View metadata, citation and similar papers at core.ac.uk
brought to you by CORE
provided by Universidade do Minho: RepositoriUM

DETECTING ABNORMALITIES IN ENDOSCOPIC CAPSULE IMAGES USING COLOR WAVELET FEATURES AND FEED-FORWARD NEURAL NETWORKS

Carlos S. Lima, Daniel Barbosa, Jaime Ramos[(1)], Adriano Tavares, Luis Carvalho and Luis Monteiro

Department of Ind. Electronics of University of Minho, Campus de Azurém, Guimarães, Portugal
carlos.lima@dei.uminho.pt

[(1)] Capuchos Hospital, Alameda Santo António dos Capuchos, Lisboa, Portugal

ABSTRACT

This paper presents a system to support medical diagnosis and detection of abnormal lesions by processing endoscopic images. Endoscopic images possess rich information expressed by texture. Texture information can be efficiently extracted from medium scales of the wavelet transform. The set of features proposed in this paper to encode textural information is named color wavelet covariance (CWC). CWC coefficients are based on the covariances of second order textural measures, an optimum subset of them is proposed. The proposed approach is supported by a classifier based on multilayer perceptron network for the characterization of the image regions along the video frames. The whole methodology has been applied on real data containing 6 full endoscopic exams and reached 87% specificity and 97.4% sensitivity.

Index Terms— Color texture, computer aided diagnosis, image analysis, medical imaging, wavelet features

1. INTRODUCTION

Conventional endoscopy is limited to the upper gastrointestinal (GI) tract, at the duodenum, and to lower GI tract, at terminal ileum. So the vast majority of the small intestine, which has a medium length of six meter, isn't seen by these techniques. Therefore the capsule endoscopy allows the visualization of the GI tract, reaching places where conventional endoscopy is unable to. Images are captured, at the rate of two frames per second, by a short-focal-length lens as the capsule is propelled by peristalsis through the gastrointestinal tract. The result is a seven hours video with more than 50.000 frames per exam. Average small bowel transit time is about 90 minutes [2], then capsule reaches the cecum and visibility is severely decreased giving a total average of 15.000 useful images. Usually the physician is required to view 60.000 images and to select the ones that he considers important. This task is boring, time consuming and prone to subjective errors since most of the frames are normal, so it claims for computational assistance.

The automatic detection of lesions can be based in textural alterations of the small bowel mucosa surface. In the proposed approach the video frame sequences are transformed in scale by using the wavelet transform, since it has been observed that the textural information is localized in the middle frequencies and lower scales of the original signal [3]. The discrimination of normal and abnormal regions relies on a texture analysis scheme, supported by the statistical color wavelet features of each frame. The construction of the texture feature space follows the multiresolution approach on the wavelets extracted from the color domain. In this study, the features were obtained from the coocurrence matrices of the wavelet transform of different color spaces, at different scales, so that we have a coocurrence matrix for each band analyzed in the color space. Then second-order-statistics are computed between color channels, for the same orientation.

The feed-forward neural networks are, perhaps, the most commonly used networks for classification purposes. They were the first type of artificial neural network devised. In this network, the information moves in only one direction, forward, from the input nodes, through the hidden nodes (if any) and to the output nodes. There are no cycles or loops in the network. The Multi-Layer Perceptron (MLP) networks are commonly used in classification problems, because they have the ability to detect complex non-linear relationships in the data. There is an extensive range of applications of these neural networks, and so, a vast theoretical and practical background in this matter[4].

This paper is focused in the features extraction process from the wireless capsule video frames, with a method based in the correlation of statistical descriptors of cooccurrence matrix calculated for midband wavelet coefficients of each color channel of a given frame. These features are the input of a MLP network, in a classification scheme used to classify real data from Capucho's Hospital pacients.

2. FEATURES EXTRACTION

This method relies on color textural features extraction process based in textural analysis. These features are estimated over the second order statistical representation of the cooccurrence matrix calculated from the wavelet transform of the colour image. The statistical descriptors calculated for each cooccurrence matrix give textural information about the properties of the decomposed subimages. These descriptors contain second order colour level information, which are mostly related to the human perception and discrimination of textures. For course textures these matrices tend to have higher values near the main diagonal whereas for a fine texture the values are scattered. The cooccurrence matrices encode the wavelet level (for each colour) spatial dependence based on the estimation of the second order joint-conditional probability density function $f(i,j,d,\theta)$, which is computed by counting all pairs of pixels at distance d having wavelet coefficients of colour levels i and j at a given direction θ. The angular displacement used is the set $\{0, \pi/4, \pi/2, 3\pi/4\}$. Let t be a translation, calculated from θ and d, then a co-occurrence matrix C_t of a image is defined for every grey-level (a, b) and the mathematical definition is given by:

$$C_t(a,b) = card\{(s,s+t) \in \mathbb{R}^2 \mid A[s]=a, A[s+t]=b\} \quad (1)$$

It is considered only 4 statistical measures among the 14 originally proposed by Haralick [5]. They are angular second moment (F1), which gives a measure of homogeneity, correlation (F2), which is a measure of directional linearity, inverse difference moment (F3) and entropy (F4) defined respectively as

$$F1 = \sum_{i=1}^{N}\sum_{j=1}^{N} p(i,j)^2 \quad (2)$$

$$F2 = \frac{\sum_{i=1}^{N}\sum_{j=1}^{N} (i.j)\,p(i,j) - \mu_x \mu_y}{\sigma_x \sigma_y} \quad (3)$$

Where

$$\mu_x = \sum_{i=1}^{N} i \sum_{j=1}^{N} p(i,j) \qquad \mu_y = \sum_{j=1}^{N} j \sum_{i=1}^{N} p(i,j) \quad (3a)$$

$$\sigma_x = \sum_{i=1}^{N} (i-\mu_x)^2 \sum_{j=1}^{N} p(i,j) \quad (3b)$$

$$\sigma_y = \sum_{j=1}^{N} (j-\mu_y)^2 \sum_{i=1}^{N} p(i,j) \quad (3c)$$

$$F3 = \sum_{i=1}^{N}\sum_{\substack{j=1 \\ i\neq j-1}}^{N} \frac{1}{1+(i-j)^2} p(i,j) \quad (4)$$

$$F4 = \sum_{i=1}^{N}\sum_{\substack{j=1 \\ p(i,j)\neq 0}}^{N} p(i,j) \log_2 p(i,j) \quad (5)$$

where $p(i,j)$ is the *ijth* entry of normalized cooccurrence matrix, N the number of levels of the wavelet and μ_x, μ_y, σ_x, σ_y are the means and standard deviations of the marginal probability $p_x(i)$ obtained by summing up the rows of the matrix p(i,j).

The proposed algorithm can be decomposed in the following categories:

A- Wavelet Domain Coefficients

Color transformations of the original image I result in three decomposed color channels, in the RGB color space:

$$I^i, \qquad i = 1,2,3. \quad (6)$$

where i stands for the color channel.

A four level discrete wavelet frame transformation is applied to each color channel (I^i). This transformation results in a new representation of the original image by a low resolution image and the detail images. The wavelet bases used were the Daubechies bases, as they were the default in the algorithm used. The performance of the proposed algorithm with different wavelet bases was not, at this point, tested. Nevertheless, it is an aspect to evaluate in the near future. Therefore the new representation is defined as:

$$I^i = \{D_n, D_l^i\} \quad i = 1,2,3 \quad l = 1,\ldots,9 \quad (7)$$

where l stands for the wavelet band and n is the decomposition level.

Since the textural information is better presented in the middle wavelet detailed channels, the second level detailed coefficients were considered. Thus, the image representation consists of the detail images produced from (7) for the values l=4, 5, 6, as shown in figure 1. This results in a set of 9 subimages, where each color channel originates 3 subimages:

$$\{D_l^i\} \quad i = 1,2,3 \qquad l = 4,5,6 \quad (8)$$

B- Coocurrence matrix and statistical descriptors

For the extraction of the second order statistical textural information, cooccurrence matrices were calculated for the nine different subimages. These matrices capture spatial interrelations among the intensities within the wavelet decomposition level, determining how often different combinations of pixel brightness values occur in an image. The cooccurrence matrices are estimated in four different directions resulting to 36 (3x3x4) matrices:

$$C_\alpha\left(D_l^i\right) \quad i = 1,2,3 \quad l = 4,5,6$$
$$\alpha = 0, \frac{\pi}{4}, \frac{\pi}{2}, 3\frac{\pi}{4}. \qquad (9)$$

Where i stands for the color channel, l for the wavelet band and α for the direction in the coocurrence computation.

Four statistical measures given by equations (2), (3), (4) and (5) are estimated for each matrix C resulting in 144 wavelet features.

$$F_m\left(C_\alpha\left(D_l^i\right)\right) \quad i = 1,2,3 \quad l = 4,5,6$$
$$\alpha = 0, \frac{\pi}{4}, \frac{\pi}{2}, 3\frac{\pi}{4} \qquad m = 1,2,3,4 \qquad (10)$$

where m stands for statistical measure.

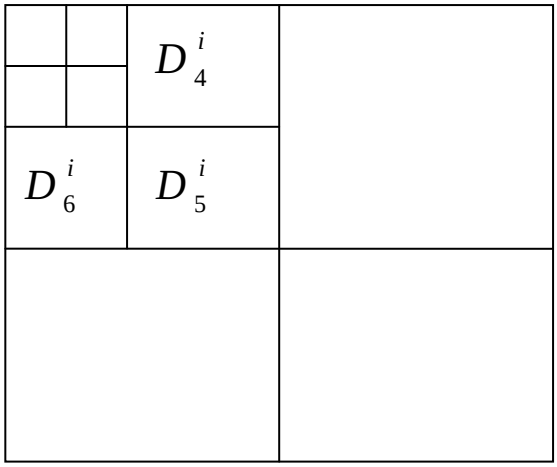

Figure 1. Four level wavelet decomposition scheme of the original image for color channel i.

C- Color Wavelet Covariance

Since each feature represents a different property of the examined region, the covariance among different statistical values between the color channels of the examined region, will statistically describe the textural behavior of the subimages, which will be very usefull information in our analysis. It is then expected that similar textures will have close statistical distributions and consequently they should have similar features. This similarity between features can be described by measuring the variance in pairs of them. Additionally the covariance between two features measures their tendency to vary together. The texture covariance has been proposed in the literature [6] as a measure used directly on image intensities or among the color intensities of the examined region. CWC coefficients can be computed based on the covariance of the same statistical descriptor, between different color channels, at different scales. This covariance can be computed as:

$$\gamma_{F_m,F_m} = \sum_\alpha F_m\left(C_\alpha\left(D_{l+3}^i\right)\right) - E\left(F_m\left(C_\alpha\left(D_{l+3}^i\right)\right)\right) \cdot F_m\left(C_\alpha\left(D_l^j\right)\right) - E\left(F_m\left(C_\alpha\left(D_l^j\right)\right)\right) \qquad (11)$$

The multiband color wavelet covariance features are then defined as

$$CWC_m^l(i,j) = \begin{cases} \gamma_{F_m,F_m}, & i < j \\ \sigma^2_{F_m,F_m} & i = j \end{cases} \qquad (12)$$

Which results in a set of 72 components per frame. These components constitute the input of the feed-forward neural network.

3. MULTILAYER PERCEPTRON

The classification scheme described in this paper used a standard MLP network, with 72 input neurons, 2 output neurons (normal and tumour) and a variable number of neurons in the hidden layer. The performance of the network was tested for different configurations in the hidden layer, in the attempt of defining the most suitable number of neurons.

The training algorithm was the well known back propagation learning process, in which the values of each connection are adjusted in order to reduce the value of the error function. The two output neurons were used to classify the data into 2 classes, namely normal and tumour.

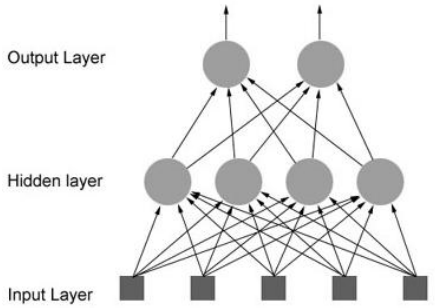

Figure 2. Example of a neural network with one hidden layer

4. EXPERIMENTAL RESULTS

The experimental set consisted of 6 full endoscopic exams taken at the Capucho's Hospital in Lisbon by Doctor Jaime Ramos. The system was trained in data that does not belong to the examined patients. The training set was constructed with images from normal segments of capsule endoscopic videos, some of them taken from exams with pathological cases. The tumour images were taken from capsule endoscopy exams with this pathology. The final training dataset was composed by 100 normal images and 73 tumour images. In figures 3 and 4 are examples of normal tissue frame and a tumour tissue frame, respectively.

Figure 3. Example of a normal intestinal tissue frame

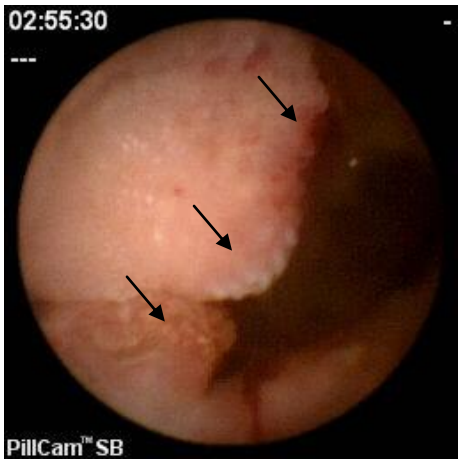

Figure 4. Example of a intestinal tumour tissue frame

Instead of measuring the rate of successful recognized patterns, more reliable measures for the evaluation of the classification performance can be achieved by using the sensitivity (true positive rate) and the specificity (100-false positive rate) measures. Therefore, sensitivity is the accuracy among positive patterns, while specificity is the accuracy among negative patterns [7]. These two measures can be calculated as:

$$Sensitivit\ \ y = \frac{d}{c + d}.100\ \ (\%) \qquad (13)$$

$$Specificit\ \ y = \left(100 - \frac{b}{a + b}.100\right)(\%) \qquad (14)$$

where a is the number of true negative patterns, b is the number of false positive patterns, c is the number of false negative patterns and d is the number of true positive patterns.

The classification performance is high when both Sensitivity and Specificity are high, in a way that their tradeoff favors true positive or false positive rate depending on the application.

A 3.2 GHz Pentium Dual Core processor-based with 256 MB of RAM was used with Matlab to run the developed algorithm. The average time processing per frame is about 2:15 minutes, which is fairly too much, but drops considerably without loss of performance if the size of the cooccurrence matrices is set to 64 X 64 instead of using 256 X 256 (full range). In any regular image, each pixel can assume 256 values (2^8 levels), that can be easily converted to 64 values (2^6 levels) with a simple multiplication, as we can see in (15). Note that with this operation, there is only a change in the number of the gradation levels for each position in the matrix. In this case the average time processing per frame is about 15 seconds.

$$V_{64}(i,j) = round\left(V_{256}(i,j) \times \frac{64-1}{256-1}\right) \qquad (15)$$

where $V_{64}(i,j)$ stands for the new value in 2^6 levels and $V_{64}(i,j)$ stands for the original value in 2^8 levels.

A mask was applied to the wavelet subimages in order to avoid computing cooccurrences in the image corners where no image information exists. The algorithm for computing cooccurrence matrices is implemented in such a way that only one passage on the matrix allows computing cooccurrences in the 4 required directions.

The first barrier in this work was the conversion of the wavelet coefficients to 256 or 64 gray levels, because the most of them are very close to zero, with some comparatively very large negative and positive values. The direct conversion into 256 levels, where the minimum wavelet coefficient is 0 and the maximum wavelet coefficient is 255, doesn't satisfy the performance criteria expected to this algorithm, because the most of the information is included in a few, very close, gray levels, implying a very sparsed coocurrence matrix, with only a few non-zero values. This can be solved with the proper dispersion of these very close wavelet coefficients to a more suitable interval. The extreme values are not appreciated, without loss of information. We can assume that the wavelet

coefficients follow a normal distribution, with zero mean, so if we shift the mean and adjust the variance, we can fit these in the interval [0,1], and then proceed with the conversion. As it is well known a random variable *y* with a given variance can be synthesized from a random variable *x* according to (16)

$$y = k \times x \quad \rightarrow \sigma_y^2 = k^2 \times \sigma_x^2 \qquad (16)$$

The mean (μ_m) and variance (σ_m) of each wavelet coefficient matrix were calculated as the mean of the mean and variance of each row of the matrix. The new value for each wavelet coefficient was then calculated as:

$$W_f(i,j) = \sqrt{\frac{\sigma}{\sigma_m}} \times (W_i(i,j) - \mu_m) + 0.5 \qquad (17)$$

where $W_f(i,j)$ is the final value of the wavelet coefficient in the (i,j) position, $W_i(i,j)$ is the initial value of the wavelet coefficient in the (i,j) position and σ is the normalized variance of the wavelet coefficients after the dispersion process.

The effects of the variation of the variance in the wavelet coefficients and the number of neurons in the hidden layer of the multilayer perceptron network were tested, searching the optimal conditions for the performance of the algorithm. Table 1 show the results for various variances of the wavelet coefficients assumed as normally distributed and also for different number of neurons in the intermediate layer. In table 1 *Nn* stands for number of neurons. The numbers between parentheses are the values for the normalized variance of the wavelet coefficients, according to (17).

Table 1. Experimental Results

Nn.	Se (0.3)	Sp	Se (0.5)	Sp	Se (1.0)	Sp
20	93%	80.5%	98.6%	79%	97.3%	71%
25	98.6%	81%	98.6%	80%	97.3%	86%
30	95.6%	83%	98.6%	81%	97.3%	87%
35	94.5%	86%	98.6%	82%	97.3%	84%

The analysis of the results in Table1, the best sensitivity (98.6%) is achieved normalizing the variance of the wavelet coefficients to 0.5, for all of the network's sizes tested, while the best specificity (87%) is achieved is achieved normalizing the variance of the wavelet coefficients to 1.0, in a MLP network with 30 neurons. The best overall results, considering the characteristic trade-off between sensitivity and specificity, is achieved in the the MLP network with 30 neurons. There isn't any strong evidence that the further increase in the number of neurons in the network would lead to better results in the classification process. So, for this specific application, it can be concluded that the optimal number of neurons in the hidden layer is in the range [25,30], since the addition of more neurons consumes more computational resources, and simultaneously, doesn't improve the performance of the classification process. However the coocurrence matrix still had many zeros, with the most of the non-zero values clustered in the center of the matrix, so we can still increase the value of the normalized variance.

5. DISCUSSION AND FUTURE WORK

The results of this paper shows that colour textural information can be adequate to classify images from endoscopic capsule. This colour textural information can be obtained from the covariances of the second-order statistical measures calculated over the wavelet frame transformation of different colour bands. The information present in the covariance of the selected features was successfully used in the classification of the images by a multilayer percetron network.

However the performance of this method can be improved with some minor modifications. For instance, the colour space used is RGB, so it will be tested soon the proposed algorithm in other colour spaces as HSV, CIE-Lab or K-L. The enlargement of the dataset for the training is another important task to improve the classification process, since there is a wide range for normal tissue frames. The use of different wavelet bases will also be considered.

At the same time, they will be considered different approaches to the features extraction, namely a multiband algorithm based in the method proposed in this paper. The utilization of different classification systems, as Radial Basis Functions neural networks and Support Vector Machine classifiers, will also be subject of investigation.

In the future, our main goal is extend our work to other pathologies and develop a tool for automatic abnormalities detection to support the medical diagnosis.

6. REFERENCES

[1] Karkanis, S. A., Iakovidis, D. K., Maroulis, D. E., Karras, and Tzivras, M. (2003). Computer-Aided Tumor Detection in Endoscopic Video Using Color Wavelet Features. IEEE- Transactions on Information Technology in Biomedicine, vol. 7, Nº3, pp. 141-151.

[2] Iddan G., Meron, G., Glukhovsky, A., and Swain,P. (2000). Wireless capsule endoscopy. Nature, pp. 415-417.

[3] Abyoto, R. W., Wirdjosoedirdjo, S. J., and Watanable R. G. (1998). Unsupervised texture segmentation using multiresolution analysis for feature extraction. J Tokyo Univ. Inform. Sci., vol. 2, no. 9, pp 49-61.

[4] Haykin, S. (1994). Neural Networks. A comprehensive foundation. Mcmillan College Publishing Company New York.

[5] Haralick, R. M., (1979). Statistical and structural approaches to texture. Proc. IEEE, vol. 67 pp. 786-804.

[6] Chen, C. H., Pau, L. F., and Wang, P. S. P. (1998).The handbook of Pattern Recognition and Computer Vision, 2nd ed., Eds., World Scientific, Singapore, pp. 207-248.

[7] Swets, J. A., Dawes, R. M., and Monahan, J. (2000). Physiological science can improve diagnostic decisions. Psych. Sci Public Interest, vol. 1 pp.1-26