

A Learning Method for Automated Polyp Detection

S.B. Göktürk¹, C. Tomasi¹, B. Acar², D. Paik², C. Beaulieu², S. Napel²

¹Robotics Lab., Computer Science Department, Stanford University

²Three Dimensional Imaging Lab., Department of Radiology, Stanford University
{gokturkb,tomasi,bacar}@Stanford.edu
{paik,cfb,snapel}@s-word.Stanford.edu

Abstract. Adenomatous polyps in the colon have a high probability of developing into subsequent colorectal carcinoma, the second leading cause of cancer deaths in United States. In this paper, we propose a new method for computer-aided diagnosis of polyps. Initial work with shape detection has shown high sensitivity for polyp detection, but at a cost of too many false positive detections. We present a statistical approach that uses support vector machines to distinguish the differentiating characteristics of polyps and healthy tissue, and subsequently uses this information for the classification of the new cases. One of the main contributions of the paper is a new 3-D pattern analysis approach, which combines the information from many random images to generate reliable signatures of the shapes. At 80% polyp detection rate, the proposed system reduces the false positive rate by 80% compared to previous work.

1 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 [1]. Previous research has shown that adenomatous polyps, particularly those larger than 1 cm in diameter, have a high probability of developing into subsequent colorectal carcinoma [2]. Detection and removal of small polyps can totally eliminate the disease. Unfortunately, colon cancer is most often discovered after the patient develops symptoms, and by then, the likelihood of a cure has diminished substantially. Therefore, a cost-effective and patient-comfortable screening procedure is desirable in order to diagnose the disease in an earlier stage.

Optical colonoscopy is considered the definitive diagnostic test as it affords direct visualization and the opportunity for biopsy or removal of suspicious lesion [3]. This method involves an optical probe being inserted into the colon in order to examine the interior. An alternative, non-invasive method has recently been proposed: Computed tomography colonography (CTC) or virtual colonoscopy is a technique that combines axial spiral CT data acquisition of the air-filled and cleansed colon with 3-dimensional imaging software to create endoscopic images of the colonic surface [4]. The initial clinical results are quite promising, yet the technique is still impractical due, in part, to the extensive amount of radiologist time involved in the process, which typically requires viewing over 300 512 by 512 images per patient study. In

order to help the radiologist be more time efficient and accurate, an automated screening method for computer-aided diagnosis of polyps is necessary.

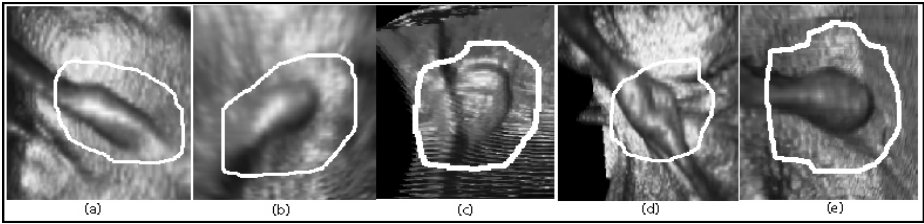


Fig. 1. (a),(b),(c) are polyps, (d) is a normal thickened fold (e) is retained stool.

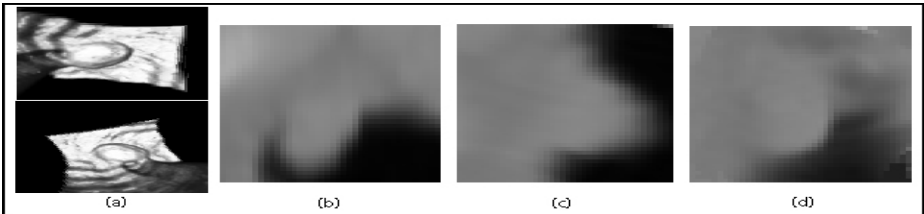


Fig. 2. (a) Two different 3-D views of a polyp. (b-d) Perpendicular images through the same polyp.

Automated polyp detection is a very recently growing area of research. The problem of identifying colonic polyps is very challenging because they come in various sizes and shapes, and because thickened folds and retained stool may mimic their shape and density. Figure 1 demonstrates the appearance of polyps and other tissue as they appear in a virtual colonoscopy study.

Initial studies concerning automated polyp detection have started from the intuitive observation that the shape of a polyp is similar to a hemisphere. In [5], Summers *et al.* computes the minimum, maximum, mean and Gaussian curvatures at all points on the colon wall. Following discrimination of polypoid shapes by their principal minimum and maximum curvatures, more restrictive criteria are applied in order to eliminate non-spherical shapes.

In [6], 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. The method detects the polyps by giving the shapes a score based on the number of intersecting normal vectors. In [7], Yoshida *et al.* use curvedness to distinguish polyps from healthy tissue.

In [8], Göktürk 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 by 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. 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 can be considered more as polyp *candidate* detectors than polyp detectors. This paper presents a statistical method to differentiate between polyps and normal tissue.

The input to the system is a set of small candidate volumes, which may or may not contain polypoid shapes. This set can be computed by one of the methods just discussed. Our novel volume processing technique 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.

The paper is organized as follows: Section 2 explains both the volume processing and support vector classifier in detail. Section 3 describes the experimental setup and discusses our preliminary results. Section 4 gives some conclusions and possible directions for future work.

2 Our Method

Many radiologists prefer to view colon CT images through three perpendicular image planes aligned with the transaxial, sagittal, and coronal anatomical directions[4]. Figure 2 shows a polyp in 3-D and 3 perpendicular views through it in 2-D. These perpendicular planes capture substantial information about the shape, yet they are incomplete by themselves. A more accurate signature of the volume can be obtained by collecting statistics over several *triples* of mutually orthogonal planes, oriented along other than the anatomical directions. We use a histogram of several geometric attributes obtained from many triples of perpendicular random images as a feature-vector to represent the shape.

Computing shape features on triples of planes, rather than on individual planes, captures 3D shape aspects more completely. Taking histograms of these features over several random triples makes the resulting signatures invariant to rotations and translations. More details on our signatures are given sections 2.1 and 2.2. Support-vector machines, described in section 2.3, are then trained with signatures computed from an initial set, and are subsequently used to classify new volumes into polyps and normal tissues.

2.1 Image Processing

Each candidate volume is sliced with several triples of perpendicular planes. A polyp may not occupy the resulting images entirely. As a consequence, images are segmented, so as to disregard tissues surrounding the putative polyp. This segmentation process is described in section 2.1.1. Shape and intensity features are then computed in the resulting sub-windows, as discussed in section 2.1.2.

2.1.1 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 3 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. Because elongated folds are common in the colon, it was found to be useful to also explicitly rule out elongated regions. To find an

optimal window size w , an image is first binarized, and the following target function is computed for each window of size w and centered in the middle of the image:

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

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

2.1.2 Image Features

Image features should capture representative information about the candidate shape. Primitive shapes such as circle, quadric curve, and line (figure 4. d,e,f) are fit 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 image 4(c). The purpose of these weights is to 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 solution gives the optimum circle. The residual to the least square solution is recorded as well.

Similarly, the residual 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, thereby helping to capture similarity to a pedunculated polyp.

In order to extract information on higher order frequency characteristics of the boundary, 3rd order moment invariants are computed as well [9]. Specifically, given the coordinates of points on the plane, a (p,q) th moment and the third order moment invariant M are given by,

$$m_{pq} = \sum_{i=1}^N (x_i - x_0)^p \cdot (y_i - y_0)^q \text{ and } M = (m_{30} + m_{12})^2 + (m_{03} + m_{21})^2$$

In addition to these shape features, intensity features are extracted from the tissue part of the image. The colon tissue is first separated away from the interior of the colon (air) by intensity thresholding. The mean and the standard deviation of the remaining intensity values are recorded as intensity measurements in the feature vector.

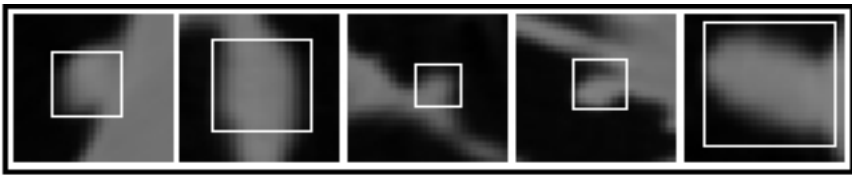


Fig. 3. 2-D images of five different suspicious lesions, with the size of the critical region varying from image to image.

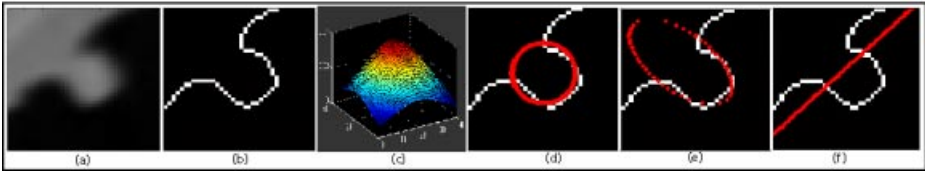


Fig. 4. (a) Sample image. (b) Edges. (c) Gaussian mask to weight the edge points on the image. (d) The circle fitted to the weighted edge points. (e) The quadratic fitted to the weighted edge points. (f) The line fitted to the edge points.

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. This vector represents the signature of the shape, relative to that particular triple of perpendicular planes.

2.2 Vector Quantization

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. Explicit histograms, on the other hand, 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 bins. This method is called *vector quantization* and is described in more detail in [10].

Let X_{ij} be the n -vector obtained from the j^{th} random triple of perpendicular planes extracted from the i^{th} shape. Having obtained X_{ij} 's from all of the shapes, the k -means algorithm [10] is used to compute vector clusters. Let X be the matrix each column of which contains one particular X_{ij} . The cluster centers are initialized to be the principal component directions of X . Subsequent iterations of the k -means algorithm then alternately reassign vectors to clusters and recompute cluster centers, resulting eventually into the optimum cluster centers B_i .

When forming 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 cluster centers. The histograms thus obtained, one per candidate volume, are the shape signatures used for classification as described in the next section.

2.3 Support Vector Machines

A classifier learning algorithm takes a *training set* as input and produces a *classifier* as its output. In our problem, a training set is a collection of candidate volumes that a radiologist has individually labeled as polyps or non-polyps. A classifier is a function that, for any new candidate volume, tells whether it is a polyp or not.

Mathematically, a classifier can also be viewed as a *hypersurface* in feature space, that separates polyps from non-polyps. Support vector machines (SVM) [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. This mapping from x to $\phi(x)$ is used implicitly in that only inner products of the form $K(x_i, x_j) = \phi(x_i)^T \phi(x_j)$ need ever to be computed, rather than the high dimensional vectors $\phi(x)$ themselves. In these so-called *kernels*, the subscripts i, j refer to vectors in the training set. 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*. Suppose that vector x_i in the training set is given (by the radiologist) a label $y_i = 1$ if it is a polyp and $y_i = -1$ if it is not. Then the optimal classifier has the form:

$$f(x) = \text{sign} \left(\sum_{SV's} \alpha_i y_i K(x_i, x) + b \right) \quad (A)$$

Where SV denotes the set of support vectors, and the constants α_i and b are computed by the classifier-learning algorithm. See [11] for details. 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. During volume classification, only the very simple expression (A) needs to be computed.

SVMs minimize the structural risk, given as the probability of misclassifying previously unseen 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 makes support vectors very appropriate for polyp recognition. More generally, using a learning method, 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.

3 Experiments

We used a data set consisting of small candidate volumes from the CT scans of subjects enrolled in our CT colonography study comprising of 47 known colonic polyps and 250 other regions containing tissue from healthy mucosal surface. These healthy tissues were all false positives obtained in previous work [6,8], and essentially look quite like the true positive polyps. All the polyps were bigger than 5 mm in their principal radius. 150 random triples of perpendicular images were extracted from each candidate shape. A 30-vector was obtained for each triple, and 32 clusters were used in k-means clustering, resulting into a 32-vector signature per candidate shape.

Figure 5 shows examples of two polyps and one thickened, yet normal fold along with their corresponding signatures.

Linear, polynomial and exponential radial basis functions [11] were used as kernel functions in different experiments with support vector machine classifiers. To test for each candidate shape, the classifier was first trained with all the other shapes and then

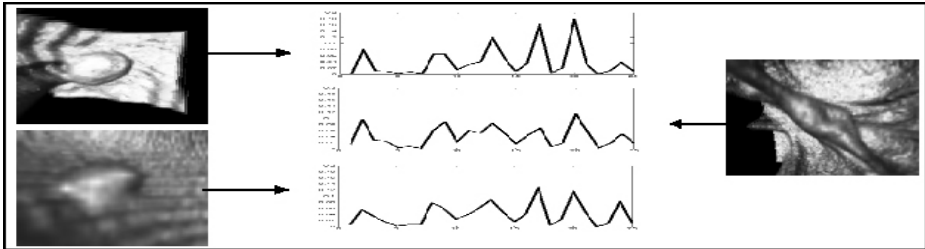


Fig. 5. Two polyps (left) and a very similarly shaped fold structure (right). As expected, the signatures are very similar, yet distinguishable.

used to test the shape. This cross-validation scheme was repeated for each candidate shape in the data set. Table 1 summarizes the results obtained from different kernel functions. The main objective of this work is to be able to achieve an accuracy of 80 or more with a minimum 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 80%, which inherently reduces the radiologist's interpretation time by the same amount.

A more quantitative analysis of the experiments can be achieved by replacing the zero-crossing ("sign") in expression (A), with 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 Fig.6(a).

Finally, we have analyzed the response of the system to different sizes of polyps. In previous studies, it has been shown that, polyps less than 1 cm in size are rarely malignant, with an incidence of carcinoma of only 1% [12]. However, the incidence of carcinoma increases to 37% in polyps greater than 2 cm in size [12]. Therefore, it is of critical importance to remove adenomatous polyps as small as 1 cm to impact the mortality of this disease. In figure 6(b), we illustrate the corresponding performance curves for polyps of size greater than 1 cm and smaller than 1 cm separately. The worse performance on smaller polyps can be explained by two reasons: First, folds in the human colon look similar to smaller polyps. Second and mainly, the resolution of current CT technology (voxel size is about $.7 \times .7 \times 3.5$ mm) is not high enough to capture shape descriptors for small objects. The results for bigger sized polyps were obtained using a 10-big-polyp training set. Since our methods rely on previous statistics, a bigger training set will improve the sensitivity of the system.

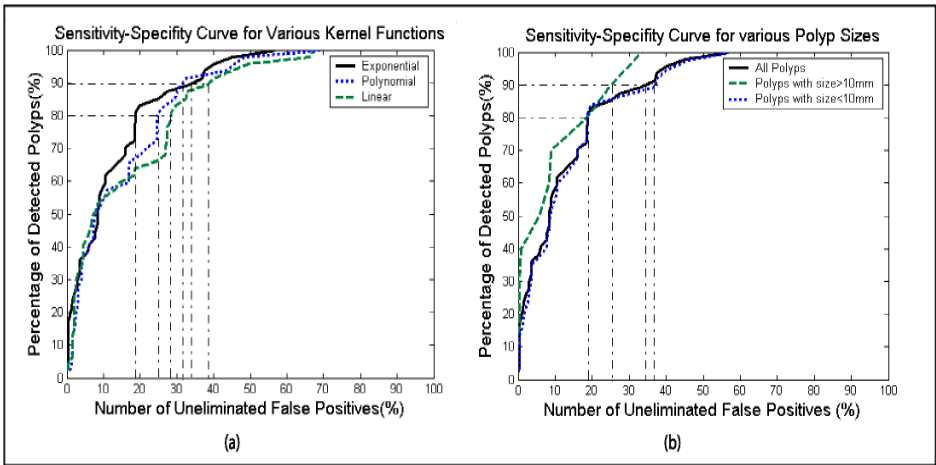


Fig. 6. Sensitivity vs. number of false positive detections. (a) Comparison between various kernel types. (b) Comparison between various polyp sizes.

Table 1. Improvements obtained by different kernel functions.

	Linear Kernel	Polynomial Kernel	Exponential Kernel
FP rate for 80% polyp detection	28.51	24.89	18.87
FP rate for 90% polyp detection	38.15	31.32	33.33
FP rate for 100% polyp detection	67.07	69.88	57.03

4 Conclusions and Future Work

Virtual colonoscopy is a promising new medical 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, computer aided diagnosis is necessary to make the approach efficient and cost-effective. Previous 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. To be more accurate, we need a method that is capable of differentiating polyps from other normal healthy structures in the colon. In this study, we proposed a learning approach that yields a good 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 new 3-D pattern analysis approach, which combines the information from many random images to generate reliable shape signatures. We also show that the use of support vector machines is capable of distinguishing implicitly the differentiating characteristics of polyps and healthy tissue, thus improving classification rates.

There are many directions for future investigation. First, we would like to analyze support vectors to observe the differentiating characteristics of polyps and healthy tissue. These observations might give valuable feedback in designing new features for

the system. In addition, while the results reported in this paper are promising, more extensive case studies need to be carried out, and more comprehensive statistics must be collected. Finally, studies integrating these computer aided detection schemes with radiologist readers will be used to measure potential improvements in sensitivity and efficiency compared with unassisted radiologist interpretation.

References

- [1] Wingo P.J., Cancer Statistics, *Ca Cancer Journal Clin*, 1995; 45:8-30.
- [2] Thoeni R.F., Laufer I. "Polyps and cancer," *Textbook of Gastrointestinal Radiology*, Philadelphia: W.B. Saunders, 1994; 1160.
- [3] Winawer S.J., Zauber A.G., Ho M.N., O'Brien M.J., Gottlieb L.S., Sternberg S.S., Waye J.D., et. al. "Prevention of colorectal cancer by colonoscopic polypectomy," The national polyp study workgroup, *N Engl J Med*. 1993; 329:1977-1981.
- [4] Dachman A.H., Kuniyoshi J.K., Boyle C.M., Samara Y., Hoffman K.R., Rubin D.T., Hanan I., "Interactive CT colonography with three-dimensional problem solving for detection of colonic polyps," *American Journal of Roentgenology*, 1998; 171:989-995.
- [5] Summers R.M., Beaulieu C.F., Pusanik L.M., Malley J.D., Jeffrey R.B., Glazer D.I., Napel S., "Automated polyp detector for CT colonography: feasibility study," *Radiology* ,216(1)284-90, 2000.
- [6] Paik D.S., Beaulieu C.F., Jeffrey R.B., Jr., Karadi C.A., Napel S., "Detection of Polyps in CT Colonography: A Comparison of a Computer-Aided Detection Algorithm to 3D Visualization Methods," *Radiological Society of North America 85th Scientific Sessions*, Chicago, November 1999.
- [7] H. Yoshida, Y. Masutani, P.M. MacEneaney, K. Doi, Y. Kim, A.H. Dachman, "Detection of colonic polyps in CT colonography based on geometric features," *Radiology*, vol. 217(SS), pp. 582-582, November 2000.
- [8] Göktürk S.B., Tomasi C., "A graph method for the conservative detection of polyps in the colon," *2nd International Symposium on Virtual Colonoscopy*, Boston , October 2000.
- [9] Hu M.K., "Visual pattern recognition by moment invariants," *IRE transactions on information theory*, vol. IT-8, pp 179-187, 1962.
- [10] A.Gersho and R.M. Gray, *Vector Quantization and Signal Compression*, Kluwer Academic Press, 1992.
- [11] Vapnik V., *Statistical Learning Theory*, New York, 1998.
- [12] Hermanek P., "Dysplasia-carcinoma sequence, types of adenomas and early colo-rectal carcinoma," *European Journal of Surgery* , 1987, 13:141-3.