Expert Systems with Applications 39 (2012) 10952–10958

Contents lists available at SciVerse ScienceDirect

Expert Systems with Applications

journal homepage: www.elsevier.com/locate/eswa

Automatic polyp detection for wireless capsule endoscopy images

Baopu Li [a,b,*], Max Q.-H. Meng [b]

[a] Shenzhen Institutes of Advanced Technology, The Chinese Academy of Sciences, Shenzhen, China
[b] Department of Electronic Engineering, The Chinese University of Hong Kong, Shatin N.T., Hong Kong SAR, China

ARTICLE INFO

Keywords:
Wireless capsule endoscopy image
Wavelet transform
Uniform local binary pattern
Polyp
Support vector machine

ABSTRACT

Wireless capsule endoscopy (WCE) opens a new stage for diagnosing gastrointestinal tract diseases since it enables direct visualization of the small intestine for the first time. However, it requires a clinician's long time inspection due to a great number of images produced by the procedure. Therefore, it may be beneficial to devise an automatic detection system to help clinicians identify problematic images. In this work, we attempt to design a computerized scheme aiming for polyp WCE image recognition though polyp in WCE images show great variations in appearance. This scheme utilizes a new texture feature to characterize WCE images, which integrates advantages of wavelet transform and uniform local binary pattern. With support vector machine (SVM) as a classifier, extensive experiments on our present image data, which consists of 600 normal WCE images and 600 polyp WCE images chosen from 10 patients, verify that it is promising to utilize the proposed scheme to detect polyp WCE images.

1. Introduction

Gastrointestinal (GI) tract related diseases, such as colon cancers and ulcerative colitis, are very common in many countries. It was reported that the bowel cancer cases in Hong Kong ranked the second of all and it was approaching the highest (Hong Kong Registry., 2007). In 2005, there were 3706 individuals that newly developed cancer in the large bowel. Most GI related cancers can be cured if they are detected at their early stages. Endoscopy is one of the most popular techniques for GI tract examination because it enables diagnosis based on analysis of real images and biopsy samples. Wireless capsule endoscopy (WCE), which is equipped with a micro-camera and wireless communication capability, eliminates a patients' pain and can directly view the entire small intestine for the first time, providing an effective approach to examine the small intestine, so it is widely applied in hospitals. Till now, WCE has shown its great value in evaluating GI bleeding, Crohn's disease, ulcer and other diseases existing in the GI tract Given Imaging Ltd. (http://www.givenimaging.com).

WCE travels through the GI tract passively propelled by the peristaltic movements or the gravity itself after it is ingested by a patient. Due to this reason, its journey in the digestive tract takes about 8 h. Visual inspection of the whole video is performed by a clinician and is time-consuming and labor-intensive. It costs an experienced clinician about two hours on average to analyze the whole video (Adeler & Gostout, 2003). In addition, abnormalities in the GI tract may be present in only one or two frames of the video, so they might be missed by physicians due to oversight sometimes. Such problems drive researchers to design computerized approaches to reduce the great burden of physicians. However, it should be pointed out that this goal is very difficult because features related to different diseases are not exactly known or well defined. Moreover, different diseases may show different symptoms in the digestive tract; even the same disease demonstrates great variations in color, shape and size.

Many studies have been investigated so as to decrease the burden of doctors. An interesting method of selecting MPEG-7 visual descriptors as feature extractor to recognize several diseases such as ulcers and bleeding in the GI tract was proposed in (Coimbra & Cunha, 2006). Based on this novel work, they continued to come up with two approaches to segment the GI tract into its four major topographic areas in a recent contributed paper (Cunha, Coimbra, Campos, & Soares, 2008), and the first software that utilizes these approaches aiming for WCE examination was also introduced in this paper. Szczypinski et al. proposed a novel model of deformable rings (Szczypinski, Sriram, Sriram, & Reddy, 2009) to interpret WCE videos, which allows a quick review of the whole video. Recently, we have investigated bleeding and ulcer region detection for WCE images in Li and Meng (2009a, 2009b), respectively, and preliminary experimental results validated that these schemes demonstrate encouraging recognition performances for bleeding and ulcer detection in WCE images.

We concentrate on polyp WCE image detection in this work since polyp is also one of the common diseases in the GI tract. To achieve this goal, we propose a new scheme that exploits color texture features and SVM. The proposed textural feature extraction

* Corresponding author at: Department of Electronic Engineering, The Chinese University of Hong Kong, Shatin N.T., Hong Kong SAR, China.
E-mail address: bpli@ee.cuhk.edu.hk (B. Li).

0957-4174/$ - see front matter © 2012 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.eswa.2012.03.029

method integrates wavelet transform and uniform local binary pattern to characterize normal WCE images and polyp WCE images. Uniform local binary pattern is robust to illumination variation and easy to implement, while wavelet transform provides a useful tool for multi-resolution analysis. Therefore, their hybrid leads to a better discrimination ability of textural features. As far as we know, we are the first to develop new color textural features for polyp detection in WCE images. Experimental results on our present data show that this new scheme achieves an encouraging performance of polyp detection when using SVM as a classifier.

The remainder of this paper is designed as follows. The new method of texture feature extraction using wavelet based uniform local binary pattern is discussed in the following section. In Section 3, SVM classifier is briefly reviewed. Section 4 presents experimental results in detail, and we draw some conclusions and make some discussions at the end of this paper.

2. Textural features

The normal regions and abnormal regions in WCE images can be differentiated by color textural features since color texture information is one of the primary features analyzed by clinicians. As illustrated in Figs. 1 and 2, respectively, WCE images with polyp and normal WCE images demonstrate different color textural characteristics on its mucosa surface, which encourage us to study color or texture features of these images. However, it should be emphasized that polyps in the GI tract exhibit great variations in appearances (Carlo et al., 2005), as can be observed in Fig. 1, which pose a great challenge to computer aided detection since a single feature may not be powerful enough to describe such a great diversity. Moreover, image resolution of M2A WCE images, data used in our experiments, is only 256×256, which may make this task more difficult to manage since traditional commercial endoscope images have a higher resolution such as 720×480 or even higher. Finally, there may be some disturbance such as bubble and faces existed in WCE images. We attempt to design a computer aided detection (CAD) system for polyp recognition for WCE images in this work. In a CAD system design, feature extraction plays relatively a more significant role compared to classifier although both

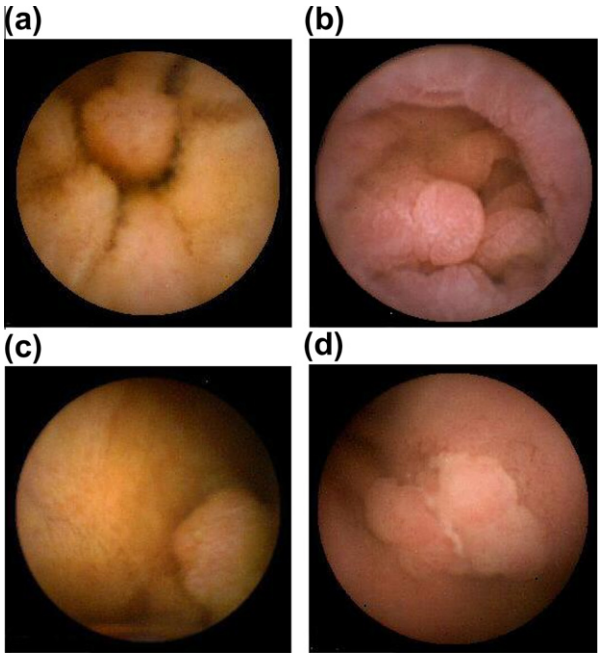

Fig. 1. Several WCE images with polyp.

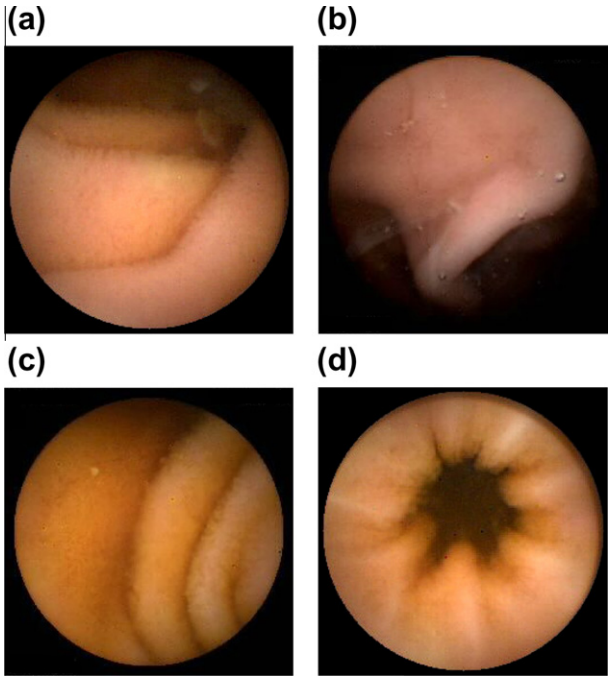

Fig. 2. Several normal WCE images.

of them are important; hence our attention is given mainly to feature extraction. In this section, we first delve into texture features for gray images and extend the proposed feature extraction scheme to color space since WCE images are color images.

2.1. Uniform binary pattern

Ojala, Pietikainen, and Harwood (1996) proposed the local binary pattern (LBP) texture operator, which is invariant against any monotonic grey scale transformation and is computationally simple. This technique is based on a two level version of the texture spectrum method (Wang & He, 1990) and describes spatial structure of the local image texture. The image pixels are first labeled by thresholding the difference between a pixel and its neighbors using the step function. For basic version LBP, neighbors mean eight direct neighbors of a pixel, and LBP has also been extended to include all circular neighborhoods with any number of pixels, which will be discussed later in this part. Then the values of pixels in the thresholded neighborhood are multiplied by binomial weights given to the corresponding pixels. Finally, values of the products are summed up to obtain an LBP number of this neighborhood. This process can be explained in Fig. 3. The basic version of LBP of a 3×3 neighborhood describes up to $2^8 = 256$ local texture patterns, and a 256-bin occurrence LBP histogram computed over a region is then employed for texture description.

In the extended case, Ojala et al. continued to come up with a simple yet efficient multi-resolution textural descriptor which is based on LBP (Ojala, Pietikainen, & Maenpaa, 2002). This operator is defined on a circular symmetric neighbor set, as illustrated in Fig. 4, where P pixels lie equally spaced on a circle of radius of R. P depicts the number of members in a circular symmetric neighbor sets and controls the quantization of an angular space, and R is the radius of a circle and determines the spatial resolution of this descriptor. By varying R, this descriptor can achieve multi-resolution analysis. The $LBP_{P,R}$ number can be defined as:

$$LBP_{P,R} = \sum_{p=0}^{P-1} s(g_p - g_c)2^p \quad (1)$$

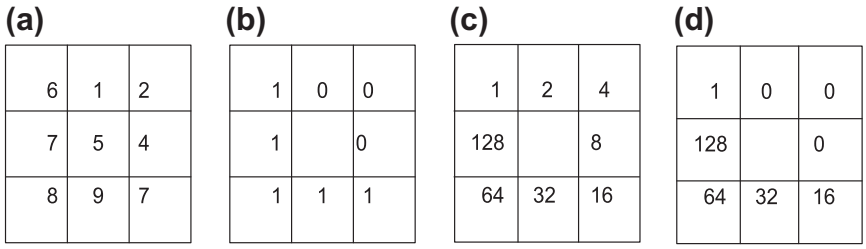

Fig. 3. One example demonstrating calculation of LBP for a pixel. (a) Example. (b)Thresholded with the center pixel. (c) The corresponding weights. (d) The product of weights and thresholded value. The final LBP value for this example is 1 + 16 + 32 + 64 + 128 = 241.

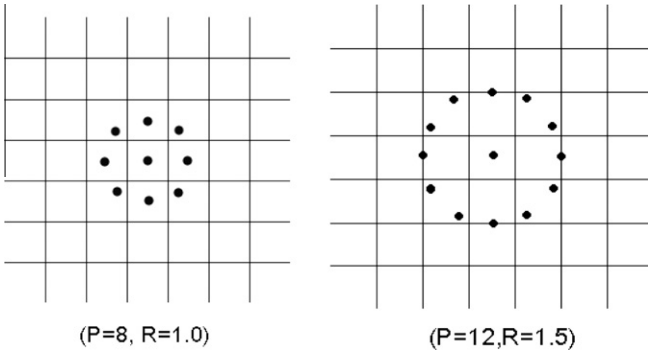

Fig. 4. Circular symmetric neighbors for different (P,R).

where g_c is the intensity of the center pixel and g_p is the intensity in the neighborhood pixel, and $s(x)$ is a function defined as below:

$$s(x) = \begin{cases} 1 & x \geqslant 0 \\ 0 & x < 0 \end{cases}$$

It is further observed that some local binary patterns are fundamental as they take up a vast majority of the whole LBP, and these fundamental patterns are called 'uniform'. The uniform pattern contains few transitions from 0 to 1. To formally define 'uniform' patterns, a uniformity measure is introduced, which corresponds to the number of spatial transitions (bitwise 0/1 changes) in the pattern. For instance, both pattern 00000000 and pattern 11111111 have transition values of 0, and patterns that have transition values of at most 2 are designated as 'uniform'. Based on the above discussions, a new operator, uniform LBP, is defined (Ojala et al., 2002):

$$LBP^{u2}_{P,R} = \begin{cases} \sum_{p=0}^{P-1} s(g_p - g_c) & \text{if } U(LBP_{P,R}) \leqslant 2 \\ P+1 & \text{otherwise} \end{cases} \quad (2)$$

where $U(LBP_{P,R}) = |s(g_{P-1} - g_c) - s(g_0 - g_c)| + \sum_{p=1}^{P-1}|s \quad (g_p - g_c) - s(g_{p-1} - g_c)|, g_c$ is intensity of the center pixel and g_p is intensity in the neighborhood pixel, P and R have the same meaning as before. This operator is a useful measurement of local image texture, and it is invariant against monotonic transformation of gray scale and simple to implement. The reason why we further turn to uniform LBP rather than LBP is for the feature dimensionality. For example, the number of features for each image using basic version LBP ($P = 8$, $R = 1$) is 256 while for the corresponding uniform LBP, it is only 10; and when using P=16, R=2, the LBP produces 65,536 features while for the corresponding uniform LBP, it is only 18. Hence, the problem of curse-of-dimensionality may be alleviated since we will apply uniform LBP to wavelet based sub-images, which yields a large number of features to represent each WCE image and will be detailed in the following subsection. So uniform LBP takes up the majority of the whole LBP and reduces the number of features needed to represent. More details about implementations of LBP and uniform LBP can be found in Ojala et al. (2002).

In the imaging process of WCE, images may suffer from illumination variations due to the specific imaging circumstances such as limited range of illuminations in the digestive tract. Furthermore, images of different patients in the database may be obtained under different illumination conditions. As such, it may be beneficial to consider illumination variation effects on textures of WCE images since common texture features may vary under illumination variations, while uniform LBP demonstrates robust performance to illumination change. Moreover, uniform LBP is computationally simple since this operator can be implemented with a few operations in a neighborhood and a lookup table Ojala et al. (2002). In addition, uniform LBP has shown its great success in classifying textures (Arasteh & Hung, 2006; Oliver, Lladó, Freixenet, & Martí, 2007; Sørensen, Shaker, & Bruijne, 2008). Due to the above reasons, we choose uniform LBP as the basis for texture feature extraction.

2.2. Discrete wavelet transform

Wavelet theory has been widely applied to analyze texture because wavelet provides a capable tool for multi-resolution analysis of an image. A lot of studies have been investigated on discrimination ability of wavelet-based features in various applications such as ultrasound image (Lee, Chen, & Hsieht, 2003) and mammogram (Wang & Karayiannis, 1998). As for WCE images, wavelet transform can provide zooming ability and local characterization to better analyze the mucosa of the inner GI tract, leading to more capable image analysis using information at different scales. Because wavelet transform for an image can be achieved with discrete wavelet transform (DWT), we briefly review DWT in this part.

The DWT is similar to a hierarchical sub-band system where sub-bands are spaced in frequency domain. For a 2D image, DWT is implemented with a separable filter-bank (Mallat, 1989), and an image is convoluted with a low pass filter L and a high pass filter H recursively. Due to the decomposition of an image using DWT, the image is transformed into four sub-images which are generally denoted as LL, LH, HL and HH. The LL sub-image comes from low pass filtering in both directions and it is the most like the original picture, so it is called the approximation component. The remaining sub-images are called detailed components. The HL is derived from low pass filtering along the vertical direction and high pass filtering along the horizontal direction and so has the label HL. The other two sub-images LH and HH have similar explanations. In our study, we apply three levels DWT to each color channel of a WCE image, and Fig. 5 illustrates such a representation of one color channel for this transformation.

As has been demonstrated that textural features are better encoded in detailed sub-images (Abyoto, Wirdjosoedirdjo, & Watanabe, 1998), we choose detailed sub-images, i.e., $\{HL_i, HH_i, LH_i, (i = 1, 2, 3)\}$ as the basis for textural feature analysis. Therefore, we further apply the aforementioned uniform LBP to each sub-image and obtain the discrete occurrence histogram of uniform

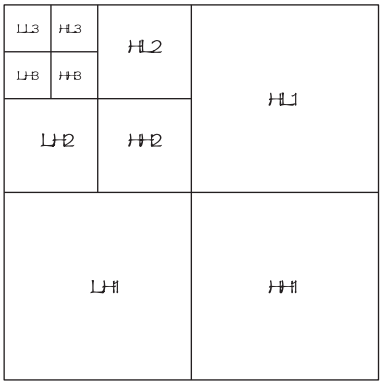

Fig. 5. Three-level image decomposed using DWT for one color channel.

LBP operator over each sub-image to describe textural features of WCE images. However, the problem of choosing which level detailed sub-images remains to be answered through experiments, which will be elaborated on later. Moreover, since motivation of the proposed scheme is multi-resolution analysis for WCE images, combination of different level detailed sub-images will also be discussed in experiments. It is also noted that in paper Ojala et al. (2002), the authors have already proposed a straightforward approach for multi-resolution analysis by combining multiple operators with different P and R. We will also compare our scheme with this method in experimental results.

Since WCE images are color images, we have to attack the problem of color textures. Though there are many algorithms for analyzing color and texture, processing of color texture images is still a relatively undeveloped field and there is no uniform and reliable model which can exploit both color and texture at the same time. Generally speaking, there are the following three strategies to handle color texture: (1) Each color channel is processed independently. (2) Using the interaction between channels to deal with color texture. (3) Color and texture are processed separately. Considering the complexity of the feature extraction method mentioned before, we will extend the proposed algorithm to RGB space and HSI space, respectively with each channel being processed separately. The reason why we choose them is because RGB color space is the most convenient color space and HSI color space is a representative space that separates color information into chromaticity and intensity.

3. Support vector machine

Support vector machine (SVM) is one kind of state-of-the-art classifiers employed in a lot of applications because they have many advantages such as minimum classification structural error, good ability to tackle the high dimensional problems and so on. Since proposed in 1990's by Vapnik (1995), this classification technique has been a hot topic in the field of machine learning. As mentioned before, the great variations of polyp appearances in WCE images inspire us to choose SVM as the classifier since it is generally believed that SVM often shows a strong generalization capability (Wang, 2005).

An SVM constructs a binary classifier from a set of labeled patterns called training set. Let $(x_i, y_i) \in R^N \times \{-1, +1\}$, $i = 1...m$ be such a set. The purpose here is to select a function $f : R^N \to \{\pm 1\}$ from a given class of functions such that f correctly classifies the test data (x, y). Based on the above assumption, SVM algorithm is able to find a hyper-plane defined by the equation

$$\omega \Phi(x) + \omega_0 = 0 \tag{3}$$

such that the margin of separation is maximized, where $\Phi(x)$ is a nonlinear mapping from the input space to the feature space and ω_0 is a scalar which can be estimated from Karush–Kuhn–Tucker complementary condition. It is shown that for the maximal margin hyper-plane (Vapnik, 1995),

$$\omega = \sum_{i=1}^{m} \lambda_i y_i \Phi^T(x) \tag{4}$$

where λ_i are the Lagrange multipliers that can be estimated through the maximization of

$$L_D = \sum_{i=1}^{m} \lambda_i - \frac{1}{2} \sum_{i=1}^{m} \sum_{j=1}^{m} \lambda_i \lambda_j K(x_i, x_j) \tag{5}$$

with respect to λ_i, and the following constraints hold at the same time: $\sum_{i=1}^{m} \lambda_i y_i = 0$ and $0 \leqslant \lambda_i \leqslant c$. In (5), $K(x_i, x_j)$ is called a kernel function and it is defined as the inner product $K(x_i, x_j) = \Phi^T(x_i)\Phi(x_j)$. Linear, polynomial, radial basis function (RBF) and sigmoid are among the most common functions used as SVM kernels, and RBF may usually perform better than the other kernels (Burges, 1998).

Given a test input vector x, the trained SVM yields an output which corresponds to the label of class, namely,

$$s = sign\left\{ \sum_{i=1}^{m} \lambda_i y_i K(x_i, x) + \omega_0 \right\} \tag{6}$$

where *sign* is the function that returns +1 for positive inputs and −1 for negative inputs.

4. Experimental results

To verify performances of the proposed scheme, especially the discrimination ability of the proposed textural features for polyp discrimination in WCE images, we carried out extensive experiments on our present data, which are discussed in detail as follows.

4.1. Experimental data

Two experts of the GI tract with about 7 years of experiences with WCE images selected a data set composed of 600 representative polyp images and 600 normal images from 10 patients' video data. These images are obtained from M2A WCE, a product of the Given imaging company, and resolution of these images is 256×256. These experts selected 60 normal images and 60 polyp images from each patient's video segments. Concerning normal images, we mainly mean those images that have healthy mucosal appearance. However, to make the normal image data more representative, some WCE images with a few bubbles and some indigested food or feces are also chosen by these experts. The original images are manually labeled to provide the ground truth. The image containing any polyp region is labeled as a positive sample; otherwise, it is labeled as a negative sample.

In order to prevent over-fitting of classification results, we exploited ten-fold cross-validation for all our classification experiments. Moreover, since the image data from the same patient may have similar appearance sometimes, all the images from one patient are put in either the training set or the test set. As such, each time 550 normal samples and 550 polyp samples from nine patients are used as training sets, while the left 50 normal samples and 50 polyp samples from the left one patients are used as test sets, and we repeat this procedure ten times. Finally, the average recognition rates are used to assess performance of classification.

4.2. Level choice and combination experiments

As mentioned before, discrimination ability of the proposed textural features depends on the choice of detailed sub-image in

different levels or combination of them in each color channel. Moreover, a straightforward idea of multi-resolution analysis based on LBP (MLBP) has been tested in paper (Ojala et al., 2002), so we will also compare our scheme (WLBP) with this method. In this part, we first carry out experiments of the proposed scheme, and then experimentally compare WLBP to MLBP.

Success of classification of WCE images using SVM is measured by accuracy, which is widely employed to assess performances of classification. The following equation explains definition of accuracy:

$$\text{Accuracy} = \frac{\textit{Number of Correct Predictions}}{\textit{Number of Positives} + \textit{Number of Negatives}} \tag{7}$$

Concerning the DWT implementation, we made use of Haar wavelet for its superior discriminating power which is demonstrated in Semler, Dettori, and Furst (2005). For implementation of SVM, we referred to LIBSVM (Chang and Lin, 2001), and RBF was found to be the optimal kernel function that yielded the best classification performance in our experiments. The optimal values of two parameters in SVM, i.e., penalty parameter and the kernel parameter of RBF, for the classification were found using a grid search approach (Chang and Lin, 2001), And the best classification results in terms of average accuracy were considered as performances of classification for SVM.

In our experiments, we first chose $P = 24$, $R = 3$ to experiment for wavelet based uniform LBP, the experiments with the other two choices for $P = 16$, $R = 2$ and $P = 8$, $R = 1$ will be reported subsequently. We record performances of the proposed textural features with $P = 24$, $R = 3$ extracted from different level detailed sub-images and their different combinations in Table 1. The standard derivation of the recognition results are also recorded in the table so as to show the stability of classification results. It is noted that level i ($i = 1,2,3$) in the first column of this table means that we choose the corresponding $\{HL_i, HH_i, LH_i\}$ sub-images in Fig. 5.

By reviewing Table 1, we find that using combination of the proposed features extracted from different level sub-images can improve the discrimination ability for polyp WCE image detection compared to features extracted only from one level. This is also expected since motivation of the proposed textural features is multi-resolution analysis and combing different textural features from sub-images in different levels may result in complementary effects for feature representation. Specifically, the best performance goes to the combination of features extracted from all three different levels, i.e., Level 1 + Level 2 + Level 3, with an encouraging accuracy of 91.6% and 91.0% in RGB and HSI space, respectively. As regards the color space choice, it can be observed that both RGB color space and HSI color space work fine for the proposed scheme since there is only slight difference between performances of the proposed schemes in these two color spaces.

The experiments of the proposed color textural features with the other two choices for $P = 16$, $R = 2$ and $P = 8$, $R = 1$ for our WCE image data are shown in Table 2 and 3, respectively. By comparing Table 2 and Table 3 to Table 1, we can conclude that the proposed color texture shows the best discrimination ability for our problem when we choose $P = 24$, $R = 3$. The reason for the better performance of $P = 24$, $R = 3$ can be ascribed to the fact that this choice provides a finer quantization of the angular space compared to $P = 16$, $R = 2$ and $P = 8$, $R = 1$, leading to a more capable characterization ability for the details in WCE images.

Table 1
Classification accuracy using wavelet based uniform local binary pattern ($P = 24$, $R = 3$) from different levels and their combinations.

	RGB	HSI
Level 1	87.2 ± 0.6	86.4 ± 0.8
Level 2	86.6 ± 0.4	85.7 ± 0.2
Level 3	85.2 ± 0.8	84.6 ± 0.3
Level 1 + Level 2	89.7 ± 0.5	89.3 ± 0.7
Level 2 + Level 3	88.7 ± 0.3	88.4 ± 0.2
Level 1 + Level 3	89.3 ± 0.1	89.1 ± 0.2
Level 1 + Level 2 + Level 3	91.6 ± 0.2	91.0 ± 0.1

Table 2
Classification accuracy using wavelet based uniform local binary pattern ($P = 16$, $R = 2$) from different levels and their combinations.

	RGB	HSI
Level 1	85.4 ± 0.2	83.5 ± 0.9
Level 2	83.4 ± 0.3	82.1 ± 0.2
Level 3	80.6 ± 0.6	80.4 ± 0.1
Level 1 + Level 2	86.4 ± 0.4	85.2 ± 0.0
Level 2 + Level 3	85.4 ± 0.3	84.2 ± 0.9
Level 1 + Level 3	87.2 ± 0.7	85.4 ± 0.3
Level 1 + Level 2 + Level 3	88.2 ± 0.6	86.7 ± 0.9

In Ojala et al. (2002), another kind of multi-resolution analysis is accomplished through integrating features provided by multiple uniform LBP operator of varying (P,R). It should be noted that uniform LBP here is applied to the original image. Moreover, we also combine the uniform LBP from each color channel in a color space to represent color texture features for a WCE image. Hence, we test this approach on our collected WCE image data and compare it to our proposed scheme. In experiments, we used $P = 8$, $R = 1$; $P = 16$, $R = 2$ and $P = 24$, $R = 3$ to implement such a multi-resolution analysis just as what has been done in Ojala et al. (2002). The recognition performance for polyp WCE images detection using $P = 8$, $R = 1$; $P = 16$, $R = 2$, and $P = 24$, $R = 3$ separately and their various integrations are listed in Table 4. It can be observed that the best performance among these choices is obtained by integrating the uniform LBP from all three scales and the best accuracy is 83.6% and 82.8% in RGB and HSI space, respectively. By comparing Table 4 to the best result of the proposed scheme, i.e., the result shown in Table 1, it can be noticed that the proposed scheme outperforms MLBP for polyp WCE image detection. And the improvement margin is 8.0% and 8.2% in RGB and HSI space, respectively, in terms of maximum average accuracy. The reason why the proposed WLBP outperforms MLBP may be due to the following two reasons. First, wavelet transform in the proposed scheme provides a more capable multi-resolution tool than the straightforward method of combination of different P and R. Secondly, the features extracted from $\{HL_i, HH_i, LH_i\}$ may be complementary since these sub-images show complementary information in different orientations, while combinations of uniform LBP with different P and R computed from the entire image may not have such complementary information.

4.3. Comparison with other methods

We compare the proposed scheme with the texture spectrum histogram (TSH) (Karkanis, Galousi, & Maroulis, 1999) and the color wavelet covariance (CWC) features (Karkanis, Iakovidis, Maroulis, & Karras, 2003) so as to further show performance of it. The reason why we choose them is that these methods have shown their capabilities in classification for traditional endoscopic images. TSH was proposed by Wang and He (1990) and it is based on texture units which characterize local texture information for a given pixel and its neighborhood. More details about this algorithm can be found in Wang and He (1990). CWC feature is a new technique to describe color textural features that are built upon the covariance of second-order textural measures in wavelet domain of color channels of an image. We use SVM as the classifier for these two features and the best parameters for SVM are also obtained using grid search method. The results of these two features are shown in Table 5.

Table 3
Classification accuracy using wavelet based uniform local binary pattern (P=8, R=1) from different levels and their combinations.

	RGB	HSI
Level 1	81.2 ± 0.5	80.4 ± 0.6
Level 2	79.2 ± 0.4	79.3 ± 0.6
Level 3	78.6 ± 0.2	78.1 ± 0.5
Level 1 + Level 2	83.4 ± 0.5	83.2 ± 0.2
Level 2 + Level 3	82.4 ± 0.1	82.3 ± 0.5
Level 1 + Level 3	83.5 ± 0.3	83.4 ± 0.4
Level 1 + Level 2 + Level 3	86.3 ± 0.6	84.6 ± 0.8

Table 4
Classification accuracy using uniform local binary pattern from different levels and their combinations.

	RGB	HSI
P = 8, R = 1	74.6 ± 0.9	73.8 ± 0.2
P = 16, R = 2	76.2 ± 0.1	75.4 ± 0.3
P = 24, R = 3	77.5 ± 0.3	77.2 ± 0.2
P = 8, R = 1 + P = 16, R = 2,	78.4 ± 0.4	78.2 ± 0.1
P = 16, R = 2 + P = 24, R = 3	81.2 ± 0.8	80.4 ± 0.3
P = 8, R = 1 + P = 24, R = 3	80.5 ± 0.5	79.5 ± 0.3
P = 8, R = 1 + P = 16, R = 2 + P = 24, R = 3	83.6 ± 0.7	82.8 ± 0.2

Table 5
Classification accuracy using CWC and TSH method.

	RGB	HSI
CWC	67.4 ± 1.5	65.2 ± 1.2
TSH	64.2 ± 1.3	62.6 ± 1.1

After reviewing Table 5, it can be noted that both CWC and TSH show inferior accuracy for polyp detection in WCE images in contrast to the proposed color textural features. Compared to TSH, the reason why the proposed features show better recognition rate for polyp detection in WCE images may be due to the fact that multi-scale based texture features can further improve the discrimination ability of the textural features compared to a single scale based texture. Compared to CWC, it may be ascribed to the fact that the proposed textural features are robust to illumination changes for WCE images obtained under different imaging circumstances in the GI tract.

5. Conclusions and discussions

A new scheme which utilizes wavelet based uniform local binary pattern and support vector machine to detect polyp WCE images has been proposed in this paper. The novel textural features combine the advantage of wavelet transform, i.e., a capable multi-resolution analysis tool, and that of uniform local binary pattern, i.e., robustness to illumination changes, leading to a more discriminative ability for polyp detection in WCE images. Comparative experiments with three other different texture features in RGB color space and HSI color space on our present WCE images show that this scheme is promising in detecting polyp WCE images and achieves the best detection accuracy of 91.6% in RGB color space despite the fact that polyp in WCE images show great variations in appearance.

However, there is still some room for detection accuracy improvement for the proposed scheme. One possible avenue of our future work along this direction is to investigate combined classifiers so as to improve detection accuracy since it is known that substantial recognition improvements may be obtained in difficult pattern recognition problems by combining or integrating outputs of multiple classifiers (Kuncheva, Bezdek, & Duin, 2001). Another possible point worth future research is to make some pre-processing for WCE images so as to filter out those WCE images that are too dark or have visual disturbances. Moreover, practical usage of the proposed whole system in hospitals remains to be tested after a larger number of the patient's video data is collected in the future.

Acknowledgements

This work is supported by the Hong Kong Research Grants Council (RGC) General Research Funds (415709 and 415611) and Innovation and Technology Support Programme (ITS/430/09) in Hong Kong, all awarded to Max Meng. We would like to thank to James Lau and Yawen Chan, two experts in Prince of Wales Hospital in Hong Kong, for providing us the WCE image data.

References

Abyoto, K. W., Wirdjosoedirdjo, S. J., & Watanabe, T. (1998). Unsupervised texture segmentation using multi-resolution analysis for feature extraction. *Journal of the Tokyo University of Information Science, 2*, 49–61.

Adeler, D. G., & Gostout, C. J. (2003). Wireless capsule endoscopy. *Hospital Physician, 5*, 14–22.

Arasteh, S., & Hung, C.-C. (2006). Color and texture image segmentation using uniform local binary patterns. *Machine Graphics & Vision International Journal, 15*, 265–274.

Burges, C. (1998). *A tutorial on support vector machines for pattern recognition*. Boston: Kluwer Academic Publishers.

Carlo, T., DeMarco, D., Smith, A., Livingston, S., Kim, B. A., Kuhn, J. A., et al. (2005). The utility of capsule endoscopy and its role for diagnosing pathology in the gastrointestinal tract. *The American Journal of Surgery, 190*, 896–901.

Chang, C. C., & Lin, C. J. (2001). LIBSVM: A library for support vector machines. http://www.csie.ntu.edu.tw/cjlin/libsvm.

Coimbra, M. T., & Cunha, J. P. S. (2006). MPEG-7 visual descriptors-contributions for automated feature extraction in capsule endoscopy. *IEEE Transactions on Circuits and Systems for Video Technology, 16*, 628–637.

Cunha, J. P. S., Coimbra, M., Campos, P., & Soares, J. M. (2008). Automated topographic segmentation and transit time estimation in endoscopic capsule exams. *IEEE Transactions on Medical Imaging, 27*, 19–27.

Given Imaging Ltd., http://www.givenimaging.com.

Hong Kong Registry. (2007). Cancer statistics, Hospital Authority of Hong Kong.

Karkanis, S. A., Galousi, K., Maroulis, D. (1999). Classification of endoscopic images based on texture spectrum. In *Proceedings of workshop on machine learning in medical applications, advance course in artificial intelligence* (pp. 63–189).

Karkanis, S. A., Iakovidis, D. K., Maroulis, D. E., & Karras, D. A. (2003). Computer-aided polyp detection in endoscopic video using color wavelet features. *IEEE Transactions on Information Technology in Biomedicine, 7*, 141–152.

Kuncheva, I. L., Bezdek, C., & Duin, P. W. (2001). Decision templates for multiple classifier fusion: An experimental comparison. *Pattern Recognition, 34*, 299–314.

Lee, W. L., Chen, Y. C., & Hsieht, K. S. (2003). Ultrasonic liver tissues classification by fractal feature vector based on M-band wavelet transform. *IEEE Transactions on Medical Imaging, 22*, 382–392.

Li, B., & Meng, M. Q.-H. (2009a). Computer-based detection of bleeding and ulcer in wireless capsule endoscopy images by chromaticity moments. *Computers in Biology and Medicine, 39*, 141–147.

Li, B., & Meng, M. Q.-H. (2009b). Computer aided detection of bleeding regions in capsule endoscopy images. *IEEE Transactions on Biomedical Engineering, 56*, 1032–1039.

Mallat, S. G. (1989). A theory for multi-resolution signal decomposition: The wavelet representation. *IEEE Transactions on PAMI, 11*, 674–693.

Ojala, T., Pietikainen, M., & Harwood, D. (1996). A comparative study of texture measures with classification based on feature distributions. *Pattern Recognition, 29*, 51–59.

Ojala, T., Pietikainen, M., & Maenpaa, T. (2002). Multi-resolution gray-scale and rotation invariant texture classification with local binary pattern. *IEEE Transactions on PAMI, 24*, 971–987.

Oliver, A., Lladó, X., Freixenet, J., & Martí, J. (2007). False positive reduction in mammographic mass detection using local binary patterns. In *Proceedings of MICCAI (10)*, (pp. 286–293).

Semler, L., Dettori, L., & Furst, J. (2005). Wavelet-based texture classification of tissues in computed tomography. In *Proceedings of the 18th international conference on computer based medical systems* (pp. 165–270).

Sørensen, L., Shaker, S. B., Bruijne, M. (2008). Texture classification in lung CT using local binary patterns. In *Proceedings of MICCAI (11)* (pp. 934–941).

Szczypinski, P. M., Sriram, R. D., Sriram, P. V. J., & Reddy, D. N. (2009). A model of deformable rings for interpretation of wireless capsule endoscopic videos. *Medical Image Analysis, 13*, 312–324.

Vapnik, V. (1995). *The nature of statistical learning theory*. New York: Springer Verlag.

Wang, L. (2005). *Support vector machines: theory and applications*. Springer.
Wang, L., & He, D. C. (1990). Texture classification using texture spectrum. *Pattern Recognition, 23*, 905–910.
Wang, T. C., & Karayiannis, N. B. (1998). Detection of micro-calcifications in digital mammograms using wavelets. *IEEE Transactions on Medical Imaging, 17*, 498–509.