Automatic Detection of Colorectal Polyps in Static Images

Da-Chuan Cheng[1*], Wen-Chien Ting[2], Yung-Fu Chen[3], and Xiaoyi Jiang[4]
[1*]*Department of Biomedical Imaging and Radiological Science, China Medical University, Xueshi Road 91, Taichung, Taiwan*
dccheng@mail.cmu.edu.tw

[2]*Colorectal Surgery, China Medical University Hospital, Taichung, Taiwan*
milka3670@pchome.com.tw

[3]*Department of Health Services Administration, China Medical University*
yungfu@mail.cmu.edu.tw

[4]*Department of Mathematics and Computer Science, University of Münster, Germany*
xjiang@math.uni-muenster.de

Received Day Month Year
Revised Day Month Year

ABSTRACT

Colorectal cancer continues to be one of the leading causes of mortality worldwide. Scanning using colorectal endoscopy is a useful and common method in clinical examinations. However, the scanning and polyps detections are performed by physicians. Failures to detect polyps might be caused due to lack of experience or knowledge. The purpose of this paper is to discover a scheme able to distinguish polyps from normal tissue in static images off-line.

Texture features are studied for the discrimination between polyps and normal tissue. Two useful and simple features are proposed. The student's t-test is applied in selecting useful features to reduce the computation time. The support vector machine is used as a classifier to identify the position of polyps. A study on the numbers in the training patterns is done in order to select an optimal ratio between the polyps and non-polyps sub-images.

Seventy-four colonoscopic images are collected to test the system. Half are used as training images and half for testing. The experimental result shows the system can identify all polyps if the colonoscopic images contain a single polyp. The sensitivity is 86.2% and the false-positive rate is 1.26 marks per image.

Keywords: colorectal polyps detection; textural features; support vector machine.

1. INTRODUCTION

Colorectal cancer was the third leading cause of cancer-related deaths in the United States in 2003 [1]. In this report, an estimated 105,500 colon and 42,000 rectal cancer cases were expected to occur and about 57,100 deaths are expected to occur in 2003. The death due to colon cancer can be prevented and cured through early detection. Therefore, early diagnosis is critically important for the patient's survival.

Nowadays, two kinds of screening are common for early polyps detection: computerized tomography (**CT**) and colonoscopy. The examination using CT has certain advantages. However, the disadvantages are:

(1) once some polyps are detected, the patient has to be re-examined via colonoscopy to remove the polyps.
(2) some small and tiny polyps are difficult to be seen in CT. If such polyps remain in the colon, they can possibly grow into malignant lesions.

Colonoscopy is an accurate screening technique for detecting polyps of all size, which also allows for biopsy of lesions and removal of most polyps [2].

The eventual goal of our study is to develop an automatic system which is able to detect the colonic polyps on colonoscopic images in real time. This system would be available to help physicians to notice some polyps which are small or tiny. Normally, large or medium-size polyps can be eas-

[‡]Corresponding author: Da-Chuan Cheng, Department of Biomedical Imaging and Radiological Science, China Medical University, Xueshi Road 91, Taichung, Taiwan. Tel.: +886-4-22053366, ext 7803; Fax: +886-4-22334175; E-mail: dccheng@mail.cmu.edu.tw.

ily found by physicians. However, during surgery the physicians have less time to notice small or tiny polyps. Therefore, a **CAD** (Computer-Aided Detection) system would be helpful in polyps detection. To achieve this goal, a fast polyps detection algorithm should be developed and connected to the video-capture system.

The goal of this paper is to explore a method which is able to detect polyps in static images. At this stage, the speed of polyp detection is not the key point. Instead, the accuracy of detecting polyps is our intended achievement. Moreover, we found that the ratio of the two training pattern numbers is a critical factor when using **SVM** (Support Vector Machine) as a classifier.

There are not too many related studies, however, they can be categorized into two kinds: 1) CT-based and 2) colonoscopy-based methods. Most of the CT-based method [3,4] uses morphology such as the curvature (2D or 3D) as the feature to recognize polyps. This is because in CT images there is no texture information. The colonoscopy-based method is able to use two kinds of features: 1) shape (or morphology) and 2) texture information of polyps. Using texture as a feature is an unique advantage comparing to the analysis in CT images. The reason is [5]: The colonic mucosal surface is granular and demarcated into small areas. Changes in the cellular pattern (pit pattern) of the colon lining might be the earliest sign of polyps [6]. These texture alterations of the colonic mucosa surface can also be used for the automatic detection of colorectal lesions.

It is a long history using texture features in image processing and pattern recognition. Haralick [7] and Cohen [8] have proposed many popular texture features for the related application areas. In which, the gray level co-occurrence matrices approach is very common and useful in texture feature extraction, texture analysis, and classification. The study related most to ours uses color wavelet features in detecting colorectal polyps [5]. They used wavelet transform because of its advantage in multi-resolution analysis. This multi-resolution analysis is in the spatial as well as the frequency domain. Two filters were used to extract the information: low-pass and band-pass filters. Both filters were applied to all levels of different resolutions. Afterwards, the co-occurrence matrices approach was used to extract texture features. Four features were derived from the co-occurrence matrices: angular second moment, correlation, inverse difference moment, entropy, contrast [9], and dissimilarity [10]. The classification is performed via linear discriminant analysis (**LDA**). They claimed that using a more complex classifier would increase the number of parameters associated with the evaluation of the proposed feature set.

Another similar work [11] used texture features as well and compared many tools of analyzing features. These tools included Backpropagation Neural Network (**BPNN**), resilient propagation (**RPROP**), scaled conjugate gradient (**SCG**), and Marquardt algorithms. They combined **PCA** (principal component analysis) and BPNN together as a classification tool for discriminating normal and abnormal colon status.

Another study is similar to colorectal polyps analysis but on gastric polyps in endoscopic video [12]. In this study, they compared four texture features to find their performance in gastric polyps detection. The features they have compared are texture spectrum histogram, texture spectrum and color histogram statistics, local binary pattern histogram, and color wavelet covariance (**CWC**). Their finding was that the CWC feature is the best feature among all, reaching $88.6 \pm 2.3\%$.

The proposed detection scheme involves a) a novel feature extraction technique based on co-occurrence matrices with a simple difference of texture features of different color channels. b) the support vector machine is applied as a classification tool in our polyps detection scheme. This paper is a revised version of our previous work in [13].

The rest of the paper is organized as follows. In Section 2 we describe the method and equipment used to acquire the colorectal images. The texture features are proposed in Section 2.3. The experimental results are shown in Section 3 followed by the discussion and conclusion.

2. MATERIALS and METHODS

2.1. Materials: Raw data

The endoscopic video system (Fujinon EPX-402 & EC-410) was used for whole colon examination. Before examinations, patients were asked to undertake colon preparation in 2 days with a laxative in order to clean the residua. The colonoscopy was intruded into the cecum via the anus and air was inflated inside via a colonoscope tube into the colon lumen to achieve a full view of the colon mucosa. During examination, videos were shown on the monitor and were recorded. The polyp images were digitally captured onto a computerized reporting system and were saved in BMP or JPG format by demand.

The image's spatial resolution is 378×254 pixels in three colors (R, G, and B). We collected 113 images. Among them, a total of seventy-four colonoscopic images were used to test our system. Thirty-seven images were used as training images and thirty-seven images were used for testing. Thirty-nine images were not used due to inconsistency in manual ground truthing, see below.

2.2. Ground Truth

We have developed a GUI (Graphic User Interface) system to help us in selecting the polyps' boundaries. Two students were trained to select polyps. Polygons are used to describe

the polyps' boundaries. The manual polyps selection was performed in all (113) collected colonoscopic images. Afterwards, two professional physicians examined the selection results and gave their judgment independently. Thirty-nine images were not selected for testing purposes because their judgments were not consistent. The remaining seventy-four images formed our database, in which all images have manual polyps boundary selections as our ground truth.

2.3. Methods: Cooccurrence matrix

The gray level co-occurrence matrix (**GLCM**) is a well-known and popular technique for extracting texture information from images. Conner and Harlow [14] have shown that GLCM is a more powerful technique than gray level difference matrix (**GLDM**), gray level run length method (**GLRLM**), and the power spectral method (**PSM**). Ohanian and Dubes [15] showed results that GLCM features perform better than fractal, Markov Random Field, and Gabor filter features for classification of texture.

Over twenty features appear in the literature which can be used to extract information from co-occurrence matrices [16]. Most of them work on gray level image in analyzing co-occurrence matrix (**CM**). The color co-occurrence matrix (**CCM**) is seldom applied because of its lower efficiency in computation time. However, colonoscopic images are color images where color is a very important feature in polyps detection tasks. Omitting the color information will cause difficulties in discriminating colonic polyps and normal tissues. Another successful technique for texture description in medical images is the texture description based on Random Sets [17]. This method offers over 170 features for texture description. The selection of the right features is done by decision tree induction in this work. In this paper, we propose a method which can apply color texture information.

2.4. Color Texture Feature

Our colonoscopic images have three color channels: R, G, and B. Based on some experiments, it was found that the green and blue channel images are good in discriminating normal and abnormal mucosal surface. The experiments were conducted on a total of thirty-four colorectal images, in which 120 sub-images (20 normal and 100 abnormal tissues) of size 64×64 are randomly selected. (The reason that only 34 images are selected for training is that the other images have smaller polyps which cannot offer sub-images of the size 64×64.) The size of co-occurrence matrices is 256×256. Four directions ($0^o, 45^o, 90^o, 135^o$) are considered when the co-occurrence matrices are calculated. Seven texture features are calculated from the co-occurrence matrices and their corresponding p-values for each image are calculated. The averaged p-values of total 34 images are shown in Table 2.4. In this table, only 0^o is shown. The results of the other three directions are similar. From this table, we conclude that the blue and green channels are more discriminable in the three color channels.

Let I denote a colonoscopic image and let I_s denote an arbitrary sub-image sampled from I. All features (F_1 - F_8) as follows are calculated based on the sub-image I_s. The first texture feature is:

$$F_1 = 1 - \frac{1}{1+\sigma_g}$$

where σ_g denotes the standard deviation of the green channel image.

It is found that the colonoscopic images have different colors. The average intensity is then a useful feature. The second texture feature is simply a combination of the average intensity of all three channel images:

$$F_2 = \frac{\bar{r}^2}{\bar{g}\bar{b}}$$

where $\bar{r}$, $\bar{g}$ and $\bar{b}$ are the mean intensity of the red, green, and blue channel images, respectively. However, the ratio of the mean intensity cannot represent the difference very well. Two measurements are added as features:

$$F_3 = \frac{\bar{r}-\bar{g}}{256}, \text{ and } F_4 = \frac{\bar{r}-\bar{b}}{256}$$

Based on the sub-images selected before, we further examine the p-values of feature $F_2 - F_4$. Their p-values are 0.1801, 0.0345, and 0.0246, respectively. Notably, these three features can be calculated faster than the features extracted from the co-occurrence matrices as listed in Table 2.4.

Let the co-occurrence matrix (CM) of a certain sub-image I_s with an offset $(0, 1)$ be denoted as $p(i, j)$. The CM made from green and red channel images are represented as $p_g(i, j)$ and $p_r(i, j)$, respectively. The contrast and energy features are utilized. However, a small modification is proposed as follows:

$$F_5 = \sum_{i,j} |i-j|^2 p_g(i,j) - \sum_{i,j} |i-j|^2 p_r(i,j)$$
$$F_6 = \sum_{i,j} p_g(i,j)^2 - \sum_{i,j} p_r(i,j)^2$$

The difference of contrast is used as a feature instead of the original one.

Another co-occurrence matrix with different offset $(-2, 0)$ is utilized and denoted as $q(i, j)$. The reason for choosing -2 instead of -1 is that we found the difference between different angles to be very limited. The co-occurrence matrix is controlled by two parameters: angle and step length. We therefore changed another parameter from 1 to

2 simultaneously. This matrix is made from I_s as well. Similar features are defined:

$$F_7 = \sum_{i,j} |i-j|^2 q_g(i,j) - \sum_{i,j} |i-j|^2 q_r(i,j)$$

$$F_8 = \sum_{i,j} q_g(i,j)^2 - \sum_{i,j} q_r(i,j)^2$$

All features $F_1, F_2, \cdots, F_8$ form a feature vector to describe a certain sub-image I_s.

2.5. Feature Normalization

Let N_1 and N_2 denote the number of sub-images of normal and abnormal tissues in the database for training. Each sub-image has a feature vector $v = [F_1 F_2 \cdots F_8] \in \mathbb{R}^{8\times 1}$. All feature vectors form a feature matrix $[v_{n,1}\ v_{n,2}\ \cdots\ v_{n,N_1}\ v_{a,1}\ v_{a,2}\ \cdots\ v_{a,N_2}] \in \mathbb{R}^{8\times(N_1+N_2)}$, where v_n and v_a denote the feature vector of normal and abnormal sub-images, respectively.

It is very common to normalize features before using them in computer vision. In this study, we simply choose the maximum value in the matrix and the feature matrix is divided by this maximum to keep all elements less or equal to one. This normalized feature matrix is used to train the classifier stated in Section 2.7.

2.6. Polyps training

The training patterns contain polyps and normal tissue sub-images. All sub-images are selected via a **GUI** (Graphic User Interface) on Matlab platform by a well-trained expert. The size of all sub-images are the same in 33×33 pixels. with R, G, and B channels. The size is determined due to the following three reasons: 1) There are different sizes of polyps in our colonoscopic images. Some polyps are very large and some are smaller than 33×33. If the window size is too large, it is possible to include the normal tissue in the window and the abnormal to normal ratio might be less than 50%. Under this situation, the feature of polyps might cause ambiguities in discrimination. On the contrary, if the window is too small, it is hard to represent the texture changes in a certain area. The larger the population of pixels in the sub-image, the more informative the features are expected to be. 2) In [12], they chosed 32×32 as a window size. In their study, they selected the abnormal tissues with exclusion of the region with strong light reflections. Based on this criterion, the window size cannot be too large. In our images, we have also strong light reflections on polyps. Although we have no such criterion, it is not good to include a large area of strong light reflections. Under this consideration, the window size cannot be too large. 3) Partial inclusion of strong light reflections in the sub-image is allowed in our scheme because the reflection is also a good feature in representing a polyp. It is found that most polyps contain more or fewer strong light reflections.

Our database contains sub-images with normal tissues and abnormal tissues (polyps). The pattern number of normal tissues is larger than the abnormal ones. This is because most images contain only one or two polyps and most regions are normal tissues in an image. Therefore, it is necessary to select more sub-images as training patterns for normal tissues.

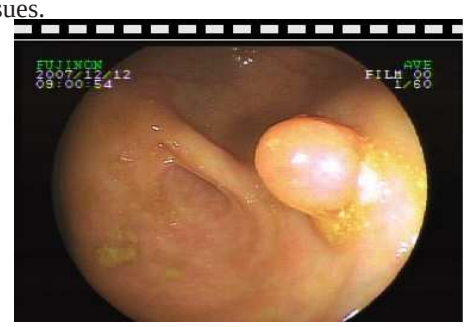

Fig. 1 A typical colonoscopic image with a polyp. There are some strong light reflections on the mucosal surface.

Some criteria for selecting training patterns are considered. The selected sub-images from a polyp image are in general not overlapping. This criterion works also in selecting polyp sub-images. Figure 1 shows one typical colonoscopic image. However, we have to note that some of our images have much smaller polyps as shown in this figure. In such cases, it is more difficult to detect them. This is also one of the reasons that the size of training patterns is chosen as 33×33, which is much smaller than 64×64.

The selection of polyps' sub-images is in general not overlapping. The strong light reflection is considered as a feature so that it can be included in sub-images. However,

Table 1 Averaged p-values of the corresponding features.

	Contrast	Homogeneity	Energy	Correlation	Entropy	Inverse difference moment	Smoothness
Red	0.1358	0.1661	0.1792	0.2297	0.2383	0.1661	0.1675
Green	0.1799	0.0798	0.0401	0.2501	0.0923	0.0798	0.2153
Blue	0.2462	0.0849	0.0256	0.1264	0.0326	0.0849	0.1222

Note: Total 34 images. 0^o of the co-occurrence matrix.

the region of the light reflection should not be too large. It is suggested to separate them into different sub-images as shown in Fig. 2. Moreover, if the polyp is large, it is suggested to select as many sub-images as possible to cover the whole polyp.

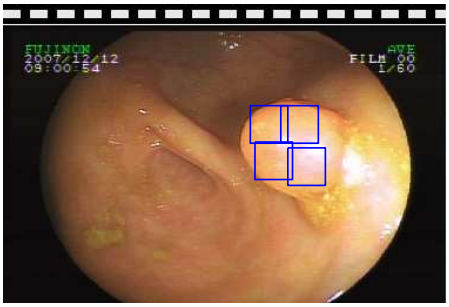

Fig. 2 Polyps sub-image selection is in general not overlapping. The strong light reflection cannot occupy too large a region in a sub-image.

The selection of sub-images of normal tissues included the dark regions, the reflections and so on. We suggest selecting partial reflections in a sub-image as well. This is because there are also reflections on normal mucosal surfaces; see Fig. 3.

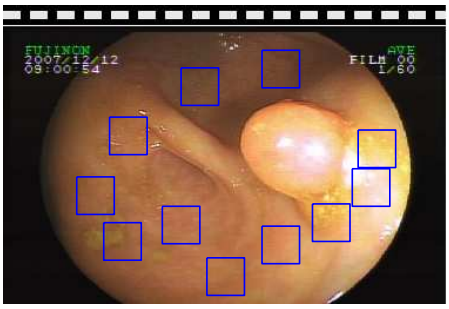

Fig. 3 The number of sub-images of normal tissue is in tendency more than the abnormal tissues.

2.7. Classification: Support Vector Machine (SVM)

Support Vector Machine (SVM) [18] is a kernel-based classifier. It is very popular in pattern recognition and computer vision. Many applications are based on SVM to have a nonlinear classification. Recently, SVM has been studied as a tool for predicting probability [19]. In this study, we utilize this tool to be a two-class classifier. The radial basis function is selected as the kernel function. The parameter of SVM using radial basis function is: $\sigma = 1$.

A rectangle active region is defined in every test image. This rectangular area is of the same size and the same position in each image and it includes most colonic mucosal surfaces in images. This definition reduces the redundant calculation on the image border without reducing useful information.

Every test image is automatically divided into several overlapping sub-images. The sub-image is of the same size as the one in the training process, i.e., 33×33 pixels with three channels (R, G, and B). The occlusion ratio is 50% in both the x- and y-direction, i.e. the sliding window has 17 pixels translation in both directions. Under this construction, smaller polyps might not be divided into two sub-images.

All sub-images are fed into our well-trained SVM classifier to test if it contains polyps or not. If one sub-image is labeled as a polyp image, a rectangle will be marked on that position for observation.

2.8. Study on training pattern numbers

During this study we found that the training number is also an important factor on sensitivity and false positive rate. In order to systematically study their relationship, we made an experiment. We divided these 74 images into two classes: one for training and another one for test. In the training class, the training patterns are randomly selected, N_1 and N_2 denote the training numbers of abnormal (containing polyps) and normal patterns, respectively. Since some polyps are small, it is possible to select an abnormal pattern in which contains some part of normal tissue. Therefore, in the abnormal pattern selection, we define that the polyp area must be more than 70% of the whole selected pattern area. On the contrary, in the normal pattern, it is not allowed to include any part of polyp. The definitions of sensitivity and false positive rate are listed as follows:

$$\text{sensitivity} = \frac{N(TP)}{N(TP) + N(FN)} \tag{1}$$

$$\text{false positive rate} = \frac{N(FP)}{N(FP) + N(TN)} \tag{2}$$

where $N(*)$ is the number function and TP, FP, TN, FN denote true positive, false positive, true negative, false negative, respectively. The SVM with a radial basis function ($\sigma = 1, 2$) is used as the classifier. The test results are in Section 3.

3. RESULTS

A total of 74 images are carefully selected and saved in our database. They are representative containing different types of polyps, different colors and so on. Half are used as training patterns (images) and half as test images.

Sixty-one sub-images are manually selected by an expert as polyps sub-images and 283 sub-images are selected as normal tissue from the 37 training images. Figure 4 shows some examples of training sub-images of polyps. There are fewer or more strong light reflection spots on polyps surfaces. Figure 5 shows some example sub-images of normal tissue in the training patterns. Most polyps sub-images have strong light reflection spots. Some sub-images of normal

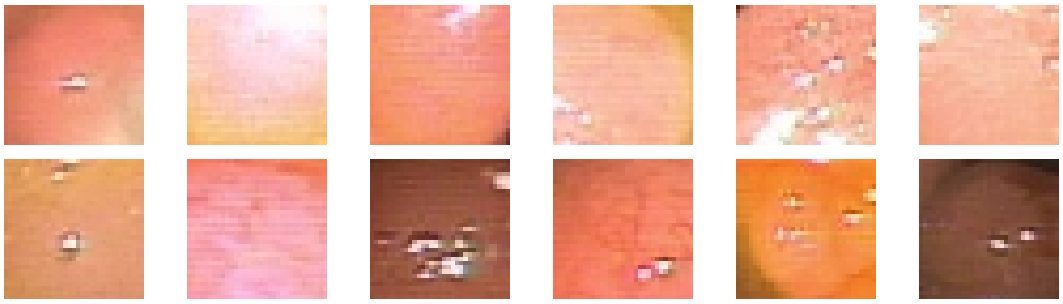

Fig. 4 Twelve from sixty-one polyps sub-images in the training patterns.

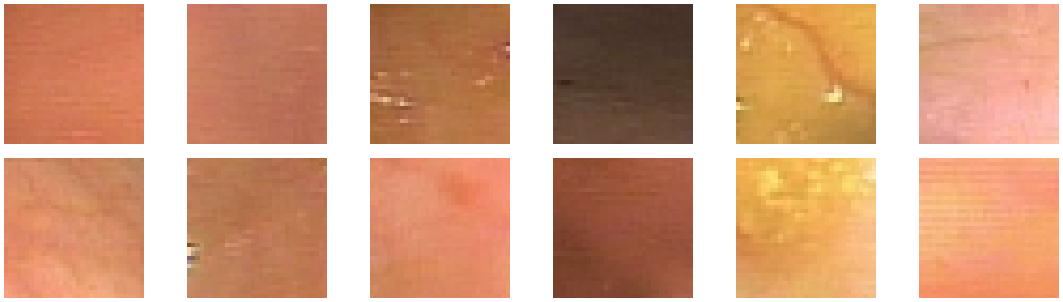

Fig. 5 Twelve from 283 sub-images of normal tissue in the training patterns.

tissue have vessels, and this is a good feature to be distinguished from polyps.

The normalized cross-correlation coefficient is a measurement if two signals are similar to each other or not. If the value between two signals is small then it is less correlated. This can be treated as a tool to see if two features are correlated to each other. The results are shown in Table 3.

Thirty-seven images are tested. Some contain a single polyp in an image and some contain multiple polyps. All polyps in single-polyp images are detected. However, some small polyps are not detected in the multiple-polyps images. One example is shown in Fig. 6. This might be due to the fact we did not choose tiny polyps in the training patterns. The sensitivity is 86.2% and the the false-positive rate per-image is 1.26.

Notably, the sensitivity calculation is slightly different to the formula used in [5]. Since the goal of our study is to identify the polyp, it is not necessary to find out every pixel of one polyp. Instead, if one pixel is correctly identified as a polyp, the rest pixels in this same polyp are considered as being identified.

Figure 7 shows some detection results, in which (a)-(i) have correct marks and false-positive and (j)-(o) have correct marks without false-positives. Two rectangles have correctly marked the polyp and one false-positive in Fig. 7(a). This might due to the color being similar to the polyp's color. Although two rectangles are marked as false-positive, we calculate only one false-positive since they are overlapping.

Figure 7(b) shows another detection result. A small polyp (bottom-right) is detected. Figure 7(c-d) and (f) detect the polyps. The false-positive might be because of the smooth mucosal surface which is like a polyp. In Figs. 7(f) and (i), the false-positives have small strong light reflection spots and the mucosal surface is smooth. Both features are similar to polyps.

In Fig. 7(k), a large polyp is detected without any false-negative. In this image, the colon is not well cleaned so that the image color is very different from most of the other images. Two marks are correct on the same polyp.

In Section 2.5, the feature normalization is done across all features. This is chosen because the observation that this type of normalization can slightly reduce the false-positive rate.

In the study on the relationship between the ROC (sensitivity versus (1-specificity)) and the training pattern number. Table 3 shows our result. N_1 and N_2 denote the number of the sub-images randomly selected from each training image from polyps and non-polyps (normal tissue), respectively. We found that an increasing ratio of N_1/N_2 will increase the sensitivity, however, the false positive rate will be increased as well. Moreover, the same N_1/N_2 ratio gets similar result. In order to get a lower false positive rate and a higher sensitivity, we have to make a compromise. We suggest selecting an N_2/N_1 ratio between 4 and 5. This is the reason why we chose 61 and 283 training sub-images ($N_2/N_1 = 4.6$) for the polyps and non-polyps, respectively.

Table 2 The normalized cross-correlation coefficients of all features.

1	2	3	4	5	6	7	8	9	
1.0	-0.2441	-0.0490	0.1008	0.1890	0.3410	-0.1878	0.3534	-0.1562	1
	1.0	0.4542	0.5738	0.4368	-0.1034	0.3621	-0.0976	0.3258	2
		1.0	0.7374	0.1987	0.0378	0.0061	0.0270	-0.0164	3
			1.0	0.8085	0.1556	0.0908	0.1166	0.0630	4
				1.0	0.1929	0.1264	0.1457	0.1056	5
					1.0	-0.1298	0.8676	-0.0741	6
						1.0	-0.0822	0.9819	7
							1.0	-0.0240	8
								1.0	9

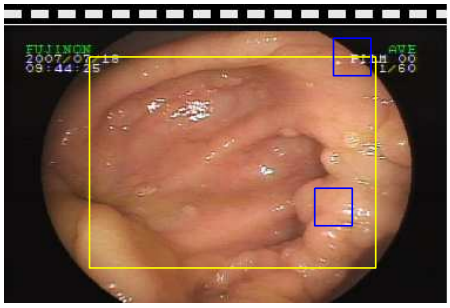

Fig. 6 The multiple polyps in the image are not well detected.

Table 3 The relationship between the sensitivity, FPR and the training pattern number (ratio).

σ	N_1 \ N_2		$2N_1$	$3N_1$	$4N_1$	$5N_1$
1	6	Sensitivity	0.9352	0.8565	0.8800	0.2083
		FPR	0.2714	0.1194	0.0970	0.0119
	8	Sensitivity	0.9352	0.9259	0.8565	0.5972
		FPR	0.2593	0.1569	0.0527	0.0298
	10	Sensitivity	0.9491	0.9259	0.8843	0.8194
		FPR	0.2562	0.1707	0.0841	0.0677
	12	Sensitivity	0.9491	0.8981	0.8515	0.5694
		FPR	0.2570	0.1358	0.0666	0.0367
2	6	Sensitivity	0.9167	0.8657	0.8519	0.7593
		FPR	0.1801	0.1187	0.0957	0.0698
	8	Sensitivity	0.8981	0.8287	0.8102	0.8148
		FPR	0.1755	0.1132	0.0916	0.0788
	10	Sensitivity	0.9491	0.8380	0.8009	0.8009
		FPR	0.1628	0.1148	0.0891	0.0752
	12	Sensitivity	0.9167	0.8380	0.8009	0.8102
		FPR	0.1690	0.1043	0.0681	0.0744

Note: Using SVM as a classifier: radial basis function.
FPR: False positive rate.

The program is written on the Matlab platform. It takes about 13 seconds to process one colorectal image of the size 378×254 at a PC with a 1.83GHz Intel Centrino Duo CPU and 2GB RAM.

4. Discussions

Regarding the training pattern selection, we found that the detection accuracy is better if the training patterns are selected manually comparing to those selected randomly. This might be due to following reasons: 1) The overlapping area can be reduced in the manual selection. 2) The larger polyps has larger area, therefore, the larger polyps get more training patterns. Thus, we can convince that most area of polyps in the training data set can be included.

We have made some experiments on different ratios of selected training pattern numbers, i.e., the ratio of polyp and normal tissue. Based on experiments, we found the false-positive is increased and the false-negative is decreased if the polyp's training pattern number is increased. This phenomenon is more significant if the overlapping area is in-

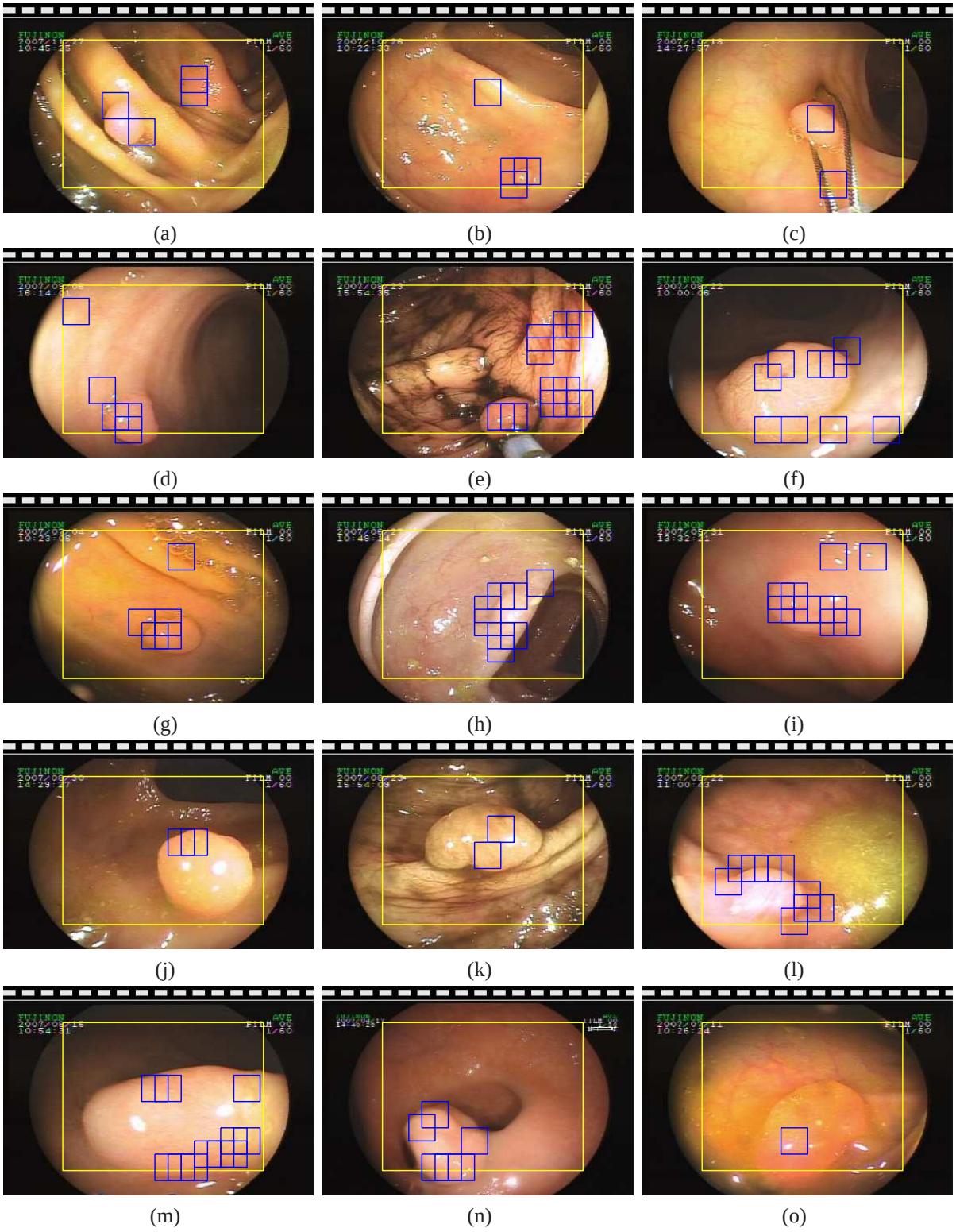

Fig. 7 Polyps detection results. (a)-(i) have correct marks and false-positive. (j)-(o) have correct marks without false-positive.

creased in polyp's training pattern selection.

According to our experiences, color is an important aspect on recognizing polyps. This might be due to the following reasons: 1) Normal colonic mucosa contains many capillaries. These capillaries are visible under the light illumination. However, the polyps contains rarely capillaries. The capollaries have good contrasts in blue and green channels. These characteristics make the two channels to be good fea-

ture sources. 2) There are mirror reflection almost in every colonic image. The strong reflection is a satuation in colors and therefore they cannot offer any information. However, we observe that almost all polyps contain mirror refletions. This provides a hint that the polyp may exist in a place where there is a mirror reflection. The mirror reflection can be detected using the satuation in all three color channels.

This system can offer the physicians an alternative way to detect polyps off-line. The goal is to confirm that all polyps have been extracted after the endoscopy examination. The white light visualization cannot offer the opportunity of recognizing malignant tumors from polyps. The future work is using narror band imaging (NBI) technique combined with image processing technique to identify malignant tumors from polyps.

5. CONCLUSION

In this study, we have proposed a new feature set in detecting colorectal polyps on colonoscopic images. Seventy-four colonoscopic images are carefully selected containing a wide range of polyps types, polyps sizes, and different colors. Half of them are used as training images and half of them are used as test images. The support vector machine is utilized as a classifier. The result shows that the sensitivity reaches 86.2% and the false-positive rate is 1.26 per image.

The current results are promising and we intend to extend our work in various ways. One idea is to combine the morphologic information with the texture information to get a more accurate detection result. In addition, it may be helpful to investigate the integration of different families of texture features, which are evaluated over multiple windows of varying sizes [20]. More generally, we expect to achieve higher detection performance by classifier ensemble techniques [21,22].

Conflict of Interest None declared.

ACKNOWLEDGMENTS

The authors would like to acknowledge China Medical University for supporting this research work under project CMU-95-280.

REFERENCES

1. Cancer Facts and Figures, American Cancer Society (2003)
2. Rex D, Weddle R, Pound D, O'Connor K, Hawes R, Dittus R, Lappas J, Lumeng L, Flexible sigmoidoscopy plus air contrast barium enema versus colonoscopy for suspected lower gastrointestinal bleeding, *Gastroenterology* **98**:855, 1990.
3. Taylor SA, Halligan S, Slater A, Goh V, Burling DN, Roddie ME, Honeyfield L, McQuillan J, Amin H, Dehmeshki J, Polyp detection with ct colonography: Primary 3d endoluminal analysis versus primary 2D transverse analysis with computer-assisted reader software, *Radiology* **239**(3):759, 2006.
4. Chowdhury TA, Whelan PF, Ghita O, The use of 3d surface fitting for robust polyp detection and classification in ct colonography, *Computerized Medical Imaging and Graphics* **30**:427, 2006.
5. Karkanis SA, Iakovidis DK, Maroulis DE, Karras DA, Tzivras M, Computer-aided tumor detection in endoscopic video using color wavelet features, *IEEE Trans. on Information Technology in Biomedicine* **7**(3):141, 2003.
6. Nagata S, Tanaka S, Haruma K, Yoshihara M, Sumii K, Kajiyama G, Shimamoto F, Pit pattern diagnosis of early colorectal carcinoma by magnifying colonoscopy: Clinical and histological implications, *Int. J. Oncol.* **16**:927, 2000.
7. Haralick RM, Shanmugam K, Dinstein I, Textural features for image classification. *IEEE Trans. on Systems, Man, and Cybernetics* **SMC-3**(6): 610, 1973.
8. Cohen P, Ledinh CT, Lacasse, V, Classification of natural textures by means of two-dimensional orthogonal masks, *IEEE Trans. on Acoustics, Speech, and Signal Processing* **37**(1):125, 1989.
9. Esgiar AN, Naguib RNG, Sharif BS, Bennett MK, Murray A, Automated feature extraction and identification of colon carcinoma, *Anal. Quant. Cytology Histology* **20**:297, 1998.
10. Esgiar AN, Naguib RNG, Sharif BS, Bennett MK, Murray A, Microscopic image analysis for quantitative measurement and feature identification of normal and cancerous colonic mucosa, *IEEE Trans. on Information Technology in Biomedicine* **2** 197, 1998.
11. Tjoa MP, Krishnan SM, Feature extraction for the analysis of colon status from the endoscopic images, *BioMedical Engineering OnLine* **2** 2003, http://www.biomedical–engineering–online.com/content/2/1/9.
12. Iakovidis DK, Maroulis DE, Karkanis SA, Brokos A, A comparative study of texture features for the discrimination of gastric polyps in endoscopic video, Los Alamitos, CA, USA, IEEE Computer Society, 575, 2005.
13. Cheng D, Ting W, Chen Y, Pu Q, Jiang X, Colorectal polyps detection using texture features and support vector machine, *In: Petra Perner, Ovidio Salvetti (Eds.) Advances in Mass Data Analysis of Images and Signals in Medicine, Biotechnology, Chemistry and Food Industry* LNAI **5108**, Springer Verlag, 62, 2008.
14. Conners RW, Harlow CA, A theoretical comparison of texture algorithms, *IEEE Trans. on Pattern Analysis and Machine Intelligence* **2**(3): 204, 1980.
15. Ohanian PP, Dubes RC, Performance evaluation for four classes of textural features, *Pattern Recognition* **25**(8):819, 1992.
16. Longstaff D, Walker R, Walker RF, Jackway P, Jackway P, Improving co-occurrence matrix feature discrimination, *In: Proc. of DICTA95, 3 rd International Conference on Digital Image Computing: Techniques and Applications*, pp. 643–648, 1995.
17. Perner P, Perner H, Mueller B, Mining knowledge for hep-2 cell image classification, *Journal Artificial Intelligence in Medicine* **26**: 161, 2002.
18. Schölkopf B, Smola AJ, Learning with Kernels *MIT Press*, Cambridge, MA, 2002
19. Luaces O, Taboada F, Albaiceta GM, Dominguez LA, Enriquez P, Bahamonde A, Predicting the probability of survival in intensive care unit patients from a small number of vari-

ables and training examples, *Artificial Intelligence in Medicine* **45**(1):63, 2009.
20. Puig D, Garcia M, Pixel-based texture classification by integration of multiple feature extraction methods evaluated over multisized windows, *International Journal of Pattern Recognition and Artificial Intelligence* **21**(7):1159, 2007.
21. Chen Y, Chen F, Yang J, Yang M, Ensemble voting system for multiclass protein fold recognition, *International Journal of Pattern Recognition and Artificial Intelligence* **22**(4):747, 2008.
22. Shih F, Fu G, Decision combination of multiple classifiers, *International Journal of Pattern Recognition and Artificial Intelligence* **22**(2):323, 2008.