

# Classification of Endoscopic Capsule Images by Using Color Wavelet Features, Higher Order Statistics and Radial Basis Functions

C. S. Lima, D. Barbosa, J. Ramos, A. Tavares, L. Monteiro and L. Carvalho

**Abstract**— This paper presents a system to support medical diagnosis and detection of abnormal lesions by processing capsule 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 code 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. Third and fourth order moments are added to cope with distributions that tend to become non-Gaussian, especially in some pathological cases. The proposed approach is supported by a classifier based on radial basis functions procedure 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 95% specificity and 93% sensitivity.

## I. INTRODUCTION

ENDOSCOPIC capsule, which allows minimally invasive imaging procedures, is ingested by the patient and films the whole gastrointestinal tract, reaching places where the conventional Endoscopy does not, such as the small bowel. Images are captured by a short-focal-length lens as the capsule is propelled by peristalsis through the gastrointestinal tract. Two frames per second are acquired for about 8 hours, resulting in more than 50.000 video frames per exam. Average small bowel transit time is about 90 minutes [1], 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. Note that having an expert physician analyzing, for a long period, a capsule endoscopic exam is also very costly, and, therefore, exists an important economic opportunity to develop a computer assisted diagnosis tool to this task.

Textural alterations of the small bowel mucosa surface can be used for the automatic detection of lesions. 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 [2]. Statistical color wavelet features have been encountered in texture analysis schemes for the discrimination of normal and abnormal (i. e. tumor) regions. The construction of the texture feature space follows the multiresolution approach on the wavelets extracted from the color domain. In this study involving color texture analysis, features were obtained based on cooccurrence matrices, obtained from the wavelet transform of different color spaces. Then second-order-statistics are computed between color channels for the same wavelet band. Third and fourth centered moments were added to deal with non-Gaussianity that tends to appear especially in some pathological cases.

Radial basis functions are the neural networks perhaps the most appropriated for statistical applications. The basic approach is viewing the design of a neural network as a curve-fitting approximation problem in a high dimensional space. Therefore learning is equivalent to finding a surface in a multidimensional space that provides a best fit to the training data. The idea behind the use of RBF's is to take advantage of the local approximations using exponentially decaying localized nonlinearities achieved by the Gaussian function, which increases the clustering power relatively to MLP's. This neural model can be advantageous over the global approximations to nonlinear input-output mappings provided by Multilayer Perceptrons (MLP's), especially when non-stationary processes need to be accurately modelled [3].

## II. FEATURES EXTRACTION

The proposed approach is based on the extraction of color textural features of the capsule endoscopic frames, in order to classify them. These features are estimated over the second order statistical representation of the wavelet transform of the color image. The cooccurrence matrices approach has been considered in this work for the description of a statistical model of the texture encoded within the decomposed subimages. It captures second order color level information, which is mostly related to the human perception and discrimination of textures. For coarse 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 color) 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 color levels  $i$  and  $j$ , at a given direction  $\theta$ . The angular

Manuscript received April 7, 2008.

C. S. Lima, D. Barbosa, A. Tavares, L. Monteiro and L. Carvalho are with the University of Minho, Industrial Electronics Department, Campus de Azurém 4800-058 Guimarães, Portugal phone: +351 253 604706; fax: +351 253 604709; e-mail: clima@dei.uminho.pt.

J. Ramos is with Capucho's Hospital, Gastroenterology department, Alameda Sto. António dos Capuchos, Lisbon, Portugal..

displacement used is the set  $\{0, \pi/4, \pi/2, 3\pi/4\}$ . Four statistical measures among the fourteen originally proposed by Haralick [4] were considered, to extract the texture information within the cooccurrence matrix. They are angular second moment, which gives a measure of homogeneity, correlation, which is a measure of directional linearity, inverse difference moment and entropy defined respectively as:

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

$$F2 = \frac{\sum_{i=1}^N \sum_{j=1}^N (i \cdot j) p(i, j) - \mu_x \mu_y}{\sigma_x \sigma_y}. \quad (2)$$

where

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

$$\sigma_x = \sum_{i=1}^N (i - \mu_x)^2 \sum_{j=1}^N p(i, j). \quad (4a)$$

$$\sigma_y = \sum_{j=1}^N (j - \mu_y)^2 \sum_{i=1}^N p(i, j). \quad (4b)$$

$$F3 = \sum_{i=1}^N \sum_{\substack{j=1 \\ i \neq j-1}}^N \frac{1}{1 + (i - j)} p(i, j). \quad (5)$$

$$F4 = \sum_{i=1}^N \sum_{\substack{j=1 \\ p(i, j) \neq 0}}^N p(i, j) \log_2 p(i, j). \quad (6)$$

where  $p(i, j)$  is the  $ij$ th entry of normalized cooccurrence matrix,  $N$  the number of levels of the wavelet and  $\mu_x, \mu_y, \sigma_x, \sigma_y$  are the means and standard deviations of the marginal probability  $p_x(i)$  obtained by summing up the rows of the matrix  $p(i, j)$ .

Color transformations of the original image  $I$  result in three decomposed color channels

$$I^i, \quad i = 1, 2, 3. \quad (7)$$

where  $i$  stands for color channel.

A three 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.

Therefore the new representation is defined as:

$$I^i = \{L_n^i, D_l^i\}, \quad i = 1, 2, 3 \quad l = 1, 2, 3, \dots \quad (8)$$

where  $n$  is the decomposition level. Since the textural information is better presented in the middle wavelet detailed channels, then second level detailed coefficients were considered. Thus, the image representation consists of the detail images produced from (8) for the values  $l=4, 5, 6$  as shown in figure 1. This results in a set of 9 subimages:

$$\{D_l^i\} \quad i = 1, 2, 3 \quad l = 4, 5, 6. \quad (9)$$

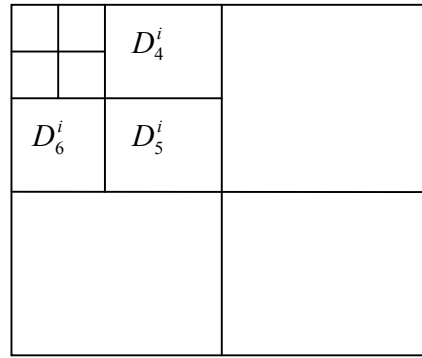


Fig. 1. Three level wavelet decomposition scheme of the original image for color channel  $i$ .

For the extraction of the second order statistical textural information cooccurrence matrices were used calculated over the nine different subimages. These matrices capture spatial interrelations among the intensities within the wavelet decomposition level. The cooccurrence matrices are estimated in four different directions resulting to 36 matrices

$$C_\alpha(D_l^i) \quad i = 1, 2, 3 \quad l = 4, 5, 6 \quad (10)$$

$$\alpha = 0, \frac{\pi}{4}, \frac{\pi}{2}, 3\frac{\pi}{4}.$$

where  $\alpha$  stands for cooccurrence direction.

Four statistical measures given by equations (1), (2), (5) and (6) are estimated for each matrix resulting in 144 wavelet features:

$$F_m(C_\alpha(D_l^i)) \quad i = 1, 2, 3 \quad l = 4, 5, 6 \quad (11)$$

$$\alpha = 0, \frac{\pi}{4}, \frac{\pi}{2}, 3\frac{\pi}{4} \quad m = 1, 2, 3, 4$$

where  $m$  stands for the four texture features of the cooccurrence matrix.

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, contain certainly valuable information. 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 [5] as a measure used directly on image intensities or among the color intensities of the examined region.

Finally the covariance of the same statistical measure between different color channels is computed and used as a textural measure

$$\gamma_{F_m, F_m} = \sum_{\alpha} \left[ F_m(C_{\alpha}(D_l^i)) - E\{F_m(C_{\alpha}(D_l^i))\} \right] \times \left[ F_m(C_{\alpha}(D_l^j)) - E\{F_m(C_{\alpha}(D_l^j))\} \right] \quad (12)$$

The color wavelet covariance is defined as

$$CWC_m^l(i, j) = \begin{cases} \gamma_{F_m, F_m}, & i < j \\ \sigma_{F_m, F_m}^2, & i = j \end{cases} \quad (13)$$

which results in a set of 72 components per frame.

Higher order statistics, namely third and fourth centered moments are then calculated for each texture feature, in order to model deviations to the Gaussian distribution. These deviations are accentuated in pathological cases for almost all features. Note that this shift to the normal Gaussian distribution does not affect preferentially any feature also. These are calculated as:

$$\gamma_m^3 = \frac{1}{N} \sum_{i,l} \left[ \sum_{\alpha} \left[ \left( F_m(C_{\alpha}(D_l^i)) - E\{F_m(C_{\alpha}(D_l^i))\} \right)^3 \right] \right] \quad (14)$$

$$\gamma_m^4 = \frac{1}{N} \sum_{i,l} \left[ \sum_{\alpha} \left[ \left( F_m(C_{\alpha}(D_l^i)) - E\{F_m(C_{\alpha}(D_l^i))\} \right)^4 \right] \right] \quad (15)$$

where N is given by the total combinations arisen from  $i$ , which stands for color channel, and  $l$ , which stands for the wavelet band. In this particular case,  $N=9$ .

Summing up the 8 higher order moments, each frame is characterized by a set of 80 components in the observation vector. These components constitute the input of the radial basis function.

### III. RADIAL BASIS FUNCTIONS

The classification scheme described in this paper used a standard radial basis function, which basic scheme is shown in figure 2, with 80 input units, 17 RBF units and 5 output neurons. One of the motivations to use radial basis functions is its ability to construct local approximations to non-linear I/O mapping, which is achieved by the Gaussian functions, and can model more accurately groups of features that have tendency to vary together.

The training algorithm was the well known hybrid learning process, where the centers were computed by clustering, the spreads of the Gaussians chosen by normalization and the Least Mean Square algorithm for computing the network weights. The 5 output neurons were used to classify the data into 5 classes namely bleeding, ulcer, polyps, normal and a type of defect not classifiable without an exhaustive analysis by the physician. The most part of the frames were later classified by the physician as a polyp with some medical particularities.

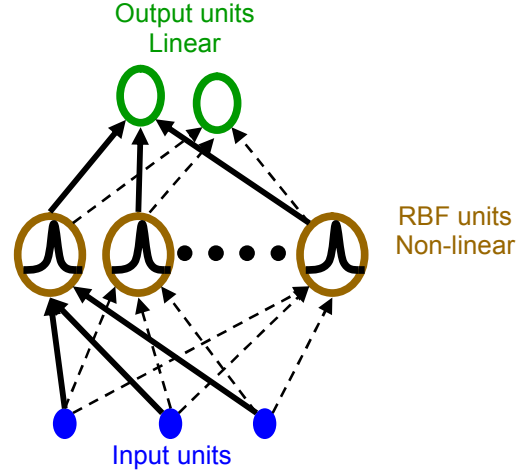


Fig. 2. Scheme of a radial basis function with three input and two output neurons.

### IV. EXPERIMENTAL RESULTS

The experimental set consisted of 6 full capsule 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 composed by more than 2.000 normal images, some of them taken from exams with pathological cases, 23 bleeding images, 54 polyp images, 123 Tumor images and 58 a priori undefined images, which includes Crohn and Jejunal lymphoma diseases. The total amount of images involved in the test set is about 85.000. Figure 3 shows some frames belonging to the training set. Note that, obviously, none of the images used in the training dataset were included in the test dataset and vice-versa.

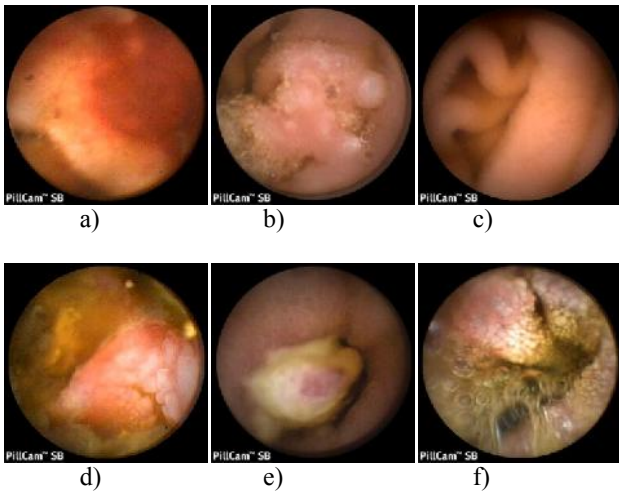


Fig. 3. Images of the small Bowel: a) Bleeding b) Polypoid mass c) normal d) Tumor e) Crohn's Disease f) Jejunal lymphoma

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 [6]. These two measures can be calculated as:

$$Sensitivity = \frac{d}{c+d} \cdot 100 (\%). \quad (14)$$

$$Specificity = \left( 100 - \frac{b}{a+b} \cdot 100 \right) (\%). \quad (15)$$

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 trade-off 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 but drops considerably without loss of performance if the size of the cooccurrence matrices is set to 64 X 64 instead of using almost 256 X 256 (full range). In this case the average time processing per frame is about 15 seconds. 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 a way that with only one passage for the matrix computes cooccurrences in all the 4 required directions.

The used color space was the RGB color space and the obtained pair (Sensitivity, Specificity) was for the described data set (93 ±0.4%, 95±0.1%). Table 1 resumes the most relevant results. The results are given in statistical terms, and each experiment was repeated four times for four different test set.

TABLE I  
CLASSIFICATION PERFORMANCE OF THE PROPOSED  
RBF CLASSIFICATION SCHEME

RBF Units	Sensitivity (%)	Specificity (%)	Accuracy (%)
20	88 ±0.7	93±0.25	91±0.5
25	92±0.4	92±0.15	92±0.3
30	94±0.32	97±0.1	96±0.2
35	96±0.24	96±0.05	96±0.2
40	95±0.15	96±0.15	96±0.2
Overall	93±0.4	95±0.1	94±0.25

## V. DISCUSSION AND FUTURE WORK

The results of this paper shows that color textural information can be adequate to classify images from endoscopic capsule. This color textural information can be obtained from the covariances of the second-order statistical measures calculated over the wavelet frame transformation of different color bands. Higher order statistics can be added, which help to model non-Gaussianity, which tends to appear especially for pathological cases. Radial basis functions can also be adequate for classification purposes using this kind of features. However the system needs improvements, especially concerning the amount of training data for pathological cases as well as the amount of testing data balanced with a greater number of pathological cases. In the near future, dimensionality studies will be made to reduce the number of features per frame. Different classifications approaches will also be subject of future investigation.

## REFERENCES

- [1] G. Idden, G. Meron, A. Glukhovskiy and P. Swain, "Wireless capsule endoscopy," *Nature*, pp. 415-417, 2000.
- [2] R. W. Abyoto, S. J. Wirdjosoedirdjo, and R. G. Watanable, "Unsupervised texture segmentation using multiresolution analysis for feature extraction," *J Tokyo Univ. Inform. Sci.*, vol. 2, no. 9, pp 49-61, (1998).
- [3] S. Haykin, *Neural Networks. A comprehensive foundation*. New York:Mcmillan College Publishing Company, 1994.
- [4] R. M. Haralick, "Statistical and structural approaches to texture," *Proc. IEEE*, vol. 67, pp. 786-804, 1979.
- [5] C. H. Chen, L. F. Pau, and P. S. P. Wang, *The Handbook of Pattern Recognition and Computer Vision*, Singapore: World Scientific, pp. 207-248, 1998.
- [6] J. A. Swets, R. M. Dawes, and J. Monahan, "Psychological science can improve diagnostic decisions," *Psychological. Science in the Public Interest*, vol. 1 pp.1-26, 2000.