x_j)\rangle$, $i, j = 1,\ldots,n_{tr},n_{tr}+1,\ldots,n_{tr}+n_t$. For the optimal kernel learning problem, the basic idea is to optimize a cost function over the training data block $K_{tr}$ in order to learn the optimal mixed block $K_{tr,t}$ and the optimal test data block $K_t$. If we have multiple kernel matrices $K_i$, $i = 1,\ldots,m$ generated from $m$ information sources, by imposing maximum margin constraints on the training samples and limit the target matrix in the linear subspace $K = span\{K_1,\ldots,K_m\} \cap \{K \succ = 0\}$ the kernel learning problem from multiple kernels can be formulated as the following semidefinite programming problem with 2-norm soft margin [13]:

$$\min_{\mu_i,t,\lambda,\nu} t, \text{ subject to}$$

$$trace\left(\sum_{i=1}^{m}\mu_i K_i\right)=c,\ \sum_{i=1}^{m}\mu_i K_i \succ =0,\ \begin{pmatrix} G\left(\sum_{i=1}^{m}\mu_i K_{i,tr}\right)+\frac{1}{C}I_{n_{tr}} & e+\nu+\lambda y \\ (e+\nu+\lambda y)^T t & \end{pmatrix} \succ =0, \text{ and} \tag{9}$$

$$\nu \succ =0,\ \text{where } G=diag\,(y)\ \times\left(\sum_{i=1}^{m}\mu_i K_i\right)\times\ diag\,(y)\ \text{ and } C \text{ is a constant.} \tag{10}$$

$y$ is the label vector of training data and $e$ is a $n_{tr} \times 1$ vector with all elements equal to one.

## 5. Experimental results

To evaluate our method, we applied it to a CTC dataset that contains 109 CT scans of 66 patients collected from three hospitals. Each scan was done during a single breath hold using a 4-channel or 8-channel CT scanner (General Electric LightSpeed or LightSpeed Ultra; GE Helthcare 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. After 3-dimensional colon surface segmentation and polyp candidate segmentation, we extracted two groups of features from each polyp candidate. Group A contains 18 traditional features used for colonic polyp detection. In Table 2, we list all 18 features and their corresponding geometric explanations. Group B contains 1400 histogram features of curvatures and CT value. We got 5763 polyp candidates which includes 142 true detections from 96 polyps whose sizes are equal to or greater than 6 mm. To focus on the classification performance of our method, here we do not consider uniqueness, meaning that several true polyp detections may come from the same physical polyp.

To evaluate the effectiveness of our method for classification of colonic polyps, we did leave-one-patient-out test to compare the performance of SVM using learned optimal kernel with that of SVM using group A features and group B features separately. For group A features, radius base function (RBF) kernel with width $\sigma = 1/18$ was used. Previous work of Chapelle et al. showed that for histogram features, Laplacian radius base function (LRBF) kernel usually works better than RBF kernel [21]. So for group 2 features, LRBF kernel with width $\sigma = 2.5$ was used. The values of $\sigma$ are selected based on the results of cross validation. For multiple kernel learning, we adopted the 2-norm soft margin solution because the classification problem is not linearly separable. The 2-norm soft margin parameter of SVMs $C$ was determined automatically using SDP by treating it as a variable in the SDP constraints. The optimal kernel learned by SDP on the two base kernels was directly fed into SVM for training and testing. Fig. 3 shows the free-response operator characteristic (FROC) curves on the testing set. Because the CTC dataset is extremely unbalanced in the number of true positive samples compared with the number of false positives, a bias on the decision boundary will be observed if all of the training samples are used during classifier training (most true polyps will be classified as false positives). To circumvent this problem, down-sampling of the negative samples was done in order to balance the true and false training cases. We used all the true samples and down-sampled the same number of false positives from the train set. The true samples and selected negative samples compose the training set (smaller than the whole train set). For each test patient in the leave-one-patient-out paradigm, we did 30 random samplings on the train set and trained 30 classifiers respectively. So for the testing patient, we will have 30 prediction values for each polyp candidates contained in the testing patient. In the last step, we did free response receiver operation characteristic (FROC) analysis and the FROC was

counted based on all the testing patients. We got 30 FROC curves in the end. The standard deviations of them are shown in the Fig. 3 (b). For clinical purposes, radiologists are mostly concerned about the sensitivity of the CAD system at low false positives per scan. When the number of false positives per scan is 5, the sensitivity improved to 0.83 using the multiple kernel learning method, whereas the sensitivities of SVMs are 0.77 and 0.73 for group A and group B features respectively. T-test hypothesis testing shows that the results of semidefinite programming result in a significant difference at the 0.05 significance level compared with that of SVM using group A feature and SVM using group B feature at false positives per scan of 5 ($p \leq 0.01$).

In the multiple kernel learning process, we used 2 base kernels. One was constructed on the group A features by using RBF kernel function; the other was constructed on the group B features by using Laplacian RBF kernel. In Fig. 4, we show the two base kernel matrices and the optimal kernel matrix of training data of a typical run which contains 140 true polyps and 140 false positives. From Fig. 4 we can find that although the base kernels show no clear separability; the optimal kernel learned by semidefinite programming shows better separability which is very helpful for the classification task.

To show that group A and group B features are complementary for the classification purpose, we selected 20 true polyp detections and compared SVM prediction values of them by using group A and group B features separately. These polyps are considered tough or easily misclassified samples by SVM using group A features. SVM prediction value provides a measure showing that how confident the SVM feels a polyp candidate is a true polyp. A true polyp with low SVM prediction values means that the SVM treats the true polyp more like a false positive in the feature space. We show comparisons of SVM prediction values of the twenty polyps in Fig. 5 and from it we can find that for some hard cases considered by SVM using group A features, SVM using histogram features gives higher prediction values and they are treated as true polyps with high confidence. In Fig. 6, we show four true polyps of them. These polyps are hard to distinguish by using group A feature because they are flat, or partially corrupted because of segmentation problem. By utilizing the statistical information inside the polyp, SVM using group B feature can still distinguish them although they are very flat or partially corrupted.

## 6. Comparison with state-of-the-art CTC CAD systems

As introduced in the introduction section, during the past ten years, many colonic polyp detection methods have been proposed. For example, in the work of Yoshida et al., they proposed the detection method based on shape index and curvedness and showed that the detection of performance of their CAD can reach 100% at false positives 2 per patient on a small CTC dataset which contains 43 CTC cases and 12 5-10mm polyps [8]. Kiss et al. proposed a new detection method via combination of surface normal and sphere fitting methods and reported 40% sensitivity at false positives of 8.16 per patient on a CTC dataset containing 18 patients [22]. CTC CAD is a very complicated system and it has many components and procedures which have significant influence to the final performance. For example, the segmentation algorithm of colon, the filter criterion for polyp candidates, classifier design, etc. will affect the sensitivity and specificity of a CTC CAD system. Usually these CTC CAD systems are not available to the public. We have conducted a literature review and summarized the performance of state-of-the-art CTC CAD systems in Table 3. We only consider the CTC CAD systems published in past two years which employ cutting edge technologies for CTC polyp detection and show better performance. For each method, we include the size of the data set used, information on the polyps, and sensitivity and false positives per patient/scan. From the table we see that these CTC CAD systems show high sensitivity with a fairly low number of false positives per patient/scan. Our method shows comparable performance to the recently

published methods listed in Table 3 with the caveat that each of these studies uses a unique dataset. This complicates a direct comparison of algorithm performances. The main contributions of this paper are the introduction and evaluation of a new set of statistical descriptors for polyp detection and a mechanism for combining geometric features and statistical features. It can be applied to any existing CTC CAD systems and be viewed as enhancing technology to CTC CAD systems.

## 7. Conclusion

In this paper, we propose a new method for colonic polyp detection from CT images based on statistical image descriptors and multi-kernel learning method. Two groups of features were used in the multi-kernel learning. Group A features are traditional and discriminative features designed according to the experience of radiologists. It contains various features to discern polyps from the view point of human specialists. Group B features are based on the statistical information of polyp candidates. Most of them are related with various curvatures which are invariant to rotation, translation and scale transformation. Experimental results on a CTC dataset containing 66 patients with >=6 mm polyps show the effectiveness of the proposed method. At a false positive per scan rate of 5, the sensitivity using the combined features improved from 0.77 (Group A) and 0.73 (Group B) to 0.83 ($p \leq 0.01$). We also analyzed the classification result of SVM using the two groups of features separately. Comparisons show that the two groups of features are complementary for a number of polyp candidates thereby explaining why combining them together leads to better detection performance.

## Acknowledgments

This research was supported by the Intramural Research Programs of the NIH Clinical Center and the U.S. Food and Drug Administration (NP). We thank Drs. Perry Pickhardt, J. Richard Choi and William Schindler for providing CT colonography data.

## Reference

1. Fletcher JG, Luboldt W. CT colonography and MR colonography: current status, research directions and comparison. Eur Radiol 2000;vol. 10:786–801. [PubMed: 10823635]
2. Summers RM, Yao J, Pickhardt PJ, Franaszek M, Bitter I, Brickman D, Krishna V, Choi JR. Computed tomographic virtual colonoscopy computer-aided polyp detection in a screening population. Gastroenterology 2005;129:1832–1844. [PubMed: 16344052]
3. Summers RM, Johnson CD, Pusanik LM, Malley JD, Youssef AM, Reed JE. Automated polyp detection at CT colonography: feasibility assessment in a human population. Radiology 2001;vol. 219:51–59. [PubMed: 11274534]
4. Summers RM, Beaulieu CF, Pusanik LM, Malley JD, Jeffrey RB Jr, Glazer DI, Napel S. Automated polyp detector for CT colonography: feasibility study. Radiology 2000;vol. 216:284–290. [PubMed: 10887263]
5. Yoshida H, Masutani Y, MacEneaney P, Rubin DT, Dachman AH. Computerized detection of colonic polyps at CT colonography on the basis of volumetric features: pilot study. Radiology 2002;vol. 222:327–336. [PubMed: 11818596]
6. Paik DS, Beaulieu CF, Rubin GD, Acar B, Jeffrey RB, Yee J, Dey J, Napel S. Surface normal overlap: A computer-aided detection algorithm, with application to colonic polyps and lung nodules in helical CT. IEEE Transactions on Medical Imaging 2004;vol. 23:661–675. [PubMed: 15191141]
7. Wang Z, Liang Z, Li L, Li X, Li B, Anderson J, Harrington D. Reduction of false positives by internal features for polyp detection in CT-based virtual colonoscopy. Med Phys 2005;vol. 32:3602–3616. [PubMed: 16475759]
8. Yoshida H, Nappi J. Three-dimensional computer-aided diagnosis scheme for detection of colonic polyps. IEEE Transactions on Medical Imaging 2001;vol. 20:1261–1274. [PubMed: 11811826]

9. Lowe D. Distinctive image features from scale-invariant keypoints. Internation Journal of Computer Vision 2004;vol. 60:91–110.
10. Belongie JMS, Puzicha J. Shape Matching and Object Recognition Using Shape Contexts. IEEE Transactions on Pattern Analysis and Machine Intelligence 2002;vol. 24:509–522.
11. Dalal, aBTN. Histograms of oriented gradients for human detection. presented at IEEE Computer Society Conference on Computer Vision and Pattern Recognition; 2005.
12. Chapelle O, Vapnik V, Bousquet O, Mukherjee S. Choosing multiple parameters for support vector machines. Machine Learning 2002;vol. 46:131–159.
13. Lanckriet GRG, Cristianini N, Bartlett P, El Ghaoui L, Jordan MI. Learning the kernel matrix with semidefinite programming. Journal of Machine Learning Research 2004;vol. 5:27–72.
14. Sonnenburg S, Ratsch G, Schafer C, Scholkopf B. Large scale multiple kernel learning. Journal of Machine Learning Research 2006;vol. 7:1531–1565.
15. Zhu H, Duan C, Pickhardt P, Wang S, Liang Z. Computer-aided detection of colonic polyps with level set-based adaptive convolution in volumetric mucosa to advance CT colonography toward a screening modality. Cancer Management and Research 2009;vol. 1:1–13. [PubMed: 20428331]
16. Suzuki K, Yoshida H, Nappi J, Armato SG, Dachman AH. Mixture of expert 3D massive-training ANNs for reduction of multiple types of false positives in CAD for detection of polyps in CT colonography. Medical Physics 2008;vol. 35:694–703. [PubMed: 18383691]
17. Li J, Van Uitert R, Yao JH, Petrick N, Franaszek M, Huang A, Summers RM. Wavelet method for CT colonography computer-aided polyp detection. Medical Physics 2008;vol. 35:3527–3538. [PubMed: 18777913]
18. van Ravesteijn VF, van Wijk C, Vos FM, Truyen R, Peters JF, Stoker J, van Vilet LJ. Computer Aided Detection of Polyps in CT Colonography Using Logistic Regression. IEEE Transactions on Medical Imaging 2009;vol. in press.
19. Porteous, IR. Geometric Differentiation: For the Intelligence of Curves and Surfaces. Second ed. Cambridge University Press; 2001.
20. Burges CJC. A tutorial on Support Vector Machines for pattern recognition. Data Mining and Knowledge Discovery 1998;vol. 2:121–167.
21. Chapelle O, Haffner P, Vapnik VN. Support vector machines for histogram-based image classification. Ieee Transactions on Neural Networks 1999;vol. 10:1055–1064. [PubMed: 18252608]
22. Kiss G, Van Cleynenbreugel J, Thomeer M, Suetens P, Marchal G. Computer-aided diagnosis in virtual colonography via combination of surface normal and sphere fitting methods. Eur Radiol 2002;vol. 12:77–81. [PubMed: 11868078]
23. Chowdhury TA, Whelan PF, Ghita O. A fully automatic CAD-CTC system based on curvature analysis for standard and low-dose CT data. Ieee Transactions on Biomedical Engineering 2008;vol. 55:888–901. [PubMed: 18334380]

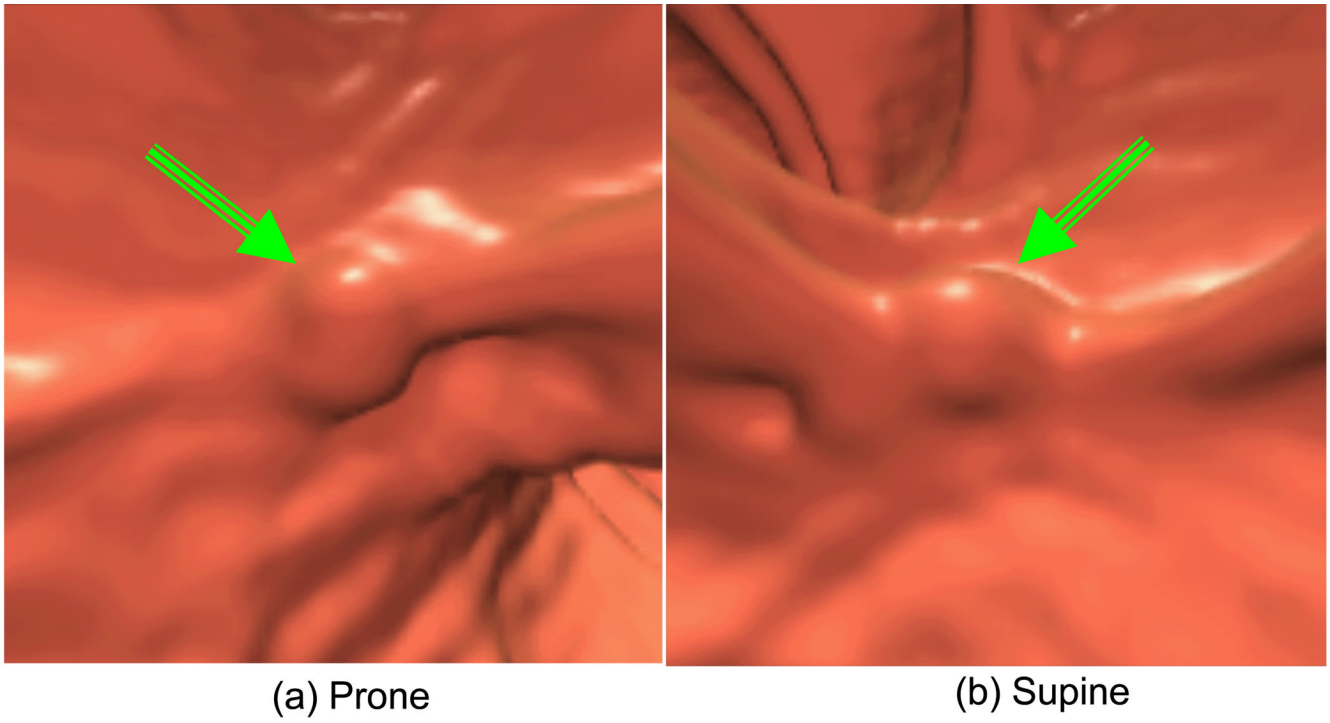



**Figure 1.**
3D endoluminal surface reconstructions of an 8 mm adenomatous polyp (arrows) which locates in the sigmoid colon. (a) Prone and (b) supine scans.

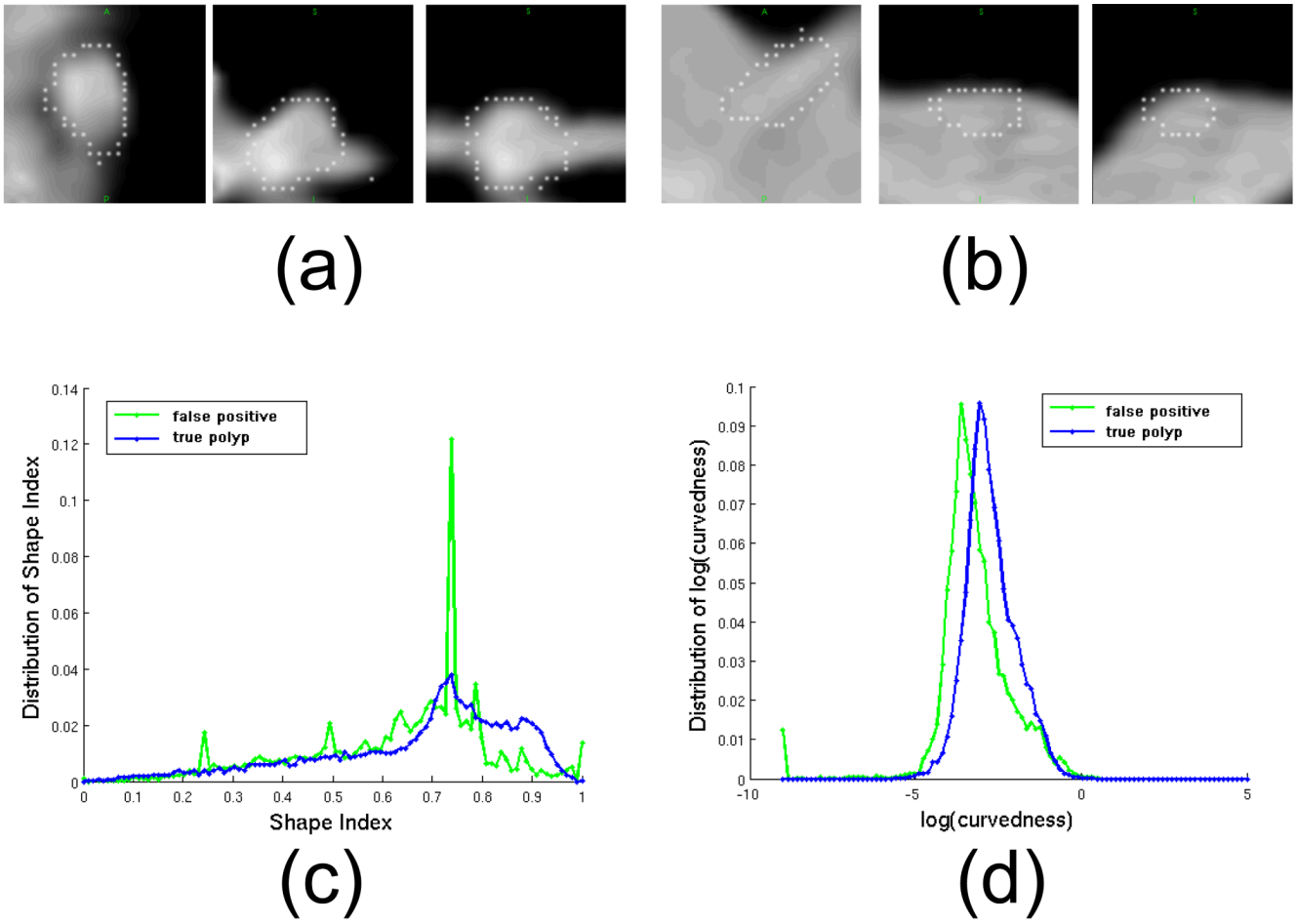
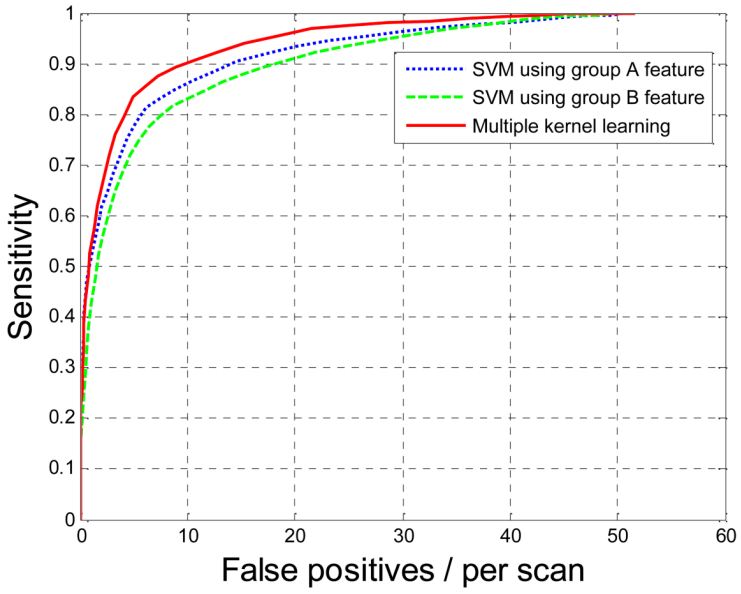
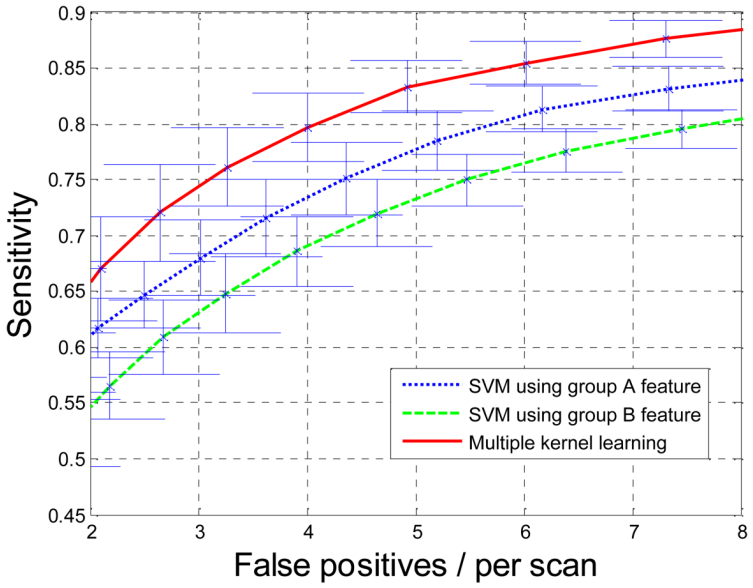



**Figure 2.**
(a) shows axial, coronal and sagittal slices of a 10 mm true polyp in which white dots mark the segmentation boundary of the candidate. (b) shows a false positive. Their distributions of shape index and curvedness are shown in (c) and (d) respectively.




(a)




(b)

**Figure 3.**

(a) Comparisons of FROC curves from SVMs using group A features alone, using group B features alone and using the semi-definite programming method to combine the two feature sets. (b) shows a magnified version of the plot in (a) for low numbers of false positives per scan. The proposed method achieves the highest sensitivity.

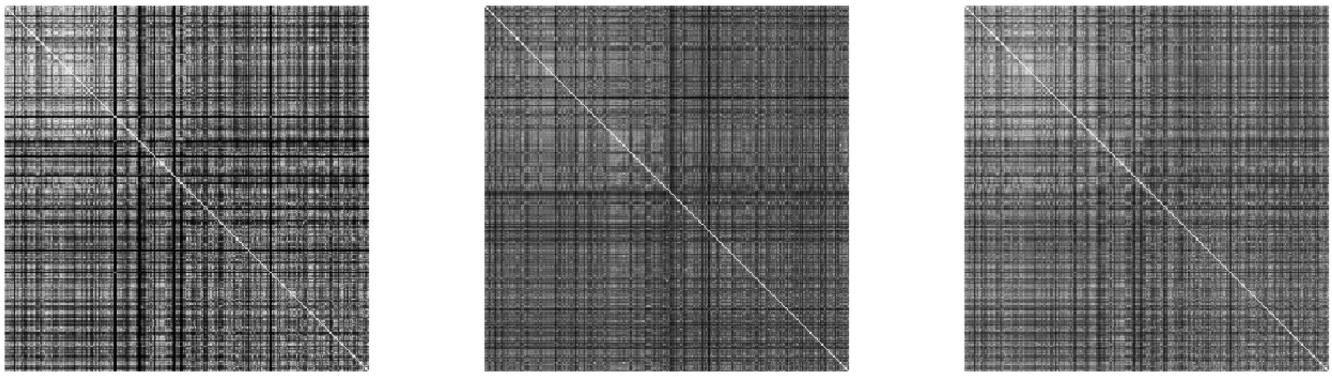



**Figure 4.**
(a) – (b) base kernels used in semidefinite programming. (c) Optimal kernel learned by semidefinite programming. In all the kernels, the first 140 samples are true polyps and other 140 samples are false positives. So the left-upper blocks show the similarities between true polyps and the right-lower blocks show the similarities between false positives. The left-upper block of (c) shows higher clustering property with high and consistent kernel values which leads to better separability.

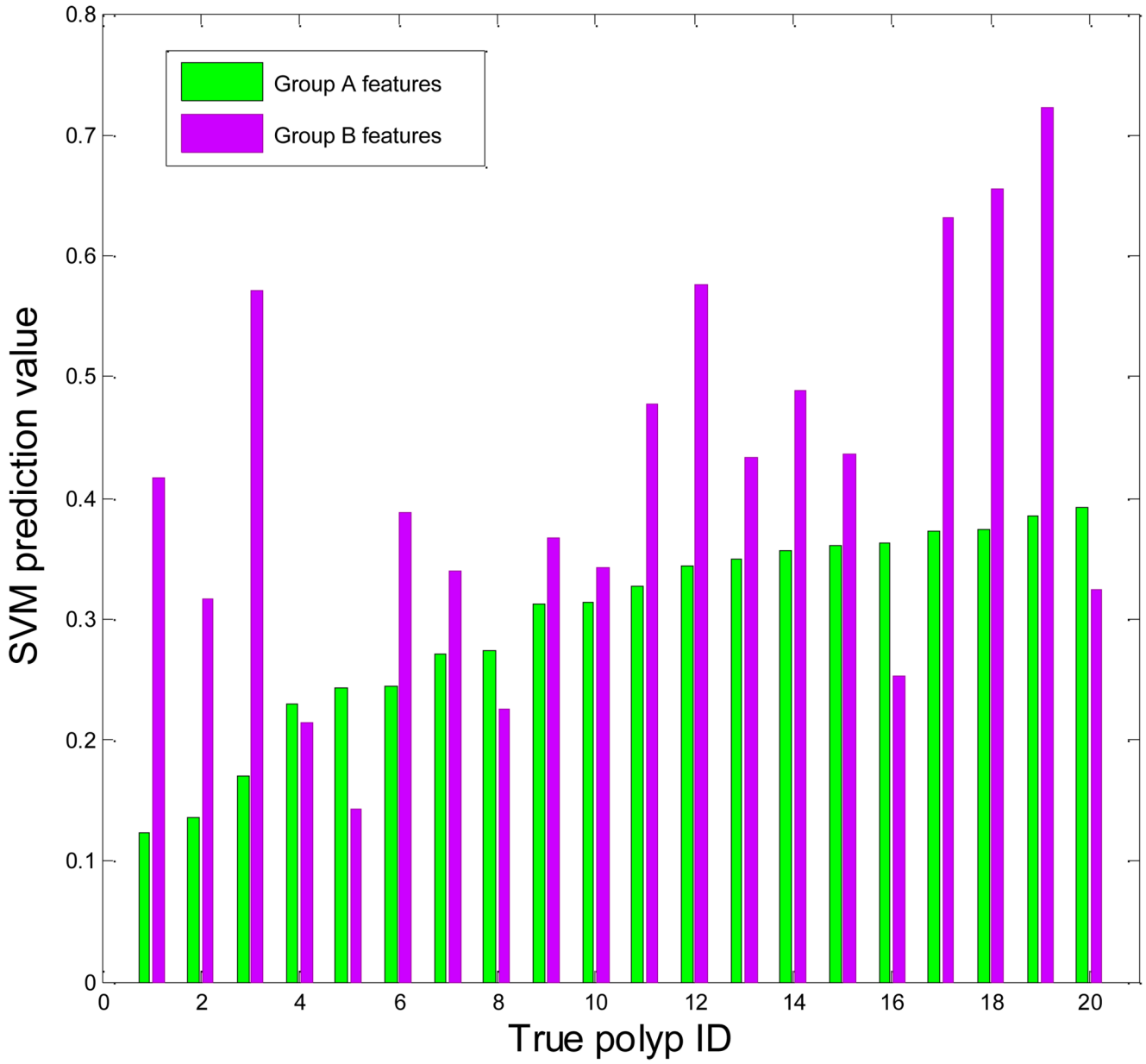



**Figure 5.**
Comparisons of SVM prediction values of 20 true polyp detections using group A features and group B features separately. This plot shows that the complementary nature between the two groups of features for a number of polyp candidates (i.e., significant difference in SVM prediction values). SVM prediction value is a measure of the likelihood that a polyp candidate is a true polyp. True polyps with low SVM prediction values mean that they are hard to be detected or similar to false positives.

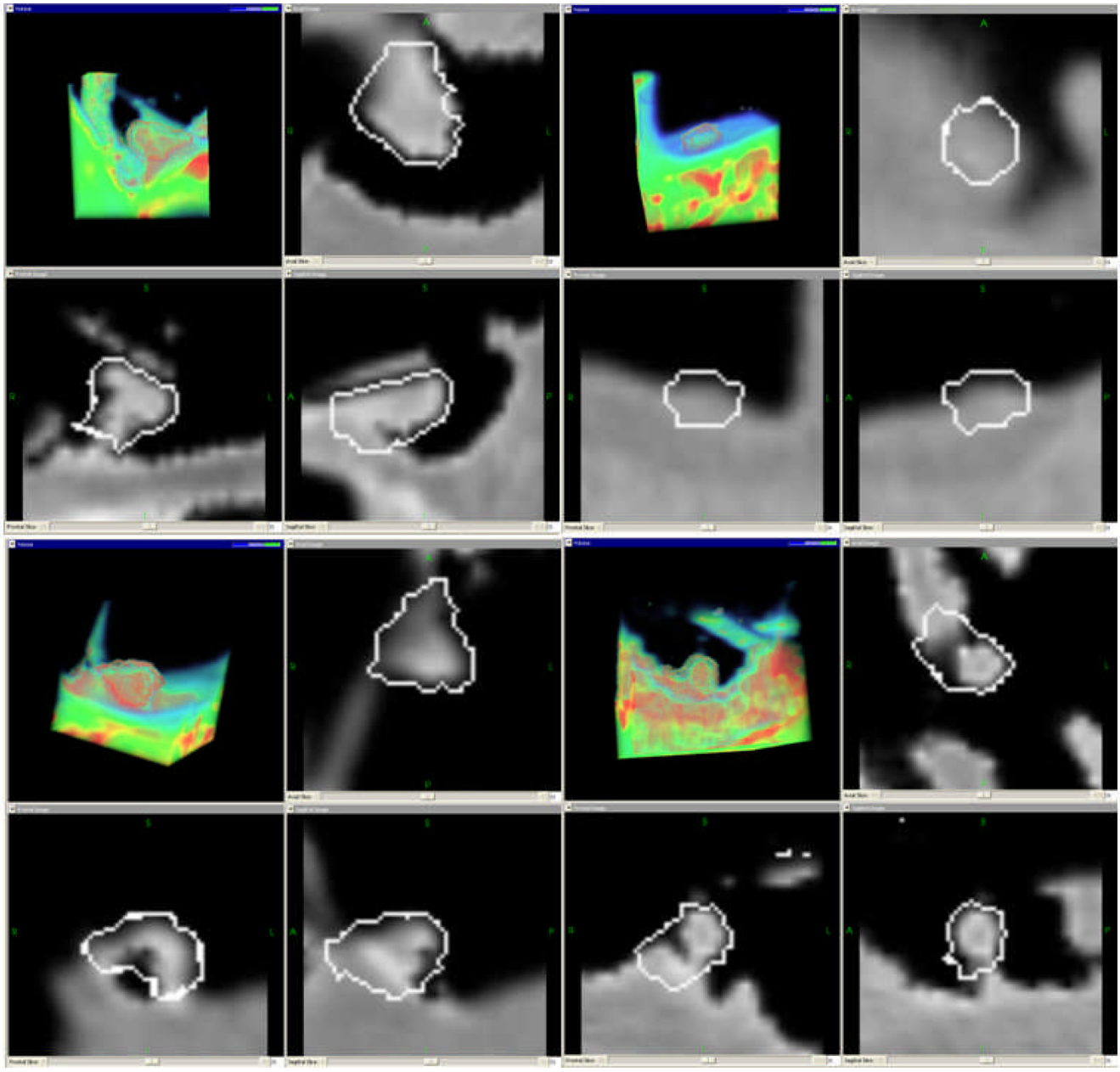


**Figure 6.**
Four true polyps considered to be false positives by SVM using group A features. SVM using HCFs gives higher prediction values and treats them as true polyps. For each polyp, we show its axial, coronal and sagittal slices and 3D endoluminal CTC volumetric reconstructions. For volumetric reconstructions, the red color corresponds to higher CT values and cyan color corresponds to lower CT values.

**Table 1**

Six curvature related features used in our HCF descriptor. Lower limits, upper limits and numbers of bins are listed for each feature and CT value.

| Feature | Lower limit | Upper limit | Number of Bins |
|---|---|---|---|
| Shape Index | 0 | +1 | 200 |
| Curvedness | 0 | +20 | 200 |
| Gaussian Curvature | −10 | +10 | 200 |
| Mean Curvature | −10 | +10 | 200 |
| Max Curvature | −10 | +10 | 200 |
| Min Curvature | −10 | +10 | 200 |
| CT Value | 0 | 1600 | 200 |

**Table 2**

18 traditional features in group A used for colonic polyp detection.

| No | Geometric Explanation |
|---|---|
| 1 | difference between max max-curvature and average max-curvature of lesion surface |
| 2 | mean curvature of lesion surface |
| 3 | (in sq cm) of all full triangles constituting surface cluster |
| 4 | width/length of bounding box containing all vertices from surface cluster |
| 5 | wall thickness measured along ray |
| 6 | std. of Gaussian curvature of vertices around the lesion |
| 7 | std of min curvature of the lesion |
| 8 | curvature relation between lesion surface and its surrounding |
| 9 | mean of wall thickness measure at different vertices |
| 10 | max curvature averaged over all vertices from neck surface cluster |
| 11 | common curvature rank between lesion surface and its surrounding |
| 12 | width/length of bounding box containing all vertices from surface cluster |
| 13 | elliptical curvature averaged over lesion segmentation |
| 14 | distance between inner and outer wall |
| 15 | 3rd order moment of given curvature/sphericity statistics |
| 16 | curvature information of voxels merged in water |
| 17 | number of segments which are part of any full triangle from surface cluster |
| 18 | Compactness: area / perimeter * 2 |

**Table 3**

Performance of state-of-the-art CTC CAD systems published in recent two years.

| Authors | Data Set Size | Polyp Size | # Polyps | Sensitivity (%) | False Positives per Patient/Scan |
|---|---|---|---|---|---|
| Li et al. [17] | 44 patients | 6–9 mm | 45 | 71 | 5.38 / patient |
| Chowdhury et al. [23] | 179 scans | >= 10 mm | 4 | 100 | 3.38 / scan |
| | | 5–10 mm | 25 | 92 | 3.38 / scan |
| | | < 5 mm | 84 | 57.14 | 3.38 / scan |
| Suzuki et al. [16] | 73 patients | >= 5 mm | 28 | 96.4 | 1.1 / patient |
| van Ravesteijn et al. [18] | 86 patients | >= 6 mm | 59 | 95 | 5 / scan |
| | 48 patients | >= 6 mm | 28 | 85 | 4 / scan |
| | 141 patients | >= 6 mm | 176 | 85 | 5 / scan |
| | 32 patients | >= 6 mm | 8 | 100 | 6 / scan |
| Zhu et al. [15] | 26 patients | >= 5 mm | 32 | 100 | 3.5 / scan |
| Our method | 109 scans | >= 6 mm | 96 | 83 | 5 / scan |