

A Statistical 3-D Pattern Processing Method for Computer-Aided Detection of Polyps in CT Colonography

Salih Burak Göktürk*, Carlo Tomasi, Burak Acar, Christopher F. Beaulieu, David S. Paik, R. Brooke Jeffrey, Jr., Judy Yee, and Sandy Napel, *Member, IEEE*

Abstract—Adenomatous polyps in the colon are believed to be the precursor to colorectal carcinoma, the second leading cause of cancer deaths in United States. In this paper, we propose a new method for computer-aided detection of polyps in computed tomography (CT) colonography (virtual colonoscopy), a technique in which polyps are imaged along the wall of the air-inflated, cleansed colon with X-ray CT. Initial work with computer aided detection has shown high sensitivity, but at a cost of too many false positives. We present a statistical approach that uses support vector machines to distinguish the differentiating characteristics of polyps and healthy tissue, and uses this information for the classification of the new cases. One of the main contributions of the paper is the new three-dimensional pattern processing approach, called random orthogonal shape sections method, which combines the information from many random images to generate reliable signatures of shape. The input to the proposed system is a collection of volume data from candidate polyps obtained by a high-sensitivity, low-specificity system that we developed previously. The results of our ten-fold cross-validation experiments show that, on the average, the system increases the specificity from 0.19 (0.35) to 0.69 (0.74) at a sensitivity level of 1.0 (0.95).

Index Terms—Computer aided diagnosis, CT colonography, pattern recognition, random orthogonal shape section (ROSS) method, support vector machines classification.

I. INTRODUCTION

COLON cancer is the second leading cause of cancer deaths in the United States. American adults have 1/20 chance of developing and 1/40 chance of dying from this disease. There are approximately 150 000 new cases diagnosed each year resulting in 56 000 deaths [1]. Previous research has shown that adenomatous polyps, particularly those larger than 1 cm in diameter, are the most likely precursor to subsequent colorectal carcinoma [2]. The National Polyp Study clearly illustrated that

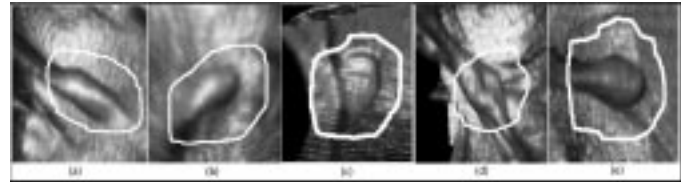


Fig. 1. (a)–(c) Examples of polyps. (d)–(e) Examples of healthy tissue that have similar shapes.

colonoscopic removal of all identifiable polyps resulted in a decline in mortality rate between 76% and 90% compared with historical controls [3]. Unfortunately, colon cancer is most often discovered after the patient develops symptoms, and by then, the likelihood of a cure has diminished substantially.

Fiberoptic colonoscopy (FOC) is considered the definitive diagnostic test [4] for the presence of colonic polyps as it affords direct visualization and the opportunity for biopsy or removal of suspicious lesions [3]. However, FOC is not feasible as a population screening test due to cost, the small but real risk of complications such as perforation, and due to the fact that there are not sufficient endoscopists in the country to accommodate all patients [4]. Moreover, the majority of colonoscopic examinations performed in the United States are negative for polyps or masses, therefore, a less invasive, more widely available procedure that is also acceptable to patients is attractive.

Computed tomography colonography (CTC) (virtual colonoscopy) is a recently proposed noninvasive technique that combines spiral CT data acquisition of the air-filled and cleansed colon with three-dimensional (3-D) imaging software to create virtual endoscopic images of the colonic surface [5]–[8]. The initial clinical results are quite promising [9]–[15], yet the technique is still impractical due, in part, to the time required to review hundreds of images/patient study. This limitation begs for a computer-aided detection (CAD) method to help the radiologist detect polyps efficiently from the acquired CTC data.

Identifying colonic polyps using CAD is very challenging because they come in various sizes and shapes, and because thickened folds and retained stool may mimic their shape and density. Fig. 1 demonstrates the appearance of polyps and other tissue as they appear in a virtual colonoscopy study.

Initial studies describing CAD for polyp detection have focused on shape analysis and started from the intuitive observation on the similarity of the polyp shape to hemispheres. Summers *et al.* characterized the colon wall by computing its

Manuscript received June 8, 2001; revised October 31, 2001. This work was supported in part by National Institutes of Health (NIH) under Grant 1R01 CA72023, in part by the Lucas Foundation, in part by the Packard Foundation, and in part by the Phil N. Allen Trust. *Asterisk indicates corresponding author.*

*S. B. Göktürk is with the Department of Electrical Engineering, Stanford University, Robotics Laboratory, Gates Building, Stanford, CA 94305-9010 USA (e-mail: gokturkb@stanford.edu).

C. Tomasi is with the Department of Computer Science, Stanford University, Stanford, CA 94305-9010 USA.

B. Acar, C. F. Beaulieu, D. S. Paik, R. B. Jeffrey, Jr., and S. Napel, are with the Department of Radiology, Stanford University, Stanford, CA 94305 USA.

J. Yee is with the San Francisco VA Medical Center, the University of California, San Francisco, CA 94143 USA.

Publisher Item Identifier S 0278-0062(01)11212-7.

minimum, maximum, mean, and Gaussian curvatures [16]. Following discrimination of polypoid shapes by their principal minimum and maximum curvatures, more restrictive criteria such as sphericity measures are applied in order to eliminate nonspherical shapes. In [17], Yoshida *et al.* use topological shape of vicinity of each voxel, in addition with a measure for the shape curvedness to distinguish polyps from healthy tissue. Gokturk and Tomasi designed a method based on the observation that the bounding surfaces of polyps are usually not exact spheres, but are often complex surfaces composed of small, approximately spherical patches [18]. In this method, a sphere is fit locally to the isodensity surface passing through every CT voxel in the wall region. Groups of voxels that originate densely populated nearby sphere centers are considered as polyp candidates. Obviously, the clusters of the sphere centers are more dense when the underlying shape is a sphere or a hemisphere. In [19] and [20], Paik *et al.* introduced a method based on the concept that normals to the colon surface will intersect with neighboring normals depending on the local curvature features of the colon. Their work uses the observation that polyps have 3-D shape features that change rapidly in many directions, so that normals to the surface tend to intersect in a concentrated area. By contrast, haustral folds change their shape rapidly when sampled transversely, resulting in convergence of normals, but change shape very slowly when sampled longitudinally. As a result, the method detects the polyps by giving the shapes a score based on the number of intersecting normal vectors. This score is higher in hemispherical polyps compared with folds.

While these methods have demonstrated promising sensitivity, they can be considered more as polyp candidate detectors than polyp detectors because of their large number of false positive detections. These methods have observed that polyps have spherical shapes, and provided different measures of sphericity. However, polyps span a large variety of shapes, and fitting spheres alone is not an accurate measure. This paper presents a statistical method to differentiate between polyps and normal tissue. Our new 3-D pattern processing method, called random orthogonal shape section (ROSS) method generates shape-signatures for small candidate volumes, which might be obtained by one of the above mentioned methods, and then feeds these signatures to a support vector machines (SVM) classifier for the final diagnosis of the volume. By using statistical learning, the different characteristics of polyps and normal tissue are automatically extracted from the data.

The paper is organized as follows: Section II describes our proposed technique, explaining both the pattern processing and the SVM classifier, and our evaluation methods. Section III presents our results, followed by discussion in Section IV and conclusions in Section V.

II. METHODS

Fig. 2 gives an overview of our system, which consists mainly of two components: A novel 3-D pattern processing approach (ROSS) for feature analysis and a SVM classifier [21]. These two components are discussed in Sections II-A and II-B, respectively.

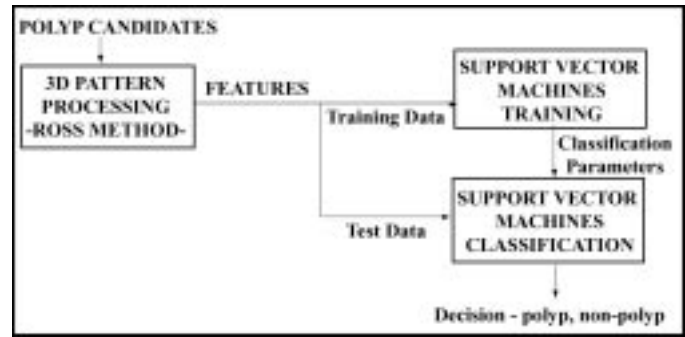


Fig. 2. Overview of the system.

A. ROSS Method

Fig. 3 gives an overview of the ROSS method. Here, we mimic one of the ways radiologists view CT images of the colon, i.e., by viewing images parallel to the axial, sagittal, and coronal planes [8], [15] (Fig. 4). While viewing these three planes gives substantial information about the 3-D shape, the information is incomplete. In our approach, we first obtain a large number of *random* triples of mutually *orthogonal sections* through a suspicious subvolume. Because other structures may co-occupy the resulting images, we next segment them to isolate the putative polyp for further processing. We next calculate a set of geometric attributes from each random plane, and generate a histogram of these geometric attributes obtained from each triple as a feature-vector to represent the shape. Accumulating histograms of these attributes over several random triples makes the resulting signatures invariant to rotations and translations. Careful choice of geometric attributes results in robust signatures. Details of the image segmentation and image-based geometric attributes are given in Sections II-A-1 and II-A-2, respectively. The histogram processing is discussed in Section II-A-3.

In the ROSS method, we have chosen to sample through the 3-D shape in the form of two-dimensional (2-D) sections. The statistics of the geometric information on these triples of sections potentially represent many variations of the shape. In order to cover as many variations by using 3-D data directly, one needs to fit prohibitive amount of ellipsoids or other primitive objects to the shape. Alternatively, one could use the histogram of local structures such as the curvature at each voxel. However, this choice is not diagnostic enough since different shapes might often have similar distribution of local information. As discussed in Section II-A-2 and II-A-3, the ROSS method goes through random samples of mutually orthogonal 2-D sections and does not involve these problems, thus, is an efficient and effective way to process the 3-D data.

1) *Image Segmentation:* In each image we first construct a square *critical window* located in the center of the image, surrounding the suspicious structure and excluding as much of the extraneous tissue as possible. The size of the critical window varies depending on the size and the shape of the suspicious structure in the image. A window is considered good when it contains a shape that is approximately circular and has a small circular radius. Because elongated folds are common in the colon, it was found to be useful to also explicitly rule out

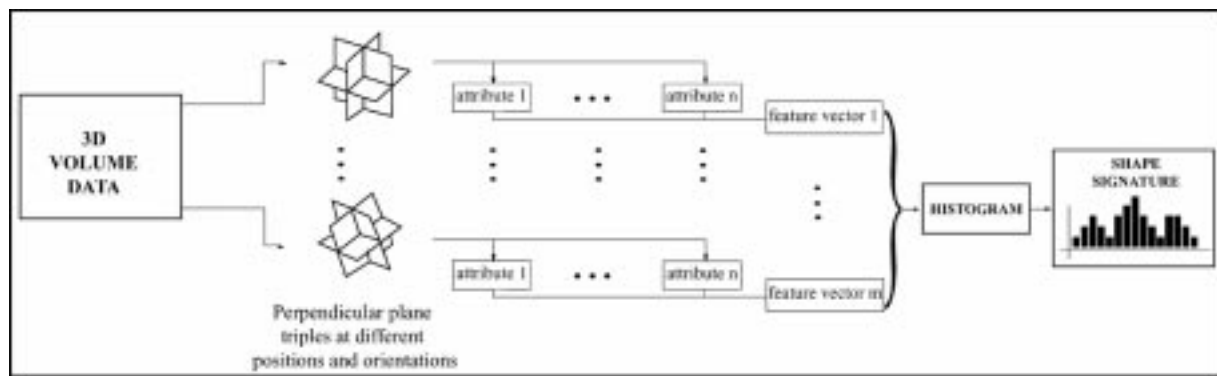


Fig. 3. Overview of the ROSS method.

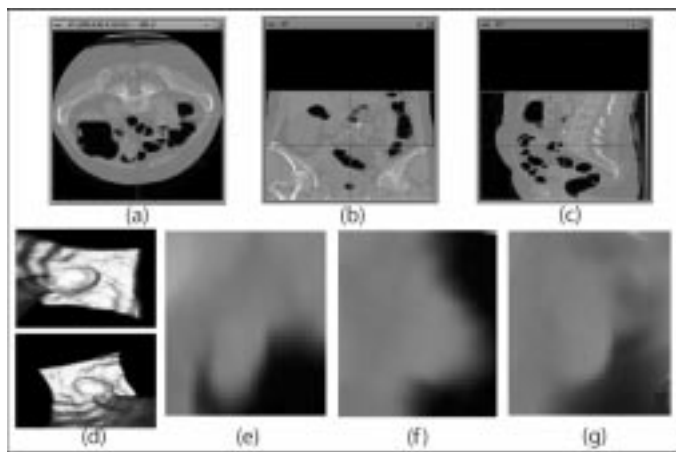


Fig. 4. (a)–(c) Three cross sections of the body through the anatomical directions: (a) axial; (b) coronal; and (c) sagittal. (d) Two 3-D renderings of a polyp with different camera views. (e)–(g) Randomly oriented, mutually orthogonal triple of planes through this polyp.

elongated regions when there is a circular region in the image. To find a critical window’s size, an image is first binarized by intensity thresholding, and the following target function is computed for each window of size w centered in the middle of the image:

$$f(w) = a_1 r(w) + e_{\text{circle}}(w) - a_2 e_{\text{line}}(w). \quad (1)$$

Here, $r(w)$ is the radius of the circle that best fits the edges in the binarized image, $e_{\text{circle}}(w)$ is the residue to the best fitting circle, $e_{\text{line}}(w)$ is the residue to the best fitting line all in the subwindow w , and a_1 and a_2 are constants. Details of the shape fitting are described in the next section. w is varied from 8 mm to 24 mm and the value of w that yields the smallest $f(w)$ is chosen as the critical window size to calculate the geometric attributes. Fig. 5 depicts the resulting critical window for several different images.

In some cases, minimizing (1) may result in a critical window that includes some portion of the surrounding colon wall. To further accommodate these cases, a Gaussian weighting function, located at the image center, is used while fitting primitive shapes. By weighting with this Gaussian mask, we aim to give more importance to boundary points of a potential polypoid shape than to those of the surrounding colon wall.

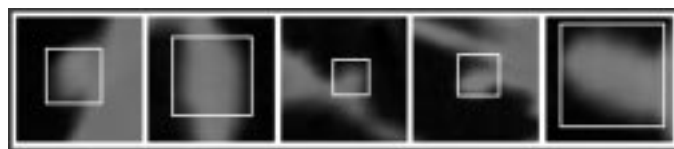


Fig. 5. Optimum segmentation window in various images.

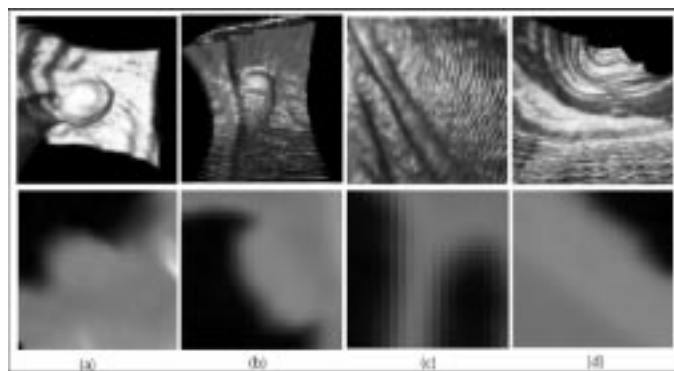


Fig. 6. Volume renderings of different shapes (top row). Bottom row gives a random cross section through these shapes. These images are specific examples where fitting (a) a circle, (b) an ellipse, (c) parallel lines, and (d) a line produces strong output about the 3-D shape.

2) *Image-Based Geometric Features*: Once we find the optimal critical subwindow in a particular image, we extract shape and intensity related geometric features in the critical window. For this purpose, we fit primitive shapes such as circle, quadratic curve, and line to the largest connected edge component, i.e., boundary, of the shape. The boundary in each random plane is determined by a 2-D Sobel derivative operation [22] followed by thresholding. The mathematical details of the primitive shape fittings are given in the Appendix.

A random slice of a sphere is a circle. Thus, fitting circles is a means of measuring the sphericity of the 3-D shape [Fig. 6(a)]. When doing so, the residuals at each pixel on the boundary are first weighted with a Gaussian located at the image center as mentioned in the previous Section. The Gaussian weighted residuals are then added together, and the least-square solution gives the optimum circle. The residual to the least-square solution is recorded as well.

Similarly, the ellipsoidal nature of the shape can be measured by fitting a quadratic to the edges. This way, we can capture

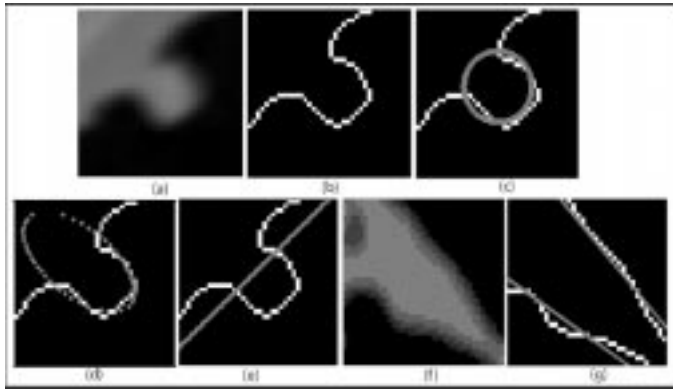


Fig. 7. Illustration of the primitive shapes: (a) a randomly oriented plane; (b) the edges in the image; (c)–(e) circle, quadratic, and line fit to the edges shown in (b). (f) Another randomly oriented plane with nearly parallel lines (g) parallel lines fit to the edges of the plane in (f).

the similarity of the shape to a pedunculated polyp [Fig. 6(b)]. The quadratic invariants given in the Appendix, as well as the residual to the least-square solution are recorded.

The cross section of a fold on the image plane may contain approximately parallel lines [Fig. 6(c)]. In order to capture this structure, we apply parallel lines analysis, which includes fitting lines to the two largest connected components of the boundary points. We record the residual to these lines and the distance and angle between these approximately parallel lines. Similarly, the residual to the optimum fitting line gives information on the flatness of the surface [Fig. 6(d)]. Fig. 7(a) and (b) gives an example of a randomly oriented plane and the edges on this plane, respectively. Fig. 7(c)–(e) illustrates three primitive shapes, circle, quadratic and line being fit to this random plane. Fig. 7(f) shows another random plane going through a thick fold and Fig. 7(g) demonstrates two approximately parallel lines being fit to its edges.

In order to extract information on higher order frequency characteristics of the boundary, third-order moment invariants are computed as well [23]. This gives information on the curviness of the boundary points.

In addition to all of these shape-based features, intensity features are extracted from the tissue part of the image. For this, the tissue part is first separated away from the background by intensity thresholding and, next, the intensity mean and standard deviation of the tissue are calculated and recorded.

All the attributes mentioned so far are calculated for each random triple of images. The three images in each triple are sorted in the order of increasing radius of curvature, and the features above are listed into a vector in this order. The resulting N -vector containing the N attributes from the three planes represents the signature of the shape, relative to that particular triple of perpendicular planes.

3) *Obtaining the Histograms in High Dimensional Space:* We next obtain histograms of feature vector distributions over the randomly oriented triples. Given enough triples, these histograms are essentially invariant to orientation and position. Note, however, that creating histograms of N -vectors is problematic for large N . Uniform distribution of, say, m bins/dimension would result in m^N -dimensional histograms and is, therefore, not tractable. A more efficient solution,

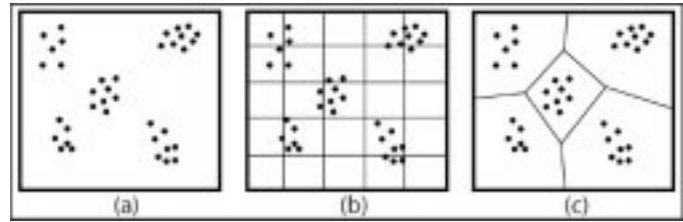


Fig. 8. Illustration of high dimensional histogram bins by an example in 2-D. (a) Data in two dimensions. (b) Uniform histogram bins. (c) Bins found when clustering is applied. The storage gain (number of bins) is fivefold in this particular example.

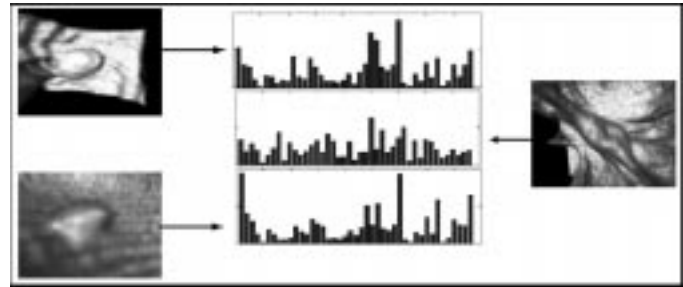


Fig. 9. Two polyps (left), a very similarly shaped fold structure (right), and their signatures (center).

proposed here, represents a histogram by first computing the representatives for the main clusters of features over a large collection of vectors. New feature vectors are then assigned to these clusters, rather than to fixed uniform bins. This method is called vector quantization and is described in more detail in [24]. Fig. 8 gives an example in 2-D space. Note that, storage efficiency increases exponentially as the number of dimensions increase beyond two.

In our implementation, suppose X_{ij} is the N -vector obtained from the j th random triple of perpendicular planes extracted from the i th shape in a training set. We obtain X_{ij} s from all of the shapes, and then invoke the k -means algorithm [24], [25] to compute vector clusters. The cluster centers are first initialized to a random selection from X_{ij} s. Subsequent iterations of the k -means algorithm then alternately reassign vectors to clusters and recompute cluster centers, resulting eventually in the optimum cluster centers.

Once the representative histogram bin centers are determined, a histogram of feature vectors belonging to each particular shape is calculated. When forming these feature histograms, the simplest strategy would be to have each vector increment a counter for the nearest cluster center. This method, however, is overly sensitive to the particular choice of clusters. We adopted a more robust solution, in which each feature vector partitions a unit vote into fractions that are inversely proportional to the vector's distances to all histogram bin centers. The histograms, thus, obtained, one/candidate volume, are the rotation and translation invariant shape signatures used for classification as described in the next section. Fig. 9 illustrates examples of signatures obtained for three different shapes.

B. SVM Classification

Given the representative vectors for polyps and healthy tissue, we computed the optimum classifier using training data and sub-

sequently used that classifier on test data. In our problem, a training set is a collection of candidate volumes that are classified by FOC and CT colonoscopy to be polyps or nonpolyps.

Proposed first by Vapnik, the SVM classifier aims to find the optimal differentiating hypersurface between the two classes [26] in the training set. Generally speaking, the optimal hypersurface is the one that not only correctly classifies the data, but also maximizes the margin of the closest data points to the hyperplane. Mathematically, we consider the problem of separating the training set S of points $x_i \in R^n$ with $i = 1, 2, \dots, N$. Each data point x_i belongs to either class and, thus, is given a label $y_i \in \{-1, 1\}$. SVM [26]–[28] implicitly transform the given feature vectors x_i into new vectors $\phi(x)$ in a space with more dimensions, such that the hypersurface that separates the x , becomes a hyperplane in the space of $\phi(x)$ s. Thus, the linearly nonseparable data in the original space becomes linearly separable in a higher dimension. Finding this hypersurface is a well-studied optimization problem [26]–[28] that maximizes the distance of the hypersurface to the margin points, so called the support vectors. Support vectors are essentially the elements that carry the differentiating characteristics between polyps and nonpolyps. In the classification process of a test vector x , only the support vectors need to be considered, and the following simple expression needs to be calculated:

$$d(x) = \sum_{x_i \in SVs} \alpha_i y_i K(x_i, x) + b \quad (2)$$

where the constants α_i, b are computed by the classifier-learning algorithm, and the x_i are the computed support vectors. K is called the Kernel function, and is used to map the data from its original dimension to higher dimensions so that the data is linearly separable in the mapped dimension. See [27] for details. Given the expression $d(x)$, the decision is given based on the sign of $d(x)$

$$\text{Decision} = \begin{cases} \text{polyp,} & \text{if } d(x) \geq 0 \\ \text{nonpolyp,} & \text{if } d(x) < 0. \end{cases} \quad (3)$$

In clear contrast with the previous classification methods, such as linear discriminant analysis (LDA), or k nearest neighbor methods that minimize the empirical risk—risk of misclassification in the training set—SVMs minimize the structural risk, given as the probability of misclassifying previously unseen data, by utilizing the optimal hyperplane between two classes of data. In addition, SVMs pack all the relevant information in the training set into a small number of support vectors and use only these vectors to classify new data. This way, the distinguishing features of the two classes are implicitly identified while the detracting less important points are discarded. In other statistical classification methods such as LDA or neural networks, the classification surfaces are obtained by taking all of the data points into account, which might statistically deteriorate the performance of these methods.

C. Evaluation

We tested our system using data acquired from 48 patients (45 male, three female, ages 27–86, mean age 60 ± 12) in the supine

or prone position. Following colon cleansing and air-insufflation, CT imaging was applied using either a single detector (or multidetector) CT system (GE Medical Systems, Milwaukee, WI) with the imaging parameters of 3 mm (2.5 mm) collimation, 2.0 (3.0) pitch, 1.0–1.5 mm reconstruction interval, 120 KVp (both SDCT and MDCT), 200 mA (60 mA). The resulting images had an average voxel spacing of $0.74 \times 0.74 \times 1.28$ mm, with a total data size of $512 \times 512 \times N$ where N is the number of slices for a particular patient.

The 3-D CT data was preprocessed by the Hough transform (HT) method as described in [19]. The outputs of this stage are the locations of densely populated surface normal intersections along with their particular HT score. This yielded a total of 31 099 hits of score 3000 or above, on the 48 patient datasets. For computational efficiency, we kept all of the polyp hits, but subsampled the nonpolyp (healthy tissue) hits, by randomly selecting 50 nonpolyp hits/dataset. We extracted a subvolume of size $32 \times 32 \times 32$ from each location, and subsequently used it as a candidate volume for the post processing stage. Some of the hits were close to the boundary of the dataset, and it was not possible to get a $32 \times 32 \times 32$ subvolume around them. Thus, these hits were eliminated, eventually resulting in 40 polyp (sizes from 2 mm to 15 mm in diameter) and 2219 nonpolyp hits. Subvolumes centered at each HT hit were used as input candidate data for our system.

Because the performance characteristics of our methods depend on accurate differentiation of polyp and nonpolyp subvolumes, here we describe how the polyp subvolumes were selected. Each patient in the study underwent FOC within 1–2 hours after acquisition of CTC data. To determine the gold standard for polyp versus nonpolyp volumes, a radiologist highly experienced in CTC reviewed the CT data with knowledge of the findings on FOC. This involved interactive viewing of the CTC data on a workstation including multiplanar reformations and selected 3-D viewing while searching for all polyps reported on FOC according to their size and location. A candidate volume was declared a polyp when the reviewer could confidently identify the polyp on CTC. If a polyp was reported on FOC but not identified on CTC, the radiologist viewed the CTC data in the opposite (supine or prone) patient acquisition to determine if it was not visible due to obscuration by incomplete distension or excess fluid in the colon. When such a polyp could not be located in the test dataset, this polyp was excluded from further analysis. In the case of polyp candidates identified by the Hough transform but not corresponding to FOC abnormalities, these were classified as nonpolyp volumes. Examples of nonpolyp volumes included normal colon, thickened haustral folds, and foci of retained stool. In cases where the radiologist identified possible polyps but there was no corresponding lesion on FOC, these areas were considered as potential false positive CTC identifications and excluded from the study. We note that the radiologist reviewer did not utilize the HT hits in reviewing the CTC data but relied on reading experience and the FOC results alone.

We applied the ROSS method to each shape in the candidate volume dataset. One hundred random triples of mutually perpendicular images were extracted from each candidate volume. Each triple yielded a 26-vector which measures the following

features on the random planes: Best-fit circle's radius, residue to the best-fit circle, line and quadratic curve, quadratic shape invariants, moment invariants, angle, and distance between nearly parallel lines (if there are such pair of lines in the image), and total residue to line fit in the pair of parallel lines. The distribution of information belonging to the 26-vector of 100 triples is estimated by high dimensional histograms as described in Section II-A-3. The k -means algorithm is executed to obtain 45-vector centers which are then used as histogram bin centers. Therefore, the high dimensional histograms obtained by vector quantization resulted in a 45-dimensional vector to represent each particular candidate volume. These vectors were used by the SVM classification algorithm with exponential radial basis functions as kernel functions [29].

We conducted a tenfold cross-validation experiment to evaluate our system. First, we randomly divided the candidate volumes into ten uniformly distributed sets of ten polyps (with repetitions, average overlap between sets is 1.8 polyps, maximum overlap is four polyps) and 221 or 222 nonpolyps (healthy tissue, without repetitions). In each of these ten experiments, one set was used as the test set, and the remaining 30 polyps and 1997 (1998) nonpolyps as training set. In our analysis, we will refer to sensitivity as the fraction of detected polyps, and to specificity as the fraction of detected normal tissue volumes among the nonpolyp locations.

To examine the tradeoff between the sensitivity and specificity more quantitatively, we substituted zero-crossing sign in (3), with a level crossing (SVM threshold, D). As the level is decreased, more true polyps are detected, but at a cost of more false positives. In this way, for any input collection of polyps and nonpolyps, we can obtain a receiver operating characteristic (ROC) curve [sensitivity versus (1-specificity) curve]. Similarly, the HT method uses a threshold, h , that denotes the minimum score necessary to declare a candidate volume to be a polyp. Thus, by varying h , one can also construct an ROC curve.

Our evaluation attempts to simulate the cascade in which HT and SVM are actually meant to be used. In this analysis, we simulated a real system, where, the HT method is first executed to achieve a predetermined sensitivity level, and the SVM method is executed afterwards to improve the specificity without sacrificing the sensitivity. To simulate this cascade system, we constructed a multidimensional ROC surface for each experiment. As shown in Fig. 10(a), two planar axes give the specificities of HT alone and the cascade system (ALL), respectively, and the height gives the sensitivity. In order to construct this surface, the HT hits with a score higher than a threshold, h , were subsequently classified using an SVM threshold, D . Varying h relates sensitivity (SnS) to HT specificity (SpS_{HT}) resulting in ROC_{HT} . Observe that the projection of this curve to the $SnS = 0$ plane results in the Π_{HT} curve on the plane. For a fixed HT threshold, h , varying D relates the sensitivity to specificity of the cascade system (SpS_{ALL}). This analysis is repeated for every (h, D) pair resulting with the ROC surface, every point of which denotes the sensitivity versus specificity for the cascade system. By definition, for every HT threshold, h , the SVM is executed to obtain the maximum specificity without sacrificing the sensitivity obtained by HT. The combination of the maximum specificity points obtained for every HT threshold, h , gives the ROC curve of the cascade system and de-

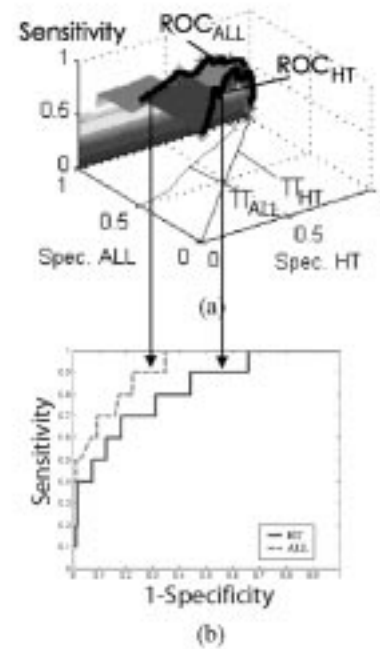


Fig. 10. Illustration of an ROC surface example and how to obtain the ROC curves from the surface.

noted by ROC_{ALL} . The projection of ROC_{ALL} to the $SnS = 0$ plane results in the Π_{ALL} curve on the plane. The differences of the Π_{HT} and Π_{ALL} curves inherently show the improvements by the cascade system over HT alone. To illustrate the ROC curves better, we plotted the ROC_{ALL} and ROC_{HT} curves in 2-D in Fig. 10(b). Our results report the resulting ROC curves for each of our ten experiments.

To assess the statistical significance of the improvements for a specific sensitivity level, we tested the null hypothesis that the average difference between the specificities obtained by HT and cascade of HT and our method is zero. For this we applied t -tests on ten values resulting from the ten experiments and calculated the p values for each sensitivity level.

III. RESULTS

Fig. 11 gives the ROC curves obtained from each of the ten experiments. There is a considerable improvement in specificity between the ROC curve of HT alone and that of the cascade system in each of the experiments. Table I summarizes the specificities obtained for sensitivity levels of 1.00 and 0.95 for each experiment (as read from the curves in Fig. 11) by our approach and the HT approach alone, along with the areas under these curves. The improvement in specificity at sensitivity level of 1.00 varied from 0.27 (Experiment #1) to 0.73 (Experiment #5) and at sensitivity level of 0.95 varies from 0.20 (Experiment #1) to 0.55 (Experiment #5). On average, the specificity of the HT was 0.19 (0.35) for a sensitivity level of 1.00 (0.95). Observe that, on average, the specificity was increased by 0.50 and 0.39 at sensitivity levels of 1.00 and 0.95, respectively, to achieve final specificity values of 0.69 and 0.74. Similarly, the areas under the curves increased from 0.80 to 0.92 on average.

We observed that the specificity level for a sensitivity of 0.28 or less was same for both HT and the cascade system and 1.00.

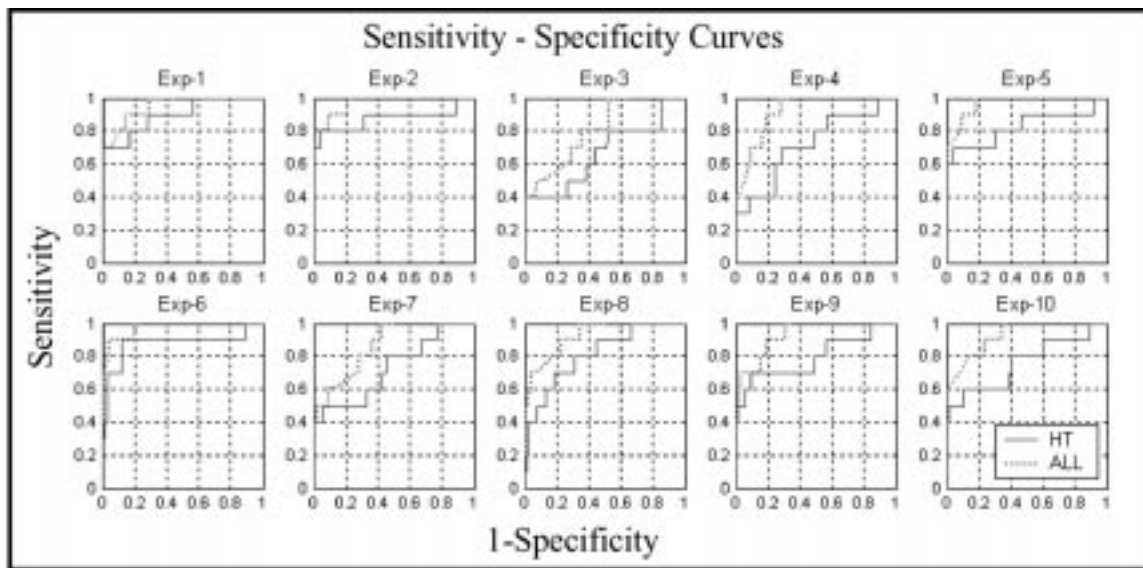


Fig. 11. The results of each of the 10-experiments, showing ROC curves for HT alone and HT followed by our method.

TABLE I

ROC AREAS (Az) AND THE SPECIFICITY (SPS) VALUES AT 100% AND 95% SENSITIVITY (SNS) LEVELS FOR THE HT AND THE TWO-STAGE METHOD (HT FOLLOWED BY OUR METHOD; DENOTED AS ALL) IN EACH OF TEN CROSS-VALIDATION EXPERIMENTS. THE SPECIFICITY VALUES OF THE TWO-STAGE METHOD ARE THE VALUES AT THE OPERATING POINT SET BY HT THAT WERE ACHIEVED WITHOUT SACRIFICING THE SENSITIVITY

#	Az_{HT}	Az_{ALL}	SpS@100% SnS		SpS@95% SnS	
			HT	ALL	HT	ALL
1	0.90	0.95	0.44	0.71	0.58	0.78
2	0.88	0.97	0.11	0.79	0.40	0.85
3	0.67	0.81	0.15	0.48	0.15	0.48
4	0.72	0.92	0.11	0.72	0.27	0.76
5	0.83	0.97	0.09	0.82	0.31	0.86
6	0.88	0.97	0.11	0.80	0.50	0.88
7	0.73	0.86	0.23	0.58	0.28	0.61
8	0.82	0.93	0.34	0.66	0.45	0.72
9	0.80	0.93	0.16	0.69	0.30	0.75
10	0.76	0.93	0.11	0.66	0.26	0.71
Mean	0.80	0.92	0.19	0.69	0.35	0.74
s.d.	0.08	0.05	0.11	0.11	0.12	0.12

The t -tests for sensitivity level of 0.46 or more resulted in p values 0.05 or less. Thus, we reject the null hypothesis and thereby accept that the improvements obtained by our system are statistically significant for sensitivity levels of 0.46 or more.

The current implementation of the algorithm is in Matlab 6.0 (The Mathworks Inc., Nattick, MA) and requires an average of 200 ms to process one triple of mutually orthogonal planes with a Pentium III 600-MHz processor PC. If 100 of triples are used, this amounts to 20 s to process each candidate volume. We also observed that the classification time is on the order of milliseconds, and is, therefore, not of practical importance when compared with the time it takes for pattern processing.

IV. DISCUSSION

Previous methods in polyp recognition suffer from low specificity mainly since they incorporate simple intuitions about the

shapes of polyps and nonpolyps. Polyp recognition is a difficult problem, and so is the manual identification of discriminating criteria between polyps and healthy tissue. Using statistical pattern processing, on the other hand, these criteria are identified automatically from training data, thereby eliminating the guess-work for obtaining the distinguishing features. More specifically, using SVM as classifiers uses the training data to extract the optimum discriminating hypersurface between the two classes and, therefore, minimizes the probability of misclassifying previously unseen (test) cases.

In addition, and in clear contrast with previous methods, the ROSS method does not explicitly make any assumptions about the polyp shape. By choosing reasonable planar attributes, it captures many characteristics of shapes accurately. Consequently, the approach is applicable to other 3-D pattern recognition problem and also to the discovery of any new polyp shapes. The generality is assured by including the new shape samples to the training set of the algorithm, and designing new planar attributes if necessary.

Our algorithm does not assume anything about the nature of the preprocessing algorithm. Any high sensitivity, low specificity preprocessing algorithm is a potential preprocessing method for our algorithm. Our experiments show that the algorithm improves the specificity for sensitivity level of 1.00 by 0.50 from 0.19 to 0.69 on average. This amounts to 62% fewer false positive detections without sacrificing sensitivity. This, in turn, results in a decrease in the radiologist's post-CAD examination time.

In the training process, the SVM classifier determines the support vectors, i.e., the data points that carry the differentiating characteristics of the two classes. Understanding the resulting support vectors could give valuable feedback for designing new features for the system. More specifically, the clusters of closeby support vectors from the two classes belong to similar shapes in polyps and nonpolyps. Further analysis of these closeby support vectors will help observe the distinguishing features of similar polyps and nonpolyps.

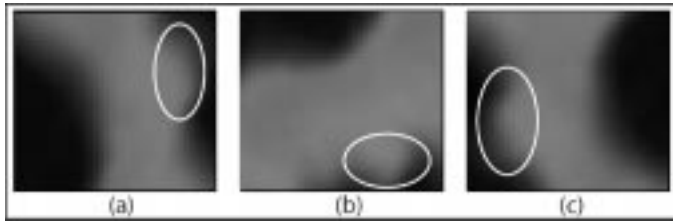


Fig. 12. Illustration of a polyp (ellipse) that has a HT hit displaced from the actual location. As seen on all three images, the edge of the fold is much closer to the center of the image; thus, in this case the fold was processed instead of the actual polyp.

In this paper, we have provided a cross-validation experiment to test our method. The most straightforward approach to a cross-validation study is to use nonoverlapping test sets in each experiment. However, this would result in test sets that contain four polyps in our case. Since we want to report the specificity at 95% sensitivity level, we would have to interpolate between the specificities at 75% and 100% sensitivity levels for obtaining the specificity at 95% sensitivity level. This would involve a large amount of bias toward the specificity level at 100%, and would not be accurate. Therefore, we chose to repeat the experiments with overlapping test sets of ten polyps each, and reduce the interpolation bias to obtain the specificity at 95% sensitivity level. As for the experiment with overlapping test sets, each fold of the tenfold cross-validation study is valid since there is no overlap between the test sets and the training sets. Finally, we report the ROC curves for all of the ten experiments, but not an average ROC curve. Therefore, we evaluate each experiment independently, and thereby avoid any bias that could occur in an average ROC curve.

As shown in the previous section, the improvements in specificities varied from experiment to experiment. In order to understand the reasons for this, we analyzed Experiments #7 and #1 where the improvements in specificities were the least 0.35 and 0.27, respectively. In experiment #7, performance was poor due to a single polyp of 6.5 mm in diameter that was located on a haustral fold. We observed that the HT hit belonging to this polyp was 6.7 mm away from the real polyp center as manually identified by the radiologist. Fig. 12 illustrates examples of random planes taken around the HT hit. Obviously, in this case, our method processed the nearby fold but not the polyp, resulting in a low score for the candidate volume. Excluding this polyp from this experiment would result in a specificity level of 0.64 for the cascade system at sensitivity level of 1.00. We observed a similar example in Experiment #1. In this experiment, a polyp of size 3 mm in diameter received the lowest SVM score. The HT hit for this polyp was 3 mm away from the actual location. Excluding this polyp from experiment #1 would result in specificity of 0.85 for sensitivity level of 1.00. We observed that this polyp recurred in Experiments #3 and #8, which also had relatively smaller improvements. Thus, the low scores that some of the polyps received in the cascade system were caused by the mismatch of locations from the preprocessing stage.

The current implementation of the ROSS method has adjustable parameters that may be optimized. Currently, the random images are centered at a random displacement, uniformly distributed between negative five and five voxels,

around the HT hit. As mentioned above, we observed that some of the HT hits were more than five voxels away from the real polyp center as identified by the radiologist, resulting in poor performance. On the other hand, examining a larger volume around the HT hit might impair the system by placing planes on nearby structures. This trade off is under investigation.

Similarly, the output of SVM classifier is strongly dependent on the choice of the Kernel function [29]. While we used exponential radial basis functions, other alternatives might be linear kernels, polynomial kernels, radial basis functions, or histogram-specific kernels. Future work will involve the analysis of different types of kernels and the design of kernels that are most suitable to our pattern processing approach.

The current implementation of the algorithm takes on average 20 s to process each candidate subvolume; We expect that optimized code could decrease the run time. Currently, 100 random triples are used to characterize a 3-D shape. A reduction in the number of random triples would reduce the time required; however the trade off with accuracy needs to be explored. If a less than adequate number of views is used, the triples could be biased toward some specific orientation in space. To find this adequate number, an easy attempt is to test the reproducibility of the ROSS method. That is, the shapes should have the same signatures regardless of the random views used in the analysis. In order to test the reproducibility, we conducted an experiment with synthetic shapes, such as cylinders, ellipsoids and spheres. Two different random sets of k orthogonal views are obtained from each shape. The value of k is varied between five and 100. At the same time, the two signatures (x and y) obtained by the two different random sets are compared by the following error function:

$$e(x, y) = \frac{1}{N} \sum_{i=1}^N \frac{|x_i - y_i|}{x_i + y_i} \quad (4)$$

where N is the number of histogram bins. Ideally, the signatures belonging to the two random sets should be identical. Fig. 13 illustrates the average error versus the time complexity. Here, we would like to remind that processing time for each random triple is approximately 200 ms, and the total processing time is a linear function of the number of planes used (k). The figure demonstrates that the number of random triples could be further reduced with only a negligible loss from the accuracy.

The algorithm is parallelizable, i.e., parallel processors could process portions of the random planes independently, further reducing the required processing time. In the extreme case of using 100 parallel processors to process the 100 planes separately, the processing time reduces to 200 ms or less. Another way to parallelize is to divide the candidate volumes among the processors, i.e., one processor/candidate volume would result in 20 s or less/patient.

V. CONCLUSION

Virtual colonoscopy is a promising new imaging technique to evaluate the human colon for precancerous polyps. Due to the large amount of radiologist time involved in reviewing hundreds of images in a search for small lesions, CAD is necessary to make the approach efficient and cost-effective. Previous

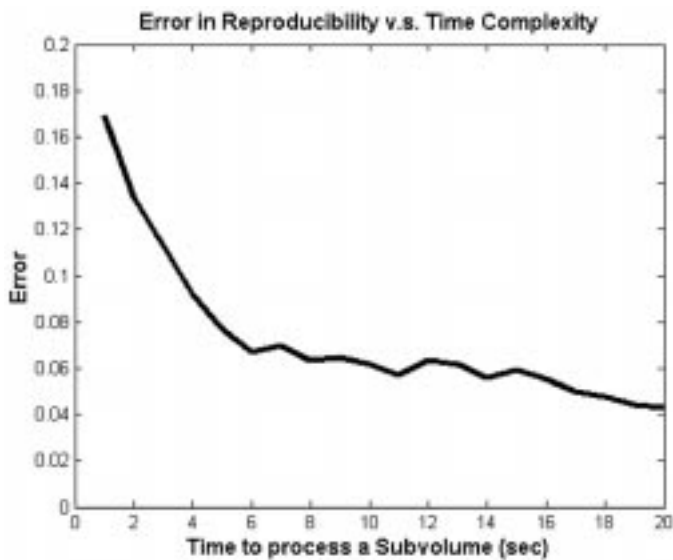


Fig. 13. Illustration of error in reproducibility versus time complexity of the method. Time to process a subvolume is a linear function of the number of random triples. Each triple takes 200 ms to process.

automated detection methods had a high sensitivity for polyp detection, but relied on human observations to differentiate polyps from normal folds or retained fecal material. In this study, we proposed a statistical pattern processing approach that yields a high polyp detection rate with a reasonable number of false positives, thereby showing the feasibility of computer-based screening. One of the main contributions of the paper is the ROSS method, which combines the information from many ROSSs to generate reliable shape signatures. We also show that the use of an SVM classifier is capable of distinguishing implicitly the differentiating characteristics of polyps and healthy tissue, thus improving specificity compared with the preprocessing stage. Studies integrating our CAD methods with radiologist readers will be used to measure potential improvements in sensitivity and efficiency compared with unassisted radiologist interpretation.

APPENDIX A

Let P_i be a collection of points in the image boundary with image coordinates (x_i, y_i) with $i = 1, \dots, N$. In this section we will describe the mathematical details to fit primitive shapes such as circle, quadric curve, and line to the collection of points P_i .

A circle is defined by its three parameters: The coordinates of its center (x_0, y_0) and its radius r . The points P_i satisfy the circle equation

$$[2x_i \ 2y_i \ 1] \begin{bmatrix} x_0 \\ y_0 \\ r^2 - x_0^2 - y_0^2 \end{bmatrix} = [x_i^2 + y_i^2]. \quad (5)$$

In order to find the three unknowns (x_0, y_0, r) , equations coming from each point are stacked and a linear least-square

solution is obtained. Let A and b be the following matrix and vector:

$$A = \begin{bmatrix} 2x_1 & 2y_1 & 1 \\ 2x_2 & 2y_2 & 1 \\ \vdots & \vdots & \vdots \\ 2x_N & 2y_N & 1 \end{bmatrix}, \quad b = \begin{bmatrix} x_1^2 + y_1^2 \\ x_2^2 + y_2^2 \\ \vdots \\ x_N^2 + y_N^2 \end{bmatrix} \quad (6)$$

then $[x_0 \ y_0 \ r^2 - x_0^2 - y_0^2]^T = (A^T A)^{-1} A^T b$ is the least-square solution to the best fitting circle. Similarly, the residue to the least-square solution is $e_{\text{circle}} = A[x_0 \ y_0 \ r^2 - x_0^2 - y_0^2]^T - b$.

A quadratic is defined by the following equation:

$$[x_i^2 \ y_i^2 \ 2x_i \ y_i \ 2x_i \ 2y_i] \begin{bmatrix} \dot{a} \\ \dot{b} \\ \dot{c} \\ \dot{d} \\ \dot{e} \end{bmatrix} = [-1]. \quad (7)$$

Similar to the circle equations, the matrix A and the vector b are built by stacking the left and right sides of the equation, thus, the least-square solution is given by $[\dot{a} \ \dot{b} \ \dot{c} \ \dot{d} \ \dot{e}]^T = (A^T A)^{-1} A^T b$. The least-square residue is given by, $e_{\text{quadratic}} = A[\dot{a} \ \dot{b} \ \dot{c} \ \dot{d} \ \dot{e}]^T - b$.

The quadratic equation can also be written in the following form:

$$[x_i \ y_i \ 1] T \begin{bmatrix} x_i \\ y_i \\ 1 \end{bmatrix} = 0 \quad (8)$$

where T is given by

$$T = \begin{bmatrix} \dot{a} & \dot{c} & \dot{d} \\ \dot{c} & \dot{b} & \dot{e} \\ \dot{d} & \dot{e} & 1 \end{bmatrix}. \quad (9)$$

The singular values of the matrix T are invariant to rotations of the quadratic shape and, thus, are recorded as quadric invariants.

The equation of a line is defined by a point on the line and its direction. Given a collection of points (x_i, y_i) for $i = 1, \dots, N$, the best fitting line (in the least-square sense) goes through the central point (x_0, y_0) . The line itself can then be written as

$$\begin{bmatrix} x_i \\ y_i \end{bmatrix} = \begin{bmatrix} x_0 \\ y_0 \end{bmatrix} + k_i \begin{bmatrix} v_x \\ v_y \end{bmatrix} \quad (10)$$

where $v = [v_x, v_y]^T$ is the direction of the line. Let $(\tilde{x}_i, \tilde{y}_i)$ be the normalized point coordinates of P_i

$$\begin{bmatrix} \tilde{x}_i \\ \tilde{y}_i \end{bmatrix} = \begin{bmatrix} x_i \\ y_i \end{bmatrix} - \begin{bmatrix} x_0 \\ y_0 \end{bmatrix} \quad (11)$$

v is then given as the first singular vector of M , where M is the following matrix:

$$M = \begin{bmatrix} \tilde{x}_1 & \tilde{x}_2 & \cdots & \tilde{x}_N \\ \tilde{y}_1 & \tilde{y}_2 & \cdots & \tilde{y}_N \end{bmatrix}. \quad (12)$$

Given v and (x_0, y_0) the best fitting line equation is complete. For any point P_i , the error to this line is given as

$$e_{\text{line}}^i = \left\| v^T \begin{bmatrix} \tilde{x}_i \\ \tilde{y}_i \end{bmatrix} v - \begin{bmatrix} \tilde{x}_i \\ \tilde{y}_i \end{bmatrix} \right\|. \quad (13)$$

REFERENCES

- [1] P. J. Wingo, "Cancer statistics," *Ca Cancer J. Clin.*, vol. 45, pp. 8–30, 1995.
- [2] R. F. Thoeni and I. Laufer, "Polyps and cancer," in *Textbook of Gastrointestinal Radiology*. Philadelphia, PA: Saunders, 1994, p. 1160.
- [3] S. J. Winawer, A. G. Zauber, M. N. Ho, M. J. O'Brien, L. S. Gottlieb, S. S. Sternberg, and J. D. Wayne, "Prevention of colorectal cancer by colonoscopy polypectomy," *N. Eng. J. Med.*, vol. 329, pp. 1977–1981, 1993.
- [4] D. M. Eddy, "Screening for colorectal cancer," *Ann. Intern. Med.*, vol. 113, pp. 373–384, 1990.
- [5] C. G. Coin, F. C. Wollett, J. T. Coin, M. Rowland, R. K. Deramos, and R. Dandrea, "Computerized radiology of the colon: A potential screening technique," *Comput. Radiol.*, vol. 7, no. 4, pp. 215–221, 1983.
- [6] A. S. Chaoui, M. A. Blake, M. A. Barish, and H. M. Fenlon, "Virtual colonoscopy and colorectal cancer screening," *Abdom. Imag.*, vol. 25, no. 2, pp. 361–367, 2000.
- [7] C. D. Johnson and A. H. Dachman, "CT colonography: The next colon screening examination?," *Radiology*, vol. 216, no. 2, pp. 331–341, 2000.
- [8] D. J. Vining, "Virtual colonoscopy," *Gastrointest. Endosc. Clin. N. Amer.*, vol. 7, no. 2, pp. 285–291, 1997.
- [9] A. K. Hara, C. D. Johnson, J. E. Reed, D. A. Ahlquist, H. Nelson, R. L. Ehman, C. H. McCollough, and D. M. Ilstrup, "Detection of colorectal polyps by computed tomographic colography: Feasibility of a novel technique," *Gastroenterology*, vol. 110, no. 1, pp. 284–290, 1996.
- [10] M. Macari, A. Milano, M. Lavelle, P. Berman, and A. J. Megibow, "Comparison of time-efficient CT colonography with twoand three-dimensional colonic evaluation for detecting colorectal polyps," *Amer. J. Roentgenol.*, vol. 174, no. 6, pp. 1543–1549, 2000.
- [11] A. K. Hara, C. D. Johnson, J. E. Reed, D. A. Ahlquist, H. Nelson, R. L. MacCarty, W. S. Harmsen, and D. M. Ilstrup, "Detection of colorectal polyps with CT colography: Initial assessment of sensitivity and specificity," *Radiology*, vol. 205, no. 1, pp. 59–65, 1997.
- [12] D. K. Rex, D. Vining, and K. K. Kopecky, "An initial experience with screening for colon polyps using spiral CT with and without CT colography," *Gastrointest. Endosc.*, vol. 50, no. 3, pp. 309–313, 1999.
- [13] American Society for Gastrointestinal Endoscopy ASGE, "Technology status evaluation: Virtual colonoscopy: November 1997," *Gastrointest. Endosc.*, vol. 48, no. 6, pp. 708–710, 1998.
- [14] A. H. Dachman, J. K. Kimiyoshi, C. M. Boyle, Y. Samara, K. R. Hoffmann, D. T. Rubin, and I. Hanan, "CT colonography with three-dimensional problem solving for detection of colonic polyps," *Amer. J. Roentgenol.*, vol. 171, no. 4, pp. 989–995, 1998.
- [15] C. F. Beaulieu, Jr., R. B. Jeffrey, C. Karadi, D. S. Paik, and S. Napel, "Display modes for CT colonography—Part II: Blinded comparison of axial CT and virtual endoscopic and panoramic endoscopic volume-rendered studies," *Radiology*, vol. 212, no. 1, pp. 203–212, 1999.
- [16] R. M. Summers, C. F. Beaulieu, L. M. Pusanik, J. D. Malley, R. B. Jeffrey, D. I. Glazer, and S. Napel, "Automated polyp detector for CT colonography: Feasibility study," *Radiology*, vol. 216, no. 1, pp. 284–290, 2000.
- [17] H. Yoshida, Y. Masutani, P. M. MacEaney, K. Doi, Y. Kim, and A. H. Dachman, "Detection of colonic polyps in CT colonography based on geometric features," *Radiology*, vol. 217, no. P, p. 582, 2000.
- [18] S. B. Göktürk and C. Tomasi, "A graph method for the conservative detection of polyps in the colon," in *Proc. 2nd Int. Symp. Virtual Colonoscopy*, Boston, MA, 2000.
- [19] D. S. Paik, C. F. Beaulieu, R. B. Jeffrey, J. Yee, A. M. Steinauer-Gebauer, and S. Napel, "Computer aided detection of polyps in CT colonography: Free response roc evaluation of performance," *Radiology*, vol. 217, no. SS, p. 370, 2000.
- [20] D. S. Paik, C. F. Beaulieu, R. B. Jeffrey, C. Karadi, and S. Napel, "Detection of polyps in CT colonography: A comparison of a computer aided detection algorithm to 3-D visualization methods," in *Radiological Society of North America 85th Scientific Sessions*. Chicago, IL: Radiological Soc. N. Amer., 1999, p. 428.
- [21] S. B. Göktürk, C. Tomasi, B. Acar, D. S. Paik, C. F. Beaulieu, and S. Napel, "A learning method for automated polyp detection," in *Proc. Medical Image Computing and Computer-Assisted Intervention (MICCAI'01)*, Utrecht, The Netherlands, pp. 85–93.
- [22] R. C. Gonzalez and R. E. Woods, *Digital Image Processing*. Reading, MA: Addison-Wesley, 1993.
- [23] M. K. Hu, "Visual pattern recognition by moment invariants," *IRE Trans. Inform. Theory*, vol. IT-8, pp. 179–187, 1962.
- [24] R. M. Gray, *Entropy and Information Theory*. Berlin, Germany: Springer-Verlag, 1990.
- [25] J. A. Hartigan and M. A. Wong, "A k-means clustering algorithm," *Appl. Statist.*, pp. 100–108, 1979.
- [26] V. N. Vapnik, *The Nature of Statistical Learning Theory*. New York: Springer, 1995.
- [27] B. Schölkopf, *Support Vector Learning*. Munich, Germany: R. Oldenbourg Verlag, 1997.
- [28] K. Veropoulos, C. Campbell, and N. Cristianini, "Controlling the sensitivity of support vector machines," presented at the Int. Joint Conf. AI (IJCAI'99), Stockholm, Sweden.
- [29] C. Campbell. Kernel methods: A survey of current techniques. [Online]. Available: <http://citeseer.nj.nec.com/campbell00kernel.html>.