Texture classification of images from Endoscopic Capsule by using MLP and SVM- A comparative approach

C. S. Lima[1], J. H. Correia[1], J. Ramos[2] and D. Barbosa[1]

[1] Industrial Electronics Department, University of Minho, Braga, Portugal
[2] Gastroenterology Department, Hospital dos Capuchos, Lisboa, Portugal

Abstract— **This article reports a comparative study of Multilayer Perceptrons (MLP) and Support Vector Machines (SVM) in the classification of endoscopic capsule images. Texture information is coded by second order statistics of color image levels extracted from co-occurrence matrices. The co-occurrence matrices are computed from images rich in texture information. These images are obtained by processing the original images in the wavelet domain in order to select the most important information concerning texture description. Texture descriptors calculated from co-occurrence matrices are then modeled by using third and forth order moments in order to cope with non-Gaussianity, which appears especially in some pathological cases. Several color spaces are used, namely the most simple RGB, the most related to the human perception HSV, and the one that best separates light and color information, which uses luminance and color differences, usually known as YCbCr.**

Keywords— **Capsule Endoscopy, texture analysis, Discrete Wavelet Transform, Multilayer Perceptrons and Support Vector Machines.**

I. Introduction

Wireless capsule endoscopy (CE) is a diagnostic procedure that allows the visualization of the whole Gastrointestinal (GI) tract, acquiring video frames, at a rate of two frames per second, while travels through the GI tract. Propelled exclusively by peristalsis, has a quite slow motion in the most part of its trajectory and therefore produces a huge amount of information, usually more than 50,000 images for a battery life of about eight hours [1]. The time required by an expert physician for the analysis of each exam is, in average, between 40 and 60 minutes [2], which increases significantly the total cost of the exam. Additionally this task is boring and therefore prone to subjective errors, since a very small amount of interesting images concerning diagnosis purposes are usually spread in thousands of unhelpful images. Therefore the development of an accurate computer assisted diagnosis tool for this task would be of greater value, not only economic but also regarding the human life, which can be at risk under misevaluation exams.

Texture analysis can be useful in the identification of abnormalities in the GI mucosa, since texture information is one of the primary clues analyzed by clinicians. Artificial methods for texture analysis can be of various natures, as statistical, geometrical, model based and signal processing based. Signal processing based methods rely on texture filtering for extracting features in the spatial or frequency domain. Gabor filter based approaches and wavelet based methods both emerged in the last decades are perhaps the most common signal processing based methods for texture analysis. New promising methods based on Color Wavelet Co-occurrence Coefficients to extract textural features in endoscopic images for tumor detection were proposed [3-6]. Color Wavelet Co-occurrence Coefficients are extracted from co-occurrence matrices computed in selected sub-bands which possess rich texture information [3]. Alternative approaches can use the same texture information, saving computational resources and diminishing the amount of coefficients extracted from capsule endoscopic images, and so time processing requirements, without loss of performance [4-6]. In any way, according to the current literature an artificial texture measure can be more appropriated for some type of lesion than for any other, even for lesions on the same organ. As an example, a comparison of four different texture features to discriminate gastric polyps in endoscopic video was reported in [7]. Comparative studies of different methods of texture encoding are out of the ambit of this paper.

Instead, we fixed the texture encoding method and try to compare the classification scheme performance, constrained to the use of Multilayer Perceptrons and Support Vector Machines. Both have several vantages and disadvantages but its performance can be dependent on the texture encoding process. Also interesting can be measure the effectiveness of the color space in the ambit of these texture coefficients and these non-linear classifiers. If we are trying to imitate the human characteristics, then HSV color space appears more appropriated than the simplest RGB, since it is more related to the human perception. However, the space YCbCr gives a better separation between light and color information and can be more robust to inaccuracies in the source light in CE frames and also in the angular and linear distance in which the light reaches the target.

II. Features extraction

The proposed method is based on a color textural features extraction process. Since the low-frequency components of the images do not contain major texture information, the most important bands in the wavelet transform are those in which are present medium and high frequency, texture encoding, information. To reduce the final number of features, a new image is synthesized from the selected wavelet coefficients, where the new image contains only the relevant texture information from the selected wavelet bands. The texture descriptors are calculated over the co-occurrence matrix calculated from the new image synthesized from the selected wavelet coefficients, for every color channels. These are statistical descriptors that contain second order color level information captured from the co-occurrence matrix, which are mostly related to the human perception and discrimination of textures.

Tone and structure of a texture are features that help to define more precisely textures [8]. Tone is based mostly in pixel intensity properties in the primitive, while structure is the spatial relationship between primitives. Repeated occurrence of some color level configuration can constitute an interesting texture description, since rapid variations with distance define fine textures while slowly variations define coarse textures. Co-occurrence matrices encode precisely this information and can be used to extract statistical descriptors for texture classification and pattern recognition systems based on textural descriptors. The statistical model is usually built by estimating 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 pixel intensity of color levels i and j at a given direction θ in the set $\{0, \pi/4, \pi/2, 3\pi/4\}$. Only the angular second moment (F1), correlation (F2), inverse difference moment (F3), and entropy (F4), representing the homogeneity, directional linearity, smoothness and randomness of the co-occurrence matrix [8] were used.

III. Image pre-processing

The image pre-processing stage synthesizes an image containing only the most relevant textural information from the source image. The most relevant texture information often appears in the middle frequency channels [9]. Texture is the discrimination information that differentiates normal from abnormal lesions, regarding colorectal diagnosis [10], [3], [11] and [12], hence it is likely to be extrapolated to small bowel diagnosis with similar characteristics.

The wavelet transform allows a spatial/frequency representation by decomposing the image in the corresponding scales. When the composition level decreases in the spatial domain it increases in the frequency domain providing zooming capabilities and local characterization of the image [13]. This spatial/frequency representation, which preserves both global and local information, seems to be adequate for texture characterization.

A new representation of the original image by a low resolution image and the detail images is obtained from the application of the wavelet transform to the three color channels and is defined as:

$$W^i = \left\{ L_n^i, D_l^i \right\}, \quad i = 1,2,3 \quad l = 1,...,6. \tag{1}$$

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

Since textural information is better presented in the middle wavelet detailed channels, then second level detailed coefficients would be considered. However, the relatively low image dimensions (256 X 256) limit the representation of the details, becoming the first level more adequate for texture representation [5]. Thus, the image representation consists of the detail images produced from (1) for the values l=1, 2, 3 as shown in figure 1. This procedure results in a set of 9 subimages, three for each channel color.

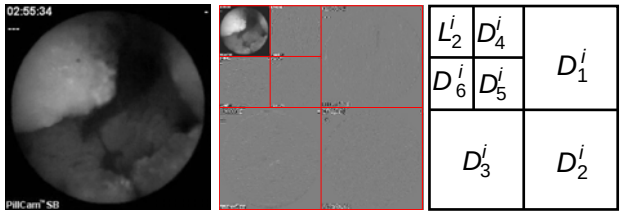

Fig. 1 – Two level DWT decomposition of a CE frame

For the extraction of the second order statistical textural information co-occurrence matrices would be used calculated over the nine different subimages. However, in order to diminish the dimension of the observation vectors the image to process can be synthesized from inverse wavelet transform with the coefficients of the large scales (lower frequencies) discarded. This procedure reduces the dimensionality of the observation vector by a factor of 3 since only three images need to be processed (color channels) instead of the nine obtained in the wavelet domain.

A new image, containing the most relevant texture information, is then synthesized from the selected wavelet bands, trough the inverse wavelet transform.

Co-occurrence matrices capture spatial interrelations among the intensities within the synthesized image and are estimated in four directions resulting in 12 matrices:

$$C_\alpha\left(N^i\right) \quad i = 1,2,3 \quad \alpha = 0, \frac{\pi}{4}, \frac{\pi}{2}, 3\frac{\pi}{4}. \tag{6}$$

where i stands for the color channel and α for the direction in the co-occurrence computation.

Four statistical measures are estimated for each matrix resulting in 48 texture descriptors:

$$F_m\left(C_\alpha\left(N^i\right)\right) \quad i=1,2,3,\ \alpha=0,\frac{\pi}{4},\frac{\pi}{2},3\frac{\pi}{4},\ m=1,2,3,4 \quad (7)$$

where m stands for statistical measure.

Since each feature represents a different property of the synthesized image, it is expected that similar textures will have close statistical distributions and, consequently, they should have similar features. This similarity between features can be statistically modeled in a tri-dimensional space, since features can be simultaneously observed in the three channel colors. Second order joint statistics can describe this similarity, however joint Gaussianity does not mean marginal Gaussianity although the reverse is true. Regarding F1, our investigations shown that normal images have approximately Gaussian distributions, while tumor images have essentially bimodal distributions.

Higher Order Statistics (HOS) is the most practical way to deal with non-Gaussianity, especially in applications requiring a significant amount of computational effort, such as the case of massive processing of capsule endoscopic images based on co-occurrence matrices. Alternative approaches can be based on Gaussian mixture modeling, which will be considered in future works.

IV. Modeling non-Gaussianity

Second order statistics is a well established theory that is completely adequate to represent random vectors. Nevertheless, it is limited by the assumptions of Gaussianity, linearity, stationarity, etc. HOS characterized by higher order moments are adequate to model non-Gaussian distributions under the assumption that all the moments are finite and so their knowledge is in practice equivalent to the knowledge of their probability function [14]. Third and fourth order moments have precisely meaning in separating Gaussian from other distributions. The third central moment gives a measure of assymmetricity of the probability density function around their mean (skewness), while the fourth central moment gives a measure of the peaky structure of the distribution when compared to the Gaussian. Higher than fourth order moments are used seldom in practice, hence not tried in the ambit of this paper. Therefore second, third, and fourth order moments, were used in the ambit of this paper. Second order moments or correlation of the same statistical measure between different color channels is computed as:

$$\phi_{F_m^i,F_m^j}=\sum_\alpha F_m\left(C_\alpha\left(I_s^i\right)\right)X\,F_m\left(C_\alpha\left(I_s^j\right)\right) \quad (9)$$

which results in the computation of 24 coefficients, six different correlations computed for each descriptor. Third and fourth order moments result in more 16 and 12 coefficients respectively. Summing up 28 higher order moments to the second order moments, each frame is characterized by a set of 52 components in the observation vector. These components constitute the input of the classifier.

V. Pattern recognition module

The classification engine uses only the features described in section IV and consists of state of the art Multilayer Perceptrons and Support Vector Machines Networks. However different color spaces are considered, namely RGB, HSV and YCbCr, which motivation was discussed in section I. No network optimizations were tried in order to increase the classifiers performance, since this procedure originates dependency on the existing database. The only concern was to be impartial regarding the comparative studies using equivalent features as the number of parameters to be estimated.

The MLP network has one hidden layer with 12 or 18 neurons in the hidden layer, as shown in tables 1, 2, and 3, the activation function of the first layer is tangent sigmoid while the one of the second layer is purely linear. The network is trained by using the standard back-propagation learning algorithm.

The experiments with SVM network were done by using software from the LIBSVM library, available at http://www.csie.ntu.edu.tw/~cjlin/libsvm/. The options (*1*), (*2*) and (*3*) of the SVM in tables 1, 2 and 3 correspond respectively to (*nu-SVC, linear*), (*nu-SVC, polynomial, d=2*) and (*nu-SVC, polynomial, d=3*) where *d* stands for the degree in the kernel function.

VI. Experimental Results

Experimental results were evaluated on real data from capsule endoscopic video segments of different patients' exams, taken at the Hospital dos Capuchos in Lisbon. The dataset consists of 400 normal frames, which were equally divided in two sets, for training and testing purposes and 200 abnormal frames, which were also equally divided in two sets. A 2.4 GHz Pentium Dual Core processor-based, with 1 GB of RAM, was used with MATLAB to run the proposed algorithm. The gradation of each color channel was reduced from 256 to 32 levels, which originates a drop in the processing time of 1 minute to one second per frame, without performance loss. Image reconstruction from only high frequency components requires the spreading of the

coefficients since the most part of them are of small values. This was achieved by a multiplication by a constant, which depends on the required dynamic range. This procedure guaranteeing that the textural information will be present in almost all the 32 gray levels. 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 and the specificity measures. From the presented results, it is clear that RGB is the color space with worst performance evaluation. However, regarding the other two color spaces no much difference was observed, although, in average, the YCbCr tends to have slightly superior results.

Table 1 comparison in the RGB color space

Classifier Parameters	MLP 12	SVM (1)	SVM (2)	SVM (3)	MLP 18
Specificity (%)	85.3	78.0	78.0	83.0	86.5
Sensitivity (%)	83.6	88.0	91.0	93.5	88.0

Table 2 comparison in the HSV color space

Classifier Parameters	MLP 12	SVM (1)	SVM (2)	SVM (3)	MLP 18
Specificity (%)	94.8	85.8	86.9	94.5	94.0
Sensitivity (%)	95.2	95.0	95.0	94.6	93.8

Table 3 comparison in the YCbCr color space

Classifier Parameters	MLP 12	SVM (1)	SVM (2)	SVM (3)	MLP 18
Specificity (%)	93.1	93.4	94.1	92.9	93.4
Sensitivity (%)	96.5	93.2	95.2	94.6	97.6

Regarding to the classifier itself and concentrating in tables 2 and 3 it is not clear which is the best. They perform quite similar and in average they are quite equivalent. However SVM have more potential, being although hard to obtain without optimizing for the existing database.

VII. Conclusion and Future Work

In this paper a comparative study of performance of MLP and SVM networks is carried out under the same database, features set and different color spaces. From this study it is clear that the RGB space is worse than the HSV and YCbCr color spaces. However no clear tendency were obtained from the comparison of the classifier itself, which implies that no strong conclusions about the preferential use of one classifier can be state, or in other words, it is not safe to say what classifier is the best, given this database and this feature set. Perhaps more accurate conclusions must be supported by a larger database, which constitutes our near future work. Nevertheless, it is important to stress out that the proposed algorithm to extract relevant information from capsule endoscopic video frames is flexible enough to achieve good classification performance in different settings in the chosen color space and in the classifier itself.

References

1. Idden G, Meron G, Glukhovsky A and Swain P (2000) Wireless capsule endoscopy. Nature 415-417
2. Pennazio M (2006) Capsule endoscopy: Where are we after 6 years of clinical use?, Digestive and Liver Disease 38:867–878
3. Karkanis S A, Iakovidis D K, Maroulis D E, Karras D A, Tzivras M (2003) Computer-aided tumor detection in endoscopic video using color wavelet features, IEEE Trans. Info. Tech. in Biomedicine, 7 (3) 141-152.
4. Lima C, Barbosa D et al. (2008) Classification of Endoscopic Capsule Images by Using Color Wavelet Features, Higher Order Statistics and Radial Basis Functions, Proceedings of the IEEE-EMBC2008, Vancouver, Canada, pp 1242-1245
5. Barbosa D, Ramos J, and Lima C (2008) Detection of Small Bowel Tumors in Capsule Endoscopy Frames Using Texture Analysis based on the Discrete Wavelet Transform, Proceedings of the IEEE-EMBC2008, Vancouver, Canada, pp 3012-3015
6. Barbosa D, Ramos J, and Lima C (2008) Wireless capsule endoscopic frame classification scheme based on higher order statistics of multiscale texture descriptors, Proceedings of the 4th European Medical and Biological Engineering Conference EMBEC'08, Antwerp, Belgium, pp 200-203
7. Iakovidis D K, Maroulis D E (2005) A comparative study of texture features for the discrimination of gastric polyps in endoscopic video, Proceedings of the 18th IEEE symposium on computer-based medical system.
8. Haralick R M, (1979) Statistical and structural approaches to texture, Proc. IEEE, 67, 786–804.
9. Van de Wouwer G , Scheunders P and Van Dyck, D (1999) Statistical texture characterization from discrete wavelet representations, IEEE Trans. Image Processing, 8, 592-598.
10. Maroulis, D., Iakovidis, D., Karkanis, A., and Karras, D., 2003, CoLD: a versatile detection system for colorectal lesions in endoscopy video frames, Computer Methods and Programs in Biomedicine, 70, 151-166.
11. Nagata S, Tanaka S, Haruma K, Yoshihara M, Summi K, Kajiyama G and Shimamoto F (2000) Pit pattern diagnosis of early colorectal carcinoma by magnifying colonoscopy: Clinical and histological implications, Int. J. Oncol., 16, 927-934.
12. Kudo S, Kashida H, Tamura T, Kogure E, Imai Y, Yamano H and Hart A R (2000) Colonoscopic diagnosis and management of nonpolypoid early colorectal cancer, World J. Surgery, 24, 1081-1090.
13. Mallat, S. 1998, *A wavelet tour of signal processing*, Academic Press.
14. Nandi, A. 1999, *Blind Estimation Using Higher-Order Statistics*, Kluwer.

Author: Carlos Lima
Institute: University of Minho
Street: B2.082 - DEI - Campus de Azurém
City: Guimarães
Country: Portugal
Email: clima@dei.uminho.pt