IDENTIFYING CANCER REGIONS IN VITAL-STAINED MAGNIFICATION ENDOSCOPY IMAGES USING ADAPTED COLOR HISTOGRAMS

A. Sousa[1], M. Dinis-Ribeiro[2], M. Areia[3], M. Coimbra[1]

[1]Instituto de Telecomunicações, Faculdade de Ciências da Universidade do Porto
[2]CINTESIS / Faculdade de Medicina do Porto and Instituto Português de Oncologia – Porto
[3]Instituto Português de Oncologia – Coimbra

ABSTRACT

In-body imaging technologies such as vital-stained magnification endoscopy pose novel image processing challenges to computer-assisted decision systems given their unique visual characteristics such as reduced color spaces and natural textures. In this paper we will show the potential of using adapted color features combined with local binary patterns, a texture descriptor that has exhibited good adaptation to natural images, for classifying gastric regions into three groups: normal, pre-cancer and cancer lesions. Results exhibit 91% accuracy, confirming that specific research for in-body imaging could be the key for future computer assisted decision systems for medicine.

Index Terms— Medical image processing, pattern recognition, local binary patterns, color features.

1. INTRODUCTION

Gastric cancer, the second most lethal cancer in the World and that with the highest mortality among digestive cancers in the Portuguese population, is one of the biggest current concerns in medicine. An early diagnosis of this disease is a vital factor in patient recovery. Nowadays, the medical community is very determined in searching methodologies for preventing the delayed detection of this disease that can only give us a reserved prognosis and, consequently, a reduced life expectancy.

Allied with emergent technologies in medicine, image processing could have a critical role in premature detection of this disease, helping doctors by providing them with another instrument to support their diagnosis. However, most previous research on statistical pattern recognition using color and texture features was based on conventional multimedia databases, which are clearly not good examples for in-body imaging scenarios [1], where we find reduced color spaces, absence of geometric textures, poor focusing and illumination conditions, etc. In this paper, we will show how adapted color histograms and uniform local binary patterns exhibit very interesting classification performances for a specific in-body imaging scenario, motivating further research in specific visual features for this imaging field.

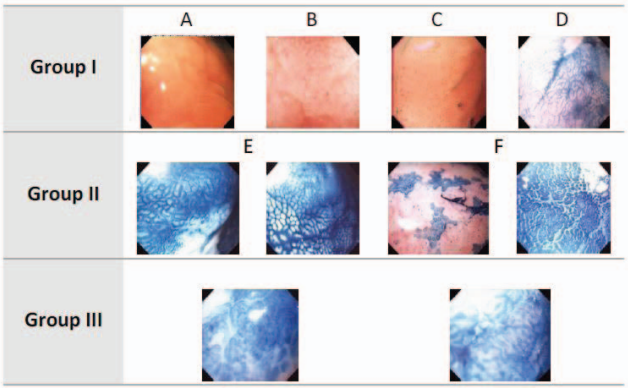

Figure 1 – Dinis-Ribeiro classification proposal [3] for vital-stained magnification endoscopy images: Group I – normal; Group II – Metaplasia (pre-cancer); Group III – Dysplasia (cancer).

For the study here presented, we have focused on the specific scenario of vital-stained magnification upper gastrointestinal endoscopy. Dinis-Ribeiro et al. [2] showed that it was possible to effectively diagnose certain gastric lesions (namely extension of intestinal metaplasia in the gastric mucosa) and adequately predict neoplasia occurrence using this technology. Dinis-Ribeiro developed an original classification for gastric mucosa, recently externally validated [3], dividing images into 3 groups based on their color and texture features as depicted in Fig.1, which were proven to be robust in intra-observer and inter-observer evaluation. In this study, we will mimic such human classification using statistical pattern recognition methodologies, focusing on adequate color and texture feature selection.

2. MATERIALS

All images used were obtained using an Olympus CV-180 endoscope at the Portuguese Institute of Oncology (IPO) Hospital in Porto, Portugal during routine clinical work. Optical characteristics of this endoscope include 140º field of view and four way angulation (180º up /down and 160º

right/left), and allows depths of field between 2 and 100 mm. The endoscopic videos used were recorded on tapes using a Digital Video (DV) recorder while performing real endoscopic examinations. Around 4 hours of video were analyzed and 176 images were initially selected given their clinical relevance. This was first determined by pre-selecting images that were annotated during the procedure by the clinician performing the exam, and later each image was individually selected for this study by an expert clinician. Images were saved as graphics files of type PNG (Portable Network Graphics) with a resolution of 518x481.

Two clinicians classified these images into three groups, following Dinis-Ribeiro's classification proposal [3], manually segmenting the image region that led them to this conclusion and labeling their choice with a confidence value (high or low confidence). Our resulting gold-standard not only uses regions where there was high-confidence annotations and agreement between the two specialists (135 images), but a second analysis was carefully performed for images where both doctors were confident and obtained different results. This typically showed that they selected different regions in an image that corresponded to different classifications (Fig. 2). The final number of high-confidence image regions used in this study was thus increased by 41 to a total of 176 (random coincidence with the initial number of images), distributed as 31.8% belonging to Group I, 54.5% to Group II, and 13.6% to Group III.

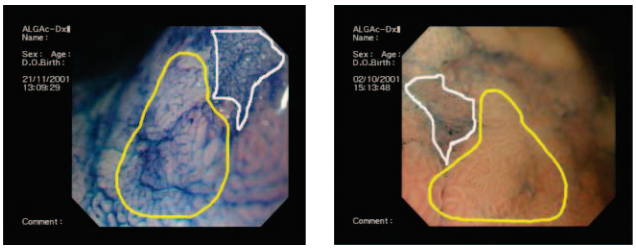

Figure 2 - Examples of situations where clinicians disagreed and both expressed high-confidence in their annotation.

3. METHODS

Classical statistical pattern recognition methodologies were used for producing image region classifiers. As such, we have extracted adequate color and texture features vectors for describing image regions, which were then concatenated, normalized and fed to statistical classifiers. A good example on this is Karkanis [4] work on tumor detection for the colon. The Weka (http://www.cs.waikato.ac.nz/ml/weka/) data mining software was used in the classifications experiments here presented. All results were obtained using 10-fold cross validation.

3.1. Adapted Color Histograms

We have shown in previous research that more adequate color histograms for in-body imaging scenarios can be obtained using simple split and merge operations on HSV (Hue-Saturation-Value) histogram bins [5][6] (Fig.3):

- *Merge* – Consists in merging the two adjacent bins that, when merged, produce the smallest resulting bin.
- *Split* – Consists in splitting the largest bin into four smaller ones, by dividing both the hue and saturation in half.

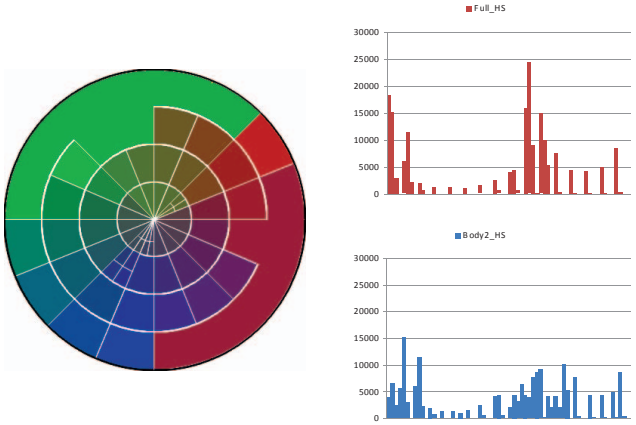

Figure 3 – Original and adapted color histogram (*Body2_HS*) for a typical in-body image of a vital-stained magnification endoscopy scenario.

Two of these new histograms will be used in this study (*Body1_HS*, *Body2_HS*), as well as two conventional ones (*Full_HSV*, *Full_HS*):

- *Full_HSV* – Obtained by quantizing the HSV space into 256 bins (16 for Hue, 4 for Saturation, 4 for Intensity).
- *Full_HS* – Obtained by collapsing the Intensity axis of the *Full_HSV* histogram, resulting in 64 bins.
- *Body1_HS* – Obtained by applying successive merge operations to the *Full_HS* histogram.
- *Body2 HS* – Obtained by applying successive iteration cycles of one split followed by three merge operations to the *Full_HS* histogram.

We refer the original papers for more details [5][6].

3.2. Local Binary Patterns (LBP)

This operator works with the eight neighbors of a pixel, using the value of this center pixel as a threshold. If a neighbor pixel has a higher gray value than the center pixel (or the same gray value) than a one is assigned to that pixel, else it gets a zero [7]. The LBP code for the center pixel is then produced by concatenating the eight ones or zeros as a binary code. The LBP operator has been extended to use neighborhoods of different sizes. In this case a circle is made with radius R from the center pixel. P sampling points on the edge of this circle are taken and compared with the value of the center pixel. To get the values of all sampling points in

the neighborhood for any radius and any number of pixels, (bilinear) interpolation is necessary. The notation (P,R) is used for defining the neighborhoods, as illustrated in Fig. 4.

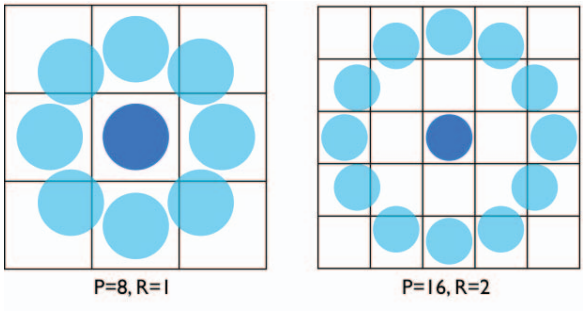

Figure 4 – Two Local Binary Pattern neighborhoods.

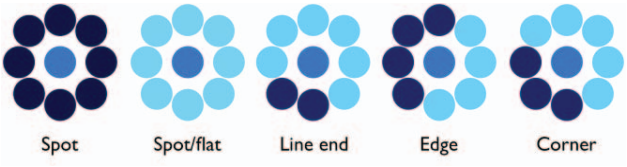

Figure 5 – Uniform Local Binary Pattern s

A LBP is called uniform if it contains at most two bitwise transitions from 0 to 1 or vice versa. The motivation for using uniform LBPs is not only higher computational efficiency, but also detecting only the important local textures, like spots, line ends, edges and corners as shown in Fig.5.

3.3. Statistical Classifiers

The Weka platform implements several statistical classifiers for pattern recognition. Different types of classifiers were chosen for this work, covering some of the most popular archetypes:

• **Decision Trees**: Uses trees that can be modeled according to a set of decision rules. Generally, Decision Tree Classification generates the output as a binary tree structure which contains rules to predict the target variable.
Weka classifier used: Logistic Model Trees – LMT.

• **Naive-Bayes**: Simple probabilistic classifier based on applying Bayes' theorem with strong (naive) independence assumptions. Generally this classifier estimates the probability of a class based on the number of instances that occur in that class;
Weka classifier used: NaiveBayes.

• **K-Nearest Neighbors**: Labels an unknown object O with the label of the majority of its k nearest neighbors. A neighbor is deemed nearest if it has the smallest distance, in a Euclidean sense, in feature space. For k = 1, this is the label of its closest neighbor in the learning set;
Weka classifier used: IB1.

• **Support Vector Machines**: Based on the concept of a decision hyperplane that maximizes the margin of separation between classes. A kernel function is used to transform the original feature space into a higher dimensional space, increasing the odds of finding hyperplanes with superior margins of separation.
Weka classifier used: Sequential Minimal Optimization (SMO).
For more details we refer to Herbrich [8] and Kuncheva [9].

4. RESULTS

After testing several combinations of color histograms and LBPs with different neighborhoods, we compiled the most significant results in Table 1.

	LMT	NaiveBayes	IB1	SMO
Full_HS+LBP(8,2)	88.1%	68.2%	84.1%	88.6%
Body2_HS+LBP(8,2)	**86.4%**	**77.3%**	**85.8%**	**90.9%**
Full_HSV+LBP(8,2)	86.9%	70.5%	80.7%	87.5%
Full_HS+LBP(8,1)	81.8%	64.7%	79.0%	83.5%
Body2_HS+LBP(8,1)	84.0%	73.8%	77.8%	83.0%
Full_HSV+LBP(8,1)	81.3%	75%	79.0%	84.7%
Full_HS+LBP(16,2)	84.1%	76.1%	76.1%	88.6%
Body2_HS+LBP(16,2)	87.5%	76.1%	77.3%	88.6%
Full_HSV+LBP(16,2)	84.7%	78.9%	77.2%	86.9%

Table 1 - Classification accuracy for the feature vectors proposed using different classifiers (LMT – Decision trees; Naïve Bayes; IB1 – K-Nearest Neighbors; SMO – Support vector machines).

Analyzing Table 1 we can observe that best results were obtained when concatenating the *Body2_HS* histogram with LBP(8,2). It is also interesting that similar results were achieved when using the *Full_HS* histogram instead of *Body2_HS*. Both histograms have the same size (123 bins) and share the same texture feature. We can now look deeper into the confusion matrix for the highest accuracy obtained in the proposed experiment – *Body2_HS + LBP(8,2)*:

		Automatic Classification		
		Group I	Group II	Group III
Manual Classification	Group I	**52**	3	1
	Group II	2	**94**	0
	Group III	**4**	**6**	**14**

Table 2 - Confusion matrix for the highest accuracy result - *Body2_HS+LBP(8,2)* - shown in Table 1.

The most immediate results are the system's ability to correctly handle Group I and Group II situations. There are very few false positives and only a couple of cases (2+3) where the automatic results fail to distinguish image patches between these two groups, obtaining an accuracy of nearly 97%. Results from Group III are not so promising given the large number of false negatives: 10 Group III cases were classified either as Group I or Group II. This is an important observation for future classification systems since the false

negatives in a computer-assisted diagnosis are very inconvenient, as the doctor is not warned when facing a potentially dangerous case. This is especially serious for the 4 image patches belonging to Group III (dysplastic lesions) that were classified as Group I (normal), which amounts to 17% 'serious' false positives. On the bright side, there is a strong probability that this problem is partially caused by the low amount of available Group III cases. As described before the used data set has only 24 Group III cases (13.6% of cases of the whole set) and it is well known that this has a strong negative impact on statistical classifiers and, consequently, their ability to correctly classify Group III.

It was decided to look a bit deeper into this problem. Support vector machines have the possibility to 'artificially' standardize the training data. In other words, the classifier can be trained while giving more weight to rare cases. Using this methodology the percentage of false negatives was reduced (Table 3), although the global accuracy of the system dropped slightly to 88.6%. This confirms our suspicions that even without altering the proposed methodology, the abstraction capability of the classifiers will surely improve given more Group III cases, thus creating a more robust automatic classification system.

		Automatic Classification		
		Group I	Group II	Group III
Manual Classification	Group I	**52**	1	4
	Group II	2	**89**	5
	Group III	**3**	**5**	**16**

Table 3 - Confusion matrix after using data standardization.

5. DISCUSSION

The classification accuracy results were better, for the majority of the cases, when using *Body2_HS* histogram as color features. For texture features, the results show that global accuracy was higher when using LBP(8,2). Other interesting point is that by significantly increasing the number of bins (for example using Full_HSV+LBP(16,2) histogram) the global accuracy does not effectively improve.

One must be careful with these observations since, as mentioned previously, it would be desirable to have a much larger dataset for robust conclusions. However, this contribution strengthens our claims that researching specific descriptors for in-body imaging might be the key for future computer assisted decision systems for medicine, making it possible to have automated classifying systems that can support doctors in a hospital environment. Comparing these results either with the percentage of data that doctors marked when they were not confident on their manual classification (7.6% for Miguel Areia and 8.9% for Mário Dinis-Ribeiro) or the percentage of cases they didn't agree (13.4 %) it is clear that results obtained are very promising.

The next non-trivial step for a computer-assisted system is the automatic segmentation of the images. This segmentation should divide the image into meaningful patches and then use them for classification. Traditional approaches such as using a simple grid are probably not good enough, paving the way for solutions involving mean-shift or normalized cuts methodologies [10].

Finally, all these patch classifications must be combined into a single classification for one image, which hopefully meets all the clinical requirements of doctors.

ACKNOWLEDGEMENTS

We would like to thank Instituto de Telecomunicações for its vital financial support of this study.

REFERENCES

[1] M. Coimbra, and J.P. Silva Cunha, "MPEG-7 visual descriptors – Contributions for automated feature extraction in capsule endoscopy", in *IEEE Trans. Circuits and Systems for Video Technology*, vol. 16/5, 2006, pp. 628-637.

[2] M. Dinis-Ribeiro et al., "Magnification Chromoendoscopy for the Diagnosis of Gastric Intestinal Metaplasia and Dysplasia", in *Gastrointestinal Endoscopy*, vol.57, pp.498-504, 2003.

[3] M. Areia et al., "External validation of a classification for methylene blue magnification chromoendoscopy in premalignant gastric lesions", in *Gastrointestinal Endoscopy*, Jan 2008.

[4] S. Karkanis, D. Iakovidis, D. Maroulis, D. Karras, and M. Tzivras, "Computer-aided tumor detection in endoscopic video using color wavelet features", in *IEEE Trans. on Inform. Tech. in Biomedicine*, vol. 7, pp.141-152, 2003.

[5] A.M. Sousa, "Analysis of Colour and Texture Features of Vital-Stained Magnification-Endoscopy images for Computer-assisted Diagnosis of Precancerous and Cancer Lesions", *FEUP, Master Thesis, 2008.*

[6] A.M. Sousa, M.D. Ribeiro, M. Areia, M. Correia, M. Coimbra, "Towards more adequate colour histograms for in-body images", *Int. Conf. IEEE EMBS,* Vancouver 2008.

[7] M. Ojala, M. Pietikainen, D. Harwood, "A comparative study of texture measures with classification based on feature distributions," *Pattern Recognition*, vol29, 1996.

[8] R. Herbrich, "Learning Kernel Classifiers: Theory and Algorithms", *MIT Press*, 2001.

[9] L. Kuncheva, "Combining Pattern Classifiers: Methods and Algorithms", *Wiley*, 2004

[10] F. Riaz, M. Dinis-Ribeiro, M.Coimbra, "Semantic relevance of current image segmentation algorithms", in *Proc. of IEEE WIAMIS 2009*, London, UK, May 2009.