A New 3-D Pattern Recognition Technique With Application to Computer Aided Colonoscopy

Salih Burak Göktürk
Stanford University
Electrical Engineering Department
Stanford, CA 94305
gokturkb@stanford.edu

Carlo Tomasi
Stanford University
Computer Science Department
Stanford, CA 94305
tomasi@stanford.edu

Abstract

To utilize CT or MRI images for computer aided diagnosis applications, robust features that represent 3-D image data need to be constructed and subsequently used by a classification method. In this paper, we present a computer aided diagnosis system for early diagnosis of colon cancer. The system extracts features by a new 3-D pattern processing method and processes them using a support vector machine classifier. Our 3-D pattern processing method, called Random Orthogonal Shape Section(ROSS) mimics the radiologist's way of viewing these images and combines information from many random triples of mutually orthogonal sections going through the volume. Another contribution of this paper is a new feedback framework between the classification algorithm and the definition of the features. This framework, called Distinctive Component Analysis combines support vector samples with linear discriminant analysis to map the features of clustered support vectors to a lower dimensional space where the two classes of objects of interest are optimally separated so as to obtain better features. We show that the combination of these better features with support vector machines classification provides a good recognition rate.

1. Introduction

Colon cancer is the second leading cause of cancer deaths in the USA [16]. Previous research has shown that adenomatous polyps in the human colon have a high probability of developing into subsequent colorectal carcinoma [13]. Detection and removal of small polyps can totally eliminate the disease. Unfortunately, endoscopic diagnosis is both invasive and expensive. Without systematic screening, colon cancer is most often discovered after the patient develops symptoms, and by then, the likelihood of a cure has diminished substantially.

Computerized tomography (CT) is a promising, non-invasive procedure for screening [1, 15], as long as, automatic diagnosis software assists the radiologist, thereby reducing the cost of analysis. Automated polyp detection in 3-D CT images is a challenging problem because of both the large volume of data and the high variety of polyp shapes. First of all, a CT scan typically involves searching through several hundreds of images. Therefore, a two stage system is necessary where the first stage efficiently preprocesses the data and generates polyp candidates, and the second stage performs an elaborate analysis of the remaining candidates for a final decision.

More fundamentally, the diversity of polyp shapes is a major challenge to computer-aided diagnosis, particularly when contrasted with the similarity that polyps can exhibit with normal anatomical structures or residues of stool in the intestine (Figure 1). A method that automatically extracts subtle differentiating characteristics of both polyps and healthy tissue is needed.

Initial studies concerning automated polyp detection have started from the intuitive observation that the shape of a polyp is similar to a hemisphere. In [12], Summers *et al.* compute the minimum, maximum, mean, and Gaussian curvatures at all points on the colon wall and then use more restrictive criteria in order to eliminate non-spherical shapes. In [9], Paik *et al.* use a Hough transform method based on the concept that normals to the colon surface intersect neighboring normals depending on the local curvature of the colon. This method detects the polyps by giving the small volumes a score based on the number of intersecting normal vectors, a criterion which gets its maximum value when the candidate shape is a perfect hemisphere. In [6], Göktürk and Tomasi propose a method based on the observation that the bounding surfaces of polyps are usually not exact spheres, but are often well described by the superposition of small, approximately spherical patches. In this method, a sphere is fit locally to the isodensity surface passing through every CT voxel in the wall region. Group of voxels that originate densely populated nearby sphere centers are considered as polyp candidates.

Due to the large number of false positive detections, all of the methods mentioned above should be considered polyp

0-7695-1272-0/01 $10.00 © 2001 IEEE

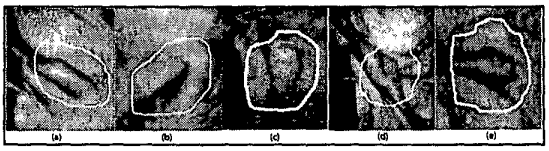

Figure 1. (a-c) Examples of polyps (d-e) examples of healthy tissue that have similar shapes.

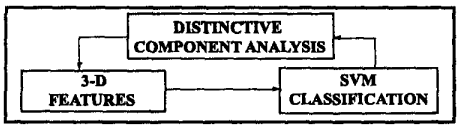

Figure 2. Flow diagram of the system

candidate detectors rather than polyp detectors, *i.e.*, they provide the first stage of a two-stage system. This paper presents a statistical method that captures the more subtle distinctions that differentiate between polyps and normal tissue, *i.e.*, the second stage of a two-stage system. The input to this stage is a set of small candidate volumes, which may or may not contain polypoid shapes. Our novel 3-D pattern processing technique (ROSS) generates shape-signatures for each candidate volume. The signatures are then fed to a Support Vector Machine (SVM) classifier for the final diagnosis of the volume. SVMs identify the elements, so called the support vectors, that carry the distinguishing characteristics between the two classes in the training set. In order to get further intuition, one way of explicitly obtaining these features is through looking at 3-D renderings of the support vectors. Here, we propose an automatic way of obtaining these features. In addition, and in clear contrast with previous statistical classification work, we present a framework where we use feedback from the classification algorithm in order to obtain distinguishing features of the volume by a concept called Distinctive Component Analysis (DCA). Using the feedback framework, we show that improved features can be attained for classification purposes. More importantly, DCA might be used to understand the differentiating characteristics between similar polyps and healthy tissue.

The paper is organized as follows: Section 2 explains our methods, including the ROSS technique, SVM classifier, and DCA in detail. Section 3 describes the experimental setup and discusses our results. Section 4 gives our conclusions and possible directions for future work.

2. Methods

Figure 2 gives the flow diagram of our system. We propose not only a good selection of features and a statistical classification method, but also present a framework where we use feedback from the classification method to obtain a better combination of these features. The system consists mainly of three components: ROSS method for feature creation, a SVM classifier as the statistical classifier, and DCA as the feedback mechanism between them.

The flow diagram of the ROSS method is given in Figure 3. Here, we mimic the way the radiologists view colon CT images. Instead of using 3-D renderings of the tissue, they prefer to view the images through three perpendicular planes aligned with the transaxial, sagittal, and coronal anatomical directions[3] (Figure 4). Viewing through these three main axes gives substantial information about the 3-D shape, but is incomplete by itself. To be more accurate, we get a sufficiently large number of random triples of mutually orthogonal planes through the volume. Having obtained many triples of planes, geometric attributes are calculated from each random plane. We use a histogram of these geometric attributes obtained from each triple as a feature-vector to represent the shape. Taking histograms of these geometric attributes over several random triples makes the resulting signatures invariant to rotations and translations. In addition, the right choice of the geometric attributes gives robust signatures of the volume. In this work we have chosen to sample through the 3-D shape in the form of 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 ellipsoids or other primitive objects to the shape. Alternatively, one could use the histogram of local structures such as the curvature. However, this choice is not diagnostic enough since different shapes might often have similar distribution of local information. As discussed in Section 2.1, 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.

SVMs are then trained with volume signatures computed from an initial set, and are subsequently used to classify new volumes into polyps and normal tissues. As discussed in Section 2.2, support vectors are essentially the vectors that carry the most distinguishing components between polyps and healthy tissue. We explicitly obtain these components by DCA, as described in Section 2.3, and subsequently improve our features using these distinctive components.

2.1. Random Orthogonal Shape Sections Method

Each candidate volume is sliced with several triples of perpendicular planes. Since a polyp may not occupy the resulting image entirely, the images are first segmented to disregard tissues surrounding the putative polyp.

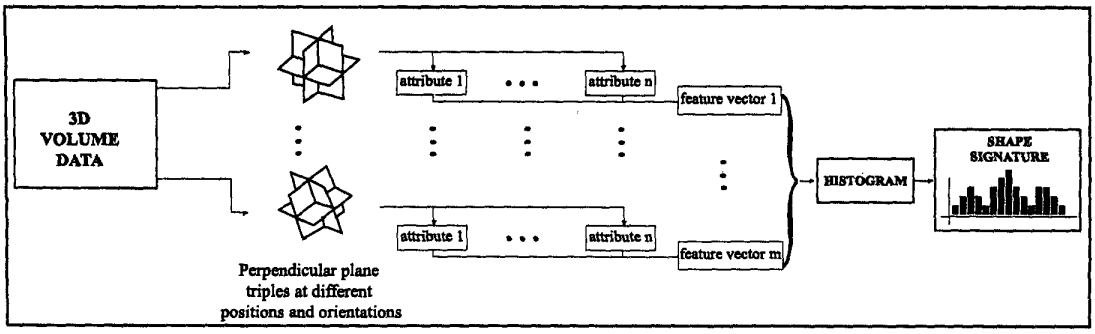

Figure 3. Overflow of the ROSS method.

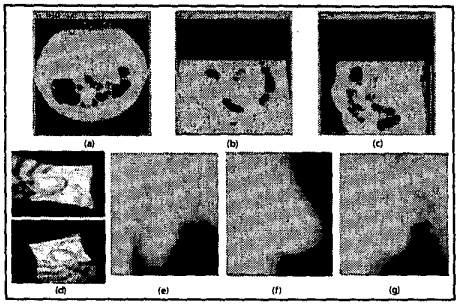

Figure 4. (a-c) Three views of the body through the anatomical directions. (d) 3-D rendering of a polyp (e-g) A randomly oriented, mutually orthogonal triple of planes

Figure 5. Optimum segmentation window in various images .

Segmentation

The size of the critical region varies depending on the size and the shape of the suspicious lesion. Here, we aim to discover the best square window that would capture the essentials of the shape. Figure 5 depicts the critical window for different shapes. A window is considered good when it contains a shape that is approximately circular and has a small radius. To find an optimal window size w, an image is first binarized, and the following function is computed for each window of size w and centered in the middle of the image:

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

Here, $r(w)$ is the radius of the circle that best fits the edges in the binarized image, $e_{circle}(w)$ is the residue to the best fitting circle, $e_{line}(w)$ is the residue to the best fitting line, and a_1 and a_2 are constants. The value of w that yields the smallest $f(w)$ is chosen as the best window size.

Image Based Geometric Features

Once we find the optimal subwindow in a particular image, we find some geometric features related to shape and intensity in the segmented image. First of all, these features should be able to capture representative information about the candidate shape. For this purpose, we fit primitive shapes such as circle, quadratic curve, and line (Figure 6 (d,e,f)) to the largest connected edge component, i.e. boundary of the shape.

A random slice of a sphere is a circle. Thus, fitting circles is a means of measuring the sphericity of the 3-D shape.When doing so, the residuals at each pixel on the boundary are first weighted with a Gaussian located at the image center, and shown in Figure 6(c). By weighting with this Gaussian mask, we give more importance to boundary points of the shape than to those of the surrounding colon wall. The weighted residuals are then added together, and the least square fit gives the optimum circle. The residual,e_{circle}, to the least square solution is recorded as well.

Similarly, the residual, e_{line} to the optimum fitting line gives information on the flatness of the surface. Quadratic curves include any second order equation system of two variables. By fitting a quadratic curve to the boundary of the image, the ellipsoidal structure of the shape can be measured.

The projection of a fold onto the image contains parallel lines.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 angle between them.

In order to extract information on higher order frequency characteristics of the boundary, 3rd order moment invariants are computed as well [7]. 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 segmented by an intensity thresholding, and next, the intensity

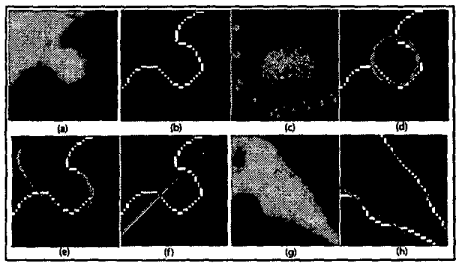

Figure 6. Illustration of the primitive shapes: (d) circle, (e) quadric (f) line (h) parallel lines. Each shape is weighted by the Gaussian given in (c)

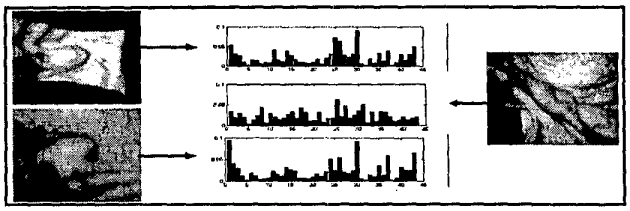

Figure 7. Two polyps (left) and a very similarly shaped fold structure (right). As expected, the signatures are very similar, yet distinguishable.

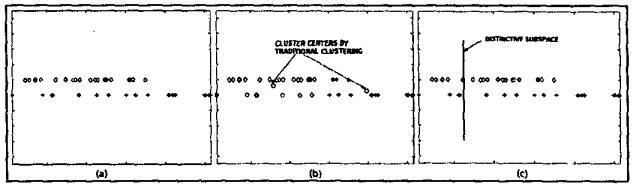

Figure 8. Illustration of a specific case where a clustering method with nearest neighbor distortion measure would fail.

mean and standard deviation of the tissue is 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. Often, one plane out of these three planes does not contain useful geometric information, i.e. it might be totally tissue, or totally air region. Thus, we eliminate the features from that plane, and the vector containing attributes from the remaining two planes represents the signature of the shape, relative to that particular triple of perpendicular planes (triple vector).

Obtaining the Histograms in High Dimensional Space

The features computed from each triple of perpendicular planes depend on the position and orientation of that particular triple. However, if histograms of feature distributions are computed from sufficiently many triples with random positions and orientations, the histograms themselves are essentially invariant to position and orientation while characterizing many variations of shapes with a few numbers.

Here, we would like to mention that getting histograms of data in high dimensional space is difficult. First of all, uniform histograms are out of the question, since the high dimension of the underlying feature space would imply prohibitive amounts of storage. A more efficient solution, 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 [5].

In order to obtain the histogram bin centers, the k-means clustering algorithm is applied on a training set of triple vectors. 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 triple 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 triple 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 per candidate volume, are the rotation and translation invariant shape signatures. In figure 7, examples of signatures obtained for various shapes are illustrated.

As one might guess, performance of the system is closely related to how well the representative histogram bin centers are chosen. A clustering algorithm with nearest neighbor distortion measure (such as k-means clustering algorithm) would fail in some situations. Figure 8 gives such an example. In (a), let the square and diamond points belong to two different classes. Although it seems to be an easy case to distinguish, when the k-means algorithm is executed, the two cluster centers are obtained as in Figure 8(b), by circles. In brief, the two clusters are not correctly identified because the spread of each cluster greatly exceeds the intercluster distance. We will address this problem by extracting the distinctive components in Section 2.3 through linear discriminant analysis.

2.2 Support Vector Machine Classification

Given a training set of representative vectors for polyps and healthy tissue, an optimum classifier is obtained, and subsequently used for the classification of new test data. A classifier learning algorithm takes this training set as input

and produces a classifier as its output. The learned classifier is then a function that, for any new candidate volume, tells whether it is a polyp or not.

The goal is then to find a separation function between the polyps and the healthy tissue that can be induced from the known data points and generalizes well on the unknown examples. Proposed first by Vapnik [14], the SVM classifier aims to find the optimal differentiating hypersurface between the two classes. 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 $i = 1, 2, \ldots, N$. Each point x_i belongs to either class and is accordingly given a label $y_i \in \{-1, 1\}$. SVMs [14, 11] implicitly transform the given feature vectors x 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. Finding the optimal hyperplane is an optimization problem where the distance of the margin points to the hyperplane is maximized. This optimization problem has been well studied in the literature [14, 11]. During the optimization problem, only inner products of the form $K(x_i, x_j) = \phi(x_i)\phi(x_j)$, called Kernels, need ever to be computed, rather than the high dimensional vectors $\phi(x)$ themselves. In the classification process, only the vectors that are very close to the separating hypersurface need to be considered when computing kernels. These vectors are called the support vectors. In other words, support vectors are essentially the closest points to the optimal separating hyperplane. An explanatory description of the support vectors then would be, 'the data points that carry the differentiating characteristics between the two classes'. Once we obtain the support vectors, the distance of a vector x from the optimal classifier has the following form:

$$d(x) = \frac{b + \sum_{x_i \in SV's} \alpha_i y_i K(x_i, x)}{|w|} \qquad (2)$$

where the constants α_i, b are computed by the classifier-learning algorithm. See [11] for details. The sign of this distance gives the decision of the SVM classifier. Computing the coefficients α_i, b is a relatively expensive (but well understood) procedure, but needs to be performed only once, on the training set.

Different from other statistical classification methods, SVMs minimize the structural risk, given as the probability of misclassifying previously unseen data. Thus SVM is theoretically well generalizable to test 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. These features can explicitly be obtained by further observations on the support vector shapes as shown in next section.

More generally, using a learning method is necessary for a complex problem such as polyp recognition where it is not quite possible to find the distinguishing components of the two classes. Using SVM rather than hand-crafting classification heuristics, exploits all of the information in the training set optimally, and eliminates the guess work from the task of defining appropriate discrimination criteria.

2.3 Feedback Framework - Distinctive Component Analysis

SVM identify the support vectors, the data that carries the distinctive characteristics in both of the classes, and implicitly uses this information in the classification process. Because of the great number of dimensions of the underlying feature space, the data vectors themselves are the result of the clustering methods described in section 2.1. Clustering, in turn, suffers from the problems shown at the end of that section.

To address these problems, we have developed a feedback framework that uses support vectors in order to obtain the distinctive components more explicitly, yielding better clusters in turn. In addition, our feedback framework provides an explanation to what aspect of a feature is particularly distinctive, and can therefore provide insights to feature designers.

Data vectors that become support vectors come in shapes that are simultaneously similar to both polyps and healthy tissue. It therefore makes sense that analyzing the differences between nearby support vectors on opposite sides of the classification boundary provides information on how features can be improved. In this section, we describe a new approach, called *Distinctive Component Analysis* (DCA), that uses linear discriminant analysis (LDA) [3] to employ the first-iteration feature vectors to produce new, low-dimensional feature vectors that are best at discriminating between positive and negative examples in small neighborhoods of the classification surface.

As an example, consider the clustering failure given in Figure 8. As it turns out, linear discriminant analysis will define a new feature as the projection of the data to the vertical line in Figure 8(c).

Without loss of generality, let us now consider the two-class clustering problem. We assume that several labeled data vectors for each class are available, and that these were constructed with the clustering methods described in section 2.1. The objective of DCA is then to find a projection of the data vectors onto a lower dimensional space where the two classes become well separated.

Mathematically, this amounts to finding the best low-dimensional space, where the intra-class variation of the projected vectors is minimized, while the inter-class variation is maximized. Linear discriminant analysis is one of the methods to obtain this low dimensional domain[3]. Without loss

of generality, let us develop the model for projecting the data from its original space (of dimension n) onto a single dimension (line). Let N_1 of x_1^i's be the instances of the first class and N_2 of x_2^i's be the instances of the second class where N_1 and N_2 are the number of instances in each class. Let $\tilde{x}_1^i$ and $\tilde{x}_2^i$ be the projection of the vectors on to the distinguishing line. That is, given an $n \times 1$ projection vector α, $\tilde{x}_1^i = \alpha^T x_1^i$ and $\tilde{x}_2^i = \alpha^T x_2^i$. The distinctive component is then determined by the value of the projection vector α, where the following cost function is minimized:

$$C(\alpha) = \xi_1 \sum_{\forall \tilde{x}_1^i} \|\tilde{x}_1^i - \tilde{x}_1\|^2 + \xi_2 \sum_{\forall \tilde{x}_2^i} \|\tilde{x}_2^i - \tilde{x}_2\|^2 - \|\tilde{x}_2 - \tilde{x}_1\|^2 \tag{3}$$

where $\tilde{x}_1 = \frac{1}{N_1} \sum_{\forall \tilde{x}_1^i} \tilde{x}_1^i$, $\tilde{x}_2 = \frac{1}{N_2} \sum_{\forall \tilde{x}_2^i} \tilde{x}_2^i$, and ξ_1 and ξ_2 are constants. We observed that $0.1 \leq \xi_1 = \xi_2 = \xi \leq 0.5$ gave similar results. The optimization problem can also be written in the following form,

$$C(\alpha) = \alpha^T \mathbf{H} \alpha \tag{4}$$

where $\mathbf{H}$ is given by:

$$\mathbf{H} = \xi_1 \frac{X_1 X_1^T}{N_1} + \xi_2 \frac{X_2 X_2^T}{N_2} + \frac{X_1 \mathbf{1}_1 \mathbf{1}_2{}^T X_2^T + X_2 \mathbf{1}_2 \mathbf{1}_1{}^T X_1^T}{N_1 N_2} - (1+\xi_1) \frac{X_1 \mathbf{1}_1 \mathbf{1}_1{}^T X_1^T}{N_1^2} - (1+\xi_2) \frac{X_2 \mathbf{1}_2 \mathbf{1}_2{}^T X_2^T}{N_2^2}$$

where X_1 is the matrix with each column containing an x_1^i and similarly for X_2. $\mathbf{1}_1$ and $\mathbf{1}_2$ are column vectors of size N_1 and N_2 with all 1's as elements. Observe that $\mathbf{H}$ is symmetric and thus has the same left and right eigenvectors, i.e. $\mathbf{H} = U^T \Lambda U$, where Λ is the diagonal matrix of sorted eigenvalues and U is the matrix of eigenvectors. Assuming that $\|\alpha\| = 1$, the expression (4) achieves its minimum value when $\alpha = U_n$, the eigenvector of $\mathbf{H}$ associated with its smallest eigenvalue. The solution of LDA proposed here is slightly different and more flexible than the traditional LDA solution. In this proposed solution, by varying ξ_1 and ξ_2, one could evaluate the trade off between the summation in equation 3 (the difference of intra-class and inter-class variations) versus the ratio of inter-class variance to intra-class variance.

One could conceivably generalize this solution to mappings to higher dimensions than individual lines, by using the last m eigenvectors, rather than the last one. However, the eigenvectors other than U_n, are perpendicular to U_n. Thus, the addition of these vectors is not of much practical importance, since U_n already captures the most distinctive information.

An analysis that identifies the distinctive components of the objects helps by not only distinguishing these important features, but also eliminating the distracting features that a clustering algorithm would misinterpret. Even more importantly, it helps us interpret those important features that distinguish between the two classes and design new systems using this information.

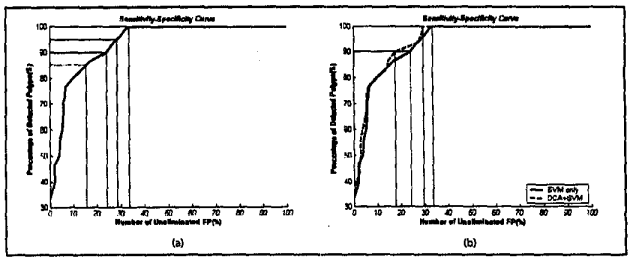

Figure 9. (a) Results obtained by SVM Classifier (b) Improvements obtained using further analysis by DCA .

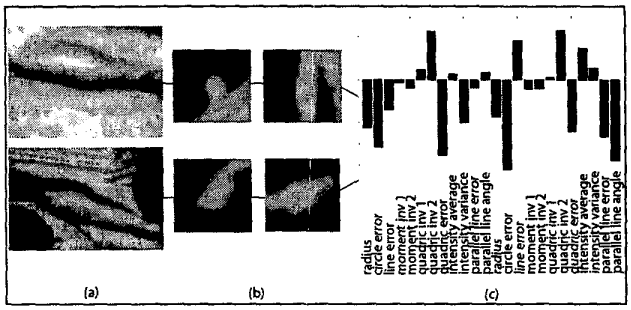

Figure 10. (a) Rendering of a polyp(top) and rendering of a similar healthy tissue(bottom) (b) Examples of two mutually orthogonal planes(the ones that were used out of the triple to obtain the planar attributes). (c) Distinctive component. First 12 attributes belong to the first (more curvy) image, and the next 12 attributes belong to the second image. The parallel line analysis in the second image and also the circle error and quadratic invariants are the distinctive components.

In the original form proposed here, the DCA uses linear discriminant analysis that extracts local components (separable components), rather than components that refer to the entire space of polyps and healthy tissue. Thus, we propose to apply DCA to small clusters of support vectors, as discussed more thoroughly in the experiments.

3. Experiments and Discussion

We used a data set consisting of small candidate volumes from the CT scans of subjects enrolled in our CT colonography study comprising 30 known polyps and 212 other regions containing tissue from healthy surface. These healthy tissues were all false positives obtained in previous work [6, 9], and essentially look quite like the true posi-

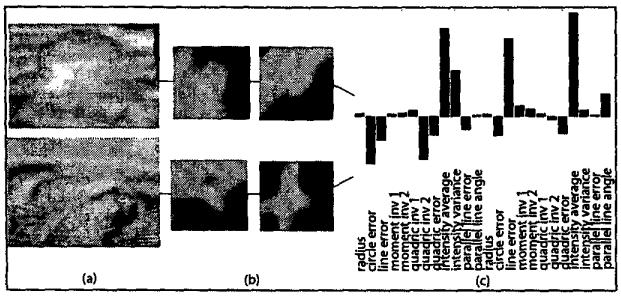

Figure 11. (a) Rendering of a polyp(top) and rendering of a similar healthy tissue(bottom) (b) Examples of two mutually orthogonal planes(the ones that were used out of the triple to obtain the planar attributes). (c) Distinctive component. The intensity average and also error to the best fitting line in the second image are the distinctive components.

	Correct	Wrong
Polyps	24	6
Normal Tissue	186	26
Total	210	32

Table 1. Correct and wrong detections in a preliminary cross-validation experiment. See text for a more informative analysis.

tive polyps (Some examples were shown in Figure 1). All the polyps had a principal radius greater than 5 mm. 150 random triples of perpendicular images were extracted from each candidate shape. A 24-vector was obtained for each triple 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 between parallel lines (if there are parallel lines in the image), total residue to line fit in the pair of parallel lines. Having obtained 24-vector for each triple of mutually orthogonal images, 43 clusters were used in k-means clustering, resulting into a 43-vector signature per candidate shape.

The resulting signature vectors were used as feature vectors by the SVM classification algorithm with exponential radial basis functions as kernels [2]. We applied a 10-fold cross validation study. In this study, we divided the polyps and the healthy tissue into 10 disjoint sets and conducted 10 experiments. In each experiment, one of the 10 sets was used as the test set, and the rest as the training set. Table 1 shows the classification results. The total performance of the system was 87%.

A more meaningful analysis of our classification results can be conducted by considering the requirements of colonoscopy screening. A moderate number of false positives is acceptable for screening: a radiologist examines each polyp reported by the automatic procedure, and false positives merely increase the time the radiologist spends on each case. A large number of false positives, however, would nullify the advantages of an automatic procedure, since it may then take as much for the radiologist to examine the reported instances as it would take her to examine a CT scan from scratch. False negatives, on the other hand, must be avoided, since a missed polyp can not be caught without the radiologist looking at the entire data set. Consequently, an automatic system must have no more false negatives than a human radiologist. [10], [4] and [8] report that radiologists examining CT scans miss about 18-35 percent of polyps of size 5mm or above. This is, therefore, our reference figure for false negatives. Algorithms that perform at this level can then be compared to each other in terms of the rate of false positives.

To examine this trade-off more quantitatively, we have analyzed what happens when the zero-crossing ("sign" function) in expression 2 is replaced by a level crossing. As the level is decreased, more true polyps are detected, but at a cost of more false positives. The percentage of true polyp detections versus false positive detections is given in Figure 9(a). The main objective of this work is to be able to achieve an accuracy of 80% or more with an acceptable number of false positives(FPs). In previous work, comparable accuracy was obtained with about 100 FPs per colon for polyps of size 5mm or greater, and our approach is shown to be able to reduce the false positive rate by 85% and 75% for accuracy levels of 85% and 95%, which inherently reduces the radiologist's interpretation time by the same amount.

The performance of the system can be further increased if the aforementioned problems with the vector quantization clustering are resolved. For this, we propose to use DCA when the system is not certain about its decision. Having obtained the support vectors from the previous analysis, the margin between the support vectors and the optimum hyperplane gives the region where the decision of the system is in the suspicious region for a test case. Thus for that particular case, we use the local distinguishing characteristics of support vectors near that test case. More explicitly, we obtain the closest k (k=2 in our experiment) polyp and non-polyp support vectors to the test case, and apply the linear discriminant analysis to the geometric attribute vectors of many triples of mutually orthogonal planes belonging to those support vector shapes. This way, a distinctive α vector is found between the support vector cases close to the test shape case. The distinctive components (α's) have been very informative in interpreting the distinguishing planar geometric attributes between the support vectors belonging to polyps and support vectors belonging to healthy tissue. For example, the distinctive components of the tissues given in Figure 10 are the angle between the parallel lines, and total fitting error to the parallel lines in the second plane (plane where parallel lines are present) and error to the best fitting circle and quadratic in the first (more curved) and sec-

ond planes. A similar example is also given in Figure 11 where the distinctive component takes in the form of the intensity attributes in the images. We think that DCA gives valuable feedback even to the radiologists regarding these distinguishing planar attributes.

The above mentioned 10 fold study is repeated for evaluating the DCA. For each experiment, if the test case falls between the support vectors(suspicious region where the expression in 2 gets values in $(-|w|, |w|)$), further local analysis of the nearby support vector cases is conducted using DCA. This analysis results in an optimum mapping α between the geometric triple attribute vectors of the nearby support vectors. Finally, this mapping is applied to the geometric triple attribute vectors of the test case and the location of the average mapping with respect to the mappings of the polyp and non-polyp nearby support vectors is used as a measure of polypness. The measure thus obtained is linearly combined with the SVM's measure obtained by expression 2 and final decision regarding the test case is given.

The results of the same 10-fold cross validation study with DCA are summarized in Figure 9(b). The addition of DCA improves the previous results by a considerable amount. The ratio of False Positive detections is further reduced from 24% to 18% and 33% to 29% for sensitivity levels of 90% and 100% respectively.

4. Summary and Conclusions

In order to process the 3-D medical images, we need to develop 3-D pattern processing methods. In this paper, we proposed a new 3-D pattern processing method called ROSS method that mimics the radiologists' way of viewing the images. We attain planar attributes from many randomly oriented triples of mutually orthogonal planes and use vector quantization to obtain histograms of these planar attributes as invariant signatures of the shape. ROSS method is potentially applicable to any shape recognition problem, since it does not make any explicit assumptions about the shape. We combine our signatures with SVM classification and improve the results obtained by the previous study on human colon cancer detection.

Another major contribution of this paper is the idea of feedback framework between the SVM classification algorithm and the features. This framework, called Distinctive Component Analysis, uses linear discriminant analysis to map the features to a lower-dimensional space where the two class of objects are optimally separated so as to obtain better features. This analysis has been informative in understanding the nature of distinguishing characteristics between polyps and normal tissue that have similar shapes. Finally, we show that the combination of these better attributes with SVM classification provides a good recognition rate with a reasonable ratio of false positive detections.

There are many possible directions for future investigation. The ROSS method can be applied to the output of any high-sensitivity algorithm to reduce the ratio of false positive detections. Thus, we would like to apply the new 3-D pattern recognition approach to other medical areas such as lung nodule detection. In addition, we would like to conduct more experiments with DCA. For this, any application area where two classes of similar objects need to be distinguished is appropriate.

Acknowledgement

We would like to thank Burak Acar, David Paik, Sandy Napel and Chris Beaulieu for inspiring discussions. We would also like to thank Chris Beaulieu and Judy Lee for the generous support of CT data. This work was supported by a seed grant from the Program for Bioengineering, Biomedicine and Biosciences at Stanford University.

References

[1] 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*, 212(1):203–212, 1999.

[2] Colin Campbell. *Kernel Methods: A Survey of Current Techniques.* citeseer.nj.nec.com/campbell00kernel.html.

[3] R.O. Duda and P.E. Hart. *Pattern classification and scene analysis.* John Wiley & Sons, New York, 1973.

[4] H.M. Fenlon, D.P. Nunes, P.C. Schroy, M.A. Barish, P.D. Clarke, and J.T. Ferrucci. A comparison of virtual and conventional colonoscopy for the detection of colorectal polyps. *N Engl J Med*, 341(20):1496–1503, 1999.

[5] A. Gersho and R.M. Gray. *Vector Quantization and Signal Compression.* Kluwer Academic Press, 1992.

[6] S.B. Göktürk and C. Tomasi. A graph method for the conservative detection of polyps in the colon. In *2nd International Symposium on Virtual Colonoscopy*, Boston, USA, 2000.

[7] M.K. Hu. Visual pattern recognition by moment invariants. *IRE transactions on information theory*, IT-8:179–187, 1962.

[8] W. Luboldt and J.F. Debatin. Virtual endoscopic colonography based on 3d mri. *Abdom Imaging*, 23(6):568–72, 1998.

[9] 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*, 217(SS):370–370, 2000.

[10] D.K. Rex. Ct and mr colography (virtual colonoscopy: status report. *J Clin Gastroenterol*, 27(3):199–203, 1998.

[11] B. Schölkopf. *Support vector learning.* R. Oldenbourg Verlag, Munich, 1997.

[12] 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*, 216(1):284–290, 2000.

[13] R.F. Thoeni and I. Laufer. Polyps and cancer. *Textbook of Gastrointestinal Radiology, Philadelphia: W.B. Saunders*, page 1160, 1994.

[14] V.N. Vapnik. *The nature of statistical learning theory.* Springer, New York, 1995.

[15] D.J. Vining. Virtual colonoscopy. *Gastrointest. Endosc. Clin. N. Am.*, 7(2):285–291, 1997.

[16] P.J. Wingo. Cancer statistics. *Ca Cancer Journal Clin*, 45:8–30, 1995.