Identifying Potentially Cancerous Tissues in Chromoendoscopy Images

Farhan Riaz, Fernando Vilarino, Mario Dinis Ribeiro, and Miguel Coimbra

Instituto de Telecomunicacoes, Universidade do Porto, Portugal
Computer Vision Center, Universitate Autnoma de Barcelona, Spain
Instituto Portugues Oncologia, Porto, Portugal
farhan.riaz@dcc.fc.up.pt

Abstract. The dynamics of image acquisition conditions for gastroenterology imaging scenarios pose novel challenges for automatic computer assisted decision systems. Such systems should have the ability to mimic the tissue characterization of the physicians. In this paper, our objective is to compare some feature extraction methods to classify a Chromoendoscopy image into two different classes: Normal and Potentially cancerous. Results show that LoG filters generally give best classification accuracy among the other feature extraction methods considered.

Keywords: Endoscopy, Computer Assisted Diagnosis, Gradient.

1 Introduction

Gastric cancer is a major cause of death worldwide. From a total of 57 million deaths worldwide in 2004, cancer accounts for 7.4 million (or 13%) of all deaths (World Health Organization, www.euro.who.int). Gastroenterology Imaging (GI) is today an essential tool for clinicians to detect cancer effectively. This is a rapidly evolving technological area with novel imaging devices such as Capsule, Narrow-Band Imaging (NBI) or High-Definition Endoscopy. In this paper, we are focused on classifying images from one of the most widely used gastrointestinal imaging modality: Chromoendoscopy (CH). It is based on using the full visible spectrum of light, accompanied by staining of the GI tissue with a dye, such as methylene blue to enhance the gastric mucosa in the images, thus helping in classifying the images as normal, pre-cancerous or cancerous. The clinical support of our work is provided by Dinis-Ribeiro classification proposal [1] which underlines the features which are supportive for classifying the images. Owing to the difficulties in the manual diagnosis systems and the training of clinicians for these novel imaging modalities, Computer Assisted Decision (CAD) systems are increasingly desirable to detect Gastrointestinal cancer effectively.

Our objective in this paper is to classify the CH images as being ‘normal’ (Group A) or ‘potentially cancerous’ (Group B). Our previous work shows the dominance of texture features in such images hinting at proficiency of methods based texture feature extraction for classifying such images (paper submitted in a journal for peer review). Many texture recognition methods are available in the

J. Vitrià, J.M. Sanches, and M. Hernández (Eds.): IbPRIA 2011, LNCS 6669, pp. 709–716, 2011.

literature, which can be mainly divided into four different categories: statistical methods, model based methods, structural methods and filter based methods [2]. Most techniques based on the former methods are more suitable for highly regular, semi-regular or micro-textures. These problems are mitigated by filter-based methods due their diversity provided by their ability to combine micro- and macro- texture features, thus giving a richer description of the images. In this paper, we focus on the extraction of interest points in the images followed by the use of Edge Histograms for finally classifying our images. The outline of the paper is as follows: We discuss the dataset (Section 2), followed by our methodology of feature extraction (Section 3). Afterwards, we describe our experimental setup (Section 4), followed by discussion and future work (Section 5).

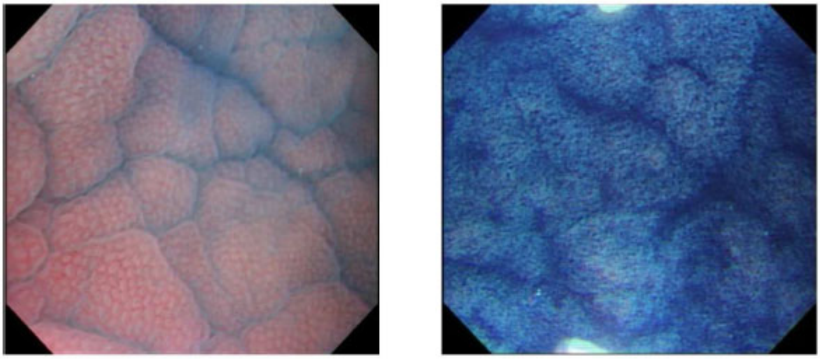
2 Materials

The CH images were obtained using an Olympus GIF-H180 endoscope at the Portuguese Institute of Oncology (IPO) Porto, Portugal during routine clinical work. Optical characteristics of this endoscope include 140 field of view and four way angulation (210 up, 90 down and 100 right/left). The endoscopic videos were recorded on tapes using a Digital Video (DV) recorder while performing real endoscopic examinations. Around 4 hours of video (360000 frames) 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 [1], 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. The final number of high-confidence image regions used in this study was thus increased by 41 to a total of 176. Finally, a careful analysis of images was carried out to remove the images belonging to the same patients, giving us a dataset of 142 images distributed as 31.6 % (45 images) belonging to Group I, 54.9 % (78 images) belonging to Group II and 13.3 % (19 images) to Group III.

3 Methods

3.1 Interest Point Detection

Conventionally, the spatial characteristics of an image are usually smooth over a particular neighborhood which makes a lot of data redundant. This enforces the



(a). Normal image (b). Cancerous image

Fig. 1. Typical normal and cancerous images. A distinct clear observation is the richness of texture for cancerous image, resulting from distortion of pattern.

need to discard the redundant data and take into account only a set of points which contain the salient features in the images. Our CH images show distinctive information between normal and potentially cancerous images, especially in terms of the distribution of interest points. For normal images this distribution is expected to be sparse in an area of annotation due to smooth texture whereas for other images we expect their dense distribution due to high texture. Also, we want to mimic the manual annotation of the physician by selecting some interest points in the images, which are representatives of important visual characteristics of the images. This motivates us to extract interest points in the images and process them to obtain salient features in images. We use three different methods for this purpose:

Harris operator: They are widely used for corner detection. For a 2-dimensional image, the Harris matrix is constructed which is a representative of the image derivatives. The magnitude of Eigen values of this matrix is a representative of potential corners in the image [3].
FAST operator: This operator considers the pixels under a Bresenham circle of a particular radius around an interest point [4] . The original detector classifies a point as a corner if there exists a set of contiguous pixels in the circle which are all brighter than the intensity of the candidate pixel plus a threshold or darker than the intensity of candidate pixel minus a threshold.
SIFT operator: This operator identifies scale-invariant features using a filtering approach through different stages [5]. In the first stage, key location in the scale space are selected by looking for location that are maxima or minima of different of Gaussian (DoG) function. The next steps consist of resampling the image and repeating the same procedure. The number of resampling stages conform the scale-space. Maxima and minima of this function are determined by comparing each pixel in a stage to its neighbors. It eventually detects key points which are translation, rotation and scale invariant.

Experiments show that the density of interest points in an annotated region could possibly be an indicator of richness of the texture which in turn can help us in characterizing the image into their respective classes.

3.2 Feature Extraction

The perception of salient points in the images is greatly affected by the amount and intensity of gradients. Even under varying lighting conditions, the most robust way to characterize an image could be the distribution and intensity of gradients in the images. It is one of the most fundamental tasks in computer vision. A variety of filters can be used this purpose, we used the following:

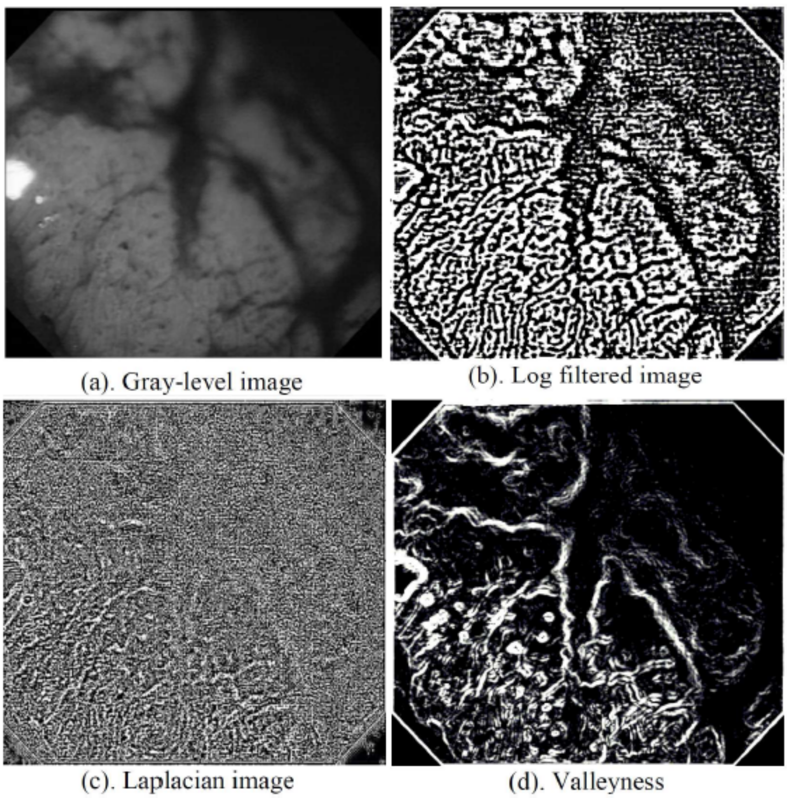


Fig. 2. Image after filtering using log, laplacian and valleyness filters

Laplacian filter: Laplacian is a differential operator given by divergence of the gradient of the function on the Euclidean space [6]. In image processing, this operator is used for various tasks such as blob detection or edge detection.
Log filter: LoG is an acronym for Laplacian of Gaussian. It is one of the most common blob detectors. It combines the use of two operators (Laplacian and

Gaussian) to improve blob detection [6]. For an input image, convolution with a Gaussian filter is done and then the Laplacian operator is applied. This operator usually gives strong positive responses for dark and strong negatives responses for bright blobs.

Valleyness filter: Ridges and valleys are useful geometric features for image analysis [7]. Researchers have characterized the mathematical model of the flow of water on the earth's surface based on the presence of ridges and valleys in the images. These models have lately been used for feature extraction and segmentation of the images by a well known method known as watershed segmentation. We use the valleyness operator to extract the ridge and valley features from the images.

3.3 Classification

The feature extraction is followed by a generation of histograms of underlying features. Those histograms are then used for classification of images into their respective classes. We classify these histograms using Support Vector Machines (linear kernel, one vs. one classification, sequential optimization) [8]. The objective of the classification task is to classify each image into two possible categories: Group A, which are images of patients which have no signs of cancer and Group B, which are images of patients who either are at initial stages of cancer or those who are suffering from cancer.

4 Experimental Setup

4.1 Parameter Setting

The first task while formulating the experimental setup is parameter setting of the methods. For this purpose, we divided the dataset into training and testing set. From a total of 142 images in the dataset, we selected 15 training images, which were used to tune the parameters of the methods used in this paper. Rest of the 127 images were used for testing.

Parameter setting for interest point detectors: The physicians provided us with ground truth data containing clinically relevant manual annotations (Region of Interest - ROI) and the respective labels for each image (manually classified). The objective of this step is to set the parameters of interest point detectors to ensure that all the interest points lie inside the ROI. We used our training set and their corresponding ROIs to tune the parameters such that all (or most) of the points lie inside the ROI. Experiments show that this is an important step as a change in the parameters of the methods can potentially result in a lot of points in clinically uninteresting regions.

Parameter setting for feature extraction methods: In the above feature extraction methods, only two of them (log filters and valleyness) take a parameter as input - the standard deviation of the Gaussian functions used. Experiments show that selection of a smaller value gives higher classification accuracy, we therefore used a small value (sigma = 0.5).

4.2 Data Normalization

To obtain meaningful and comparable results, we need to normalize the data. It involves eliminating the outliers, normalization of histogram and defining the centers of bins, which are to be used for feature extraction. For each of the feature extraction methodology, we sort the features ascendingly and discard 5% of the features (2.5% of the features at the lower extrema and 2.5% at the upper) to cater for the outliers. Mean and standard deviation from the rest of the data are extracted and the remaining data is normalized to zero mean and unit variance. Afterwards, 15 equally spaced bins between extrema of the rest of the data are created. Our experiments show that using a smaller number of bins depreciates the performance of the system, whereas using a higher number of bins does not have a change on the average output any further. For every novel image, the feature vectors are calculated and the outliers are discarded using the same procedure and the remaining data is normalized by the mean and standard deviation of the training set. Afterwards, the representative histogram of an image is generated using the bins which were calculated using the training set and the final feature vectors are obtained.

4.3 Classification Results

For analyzing the overall performance of the system, leave-one-out cross validation is used on 127 images which were saved for testing. This is because due to the lack of sufficient amount of data, we should do adequate training of the classifier to get consistent results. A visual illustration of our classification results is shown in Figure 3. The most notable observation in the graph is a higher discrimination power of gray level histograms for classifying CH images. This effectively means the classification is a function of image brightness. A deeper analysis (Fig. 4a) of histogram distribution using gray level values with the application of manual annotation mask shows that Group A images have an overall brighter intensity as compared with Group B images. We suspect that this happens because of very high texture generally for Group B images which tends to make the images to appear darker as compared with normal images. The need of feature detection is emphasized by the fact that when using full images, we have higher error rates for classification of both, Group A and Group B images (Fig. 4b). When feature detection is used, low error rates are obtained thus emphasizing the need to pre-process the image using one of the feature detection methods.

Relative conclusion does not change by altering feature detection methods, therefore results with only Harris corners is presented (Fig. 4b). Another important observation is that although gray level histogram performs reasonably well for CH images, LoG filter tend to perform better which can be observed by visualizing lower error rates for LoG filters for both Group A and Group B images.

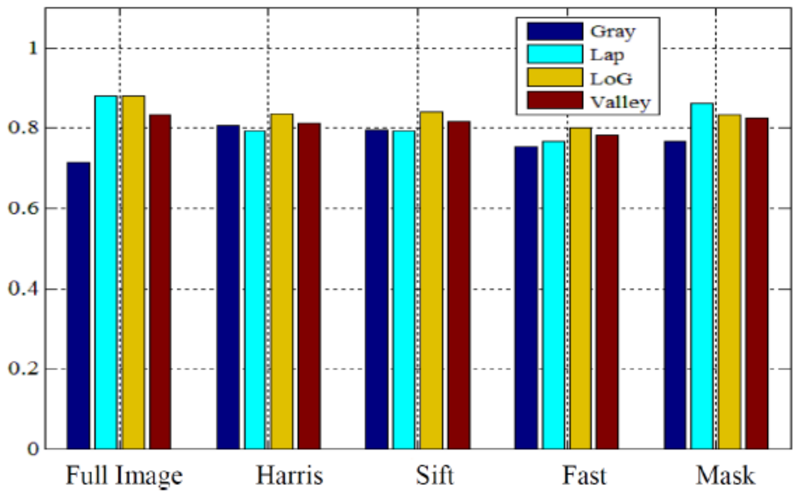


Fig. 3. Classification rates achieved for different feature extraction methods

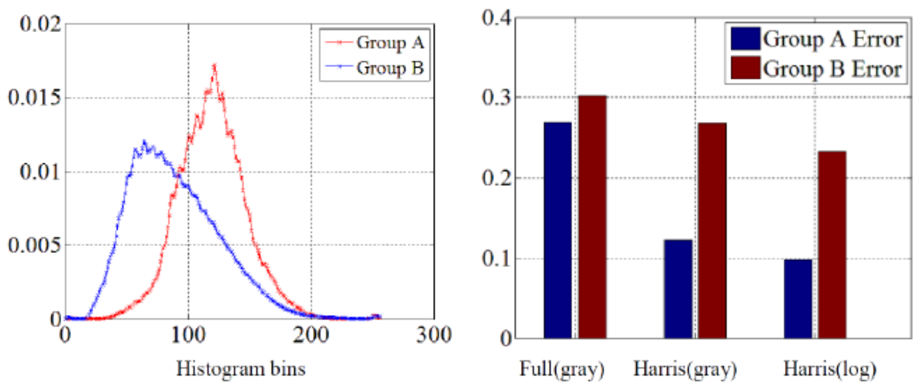


Fig. 4. Detailed analysis of classification performance. (a). Significance of gray level histograms (b). Advantages of feature detection.

5 Discussion

In this paper, we have used gradient based features to characterize Chromoendoscopy images into two different classes: Normal (Group A) and potentially cancerous (Group B). An important observation from our experiments was that the potential of gray level histograms of the images give unexpectedly good results. This is attributed to the richness of texture of images, which tends to darken the image for Group B patterns. Importance of feature detection is highlighted by an analysis of classification errors for Group A and Group B images.

When using full image for feature extraction, higher error rates are achieved however using any on the feature detection methods comparatively improved the results. This emphasizes the need to pre-process the images to select a few interest points, which improve the final classification of the images. Future work hints at studying the anatomy of the images in detail and trying to find ways to extract invariant features, which are expected to give better performance for our dataset. An interesting study would be the effects of reduction of simplification by classifying the images into 3 classes (normal, pre-cancerous and cancerous).

Acknowledgments. We would like to thank Portuguese Institute of Oncology, Porto for providing us clinical support for this work. This work was financially supported in parts by FCT (Fundao para Cincia e a Tecnologia) individual grant SFRH / BD / 45066 / 2008, project of Portuguese Government PTDC/EIA-CCO/109982/2009 and the projects of the Spanish Government COLON-QA (TIN2009-10435) and MIPRCV (CSD2007-00018). I am also very thankful to Jorge Bernal from Computer Vision Center, Barcelona for his technical support in carrying out this research. I would also like to thank the Computer Vision Center, Barcelona for providing me with logistic support for this research.

References

1. Ribeiro, M.D.: Clinical, Endoscopic and Laboratorial Assessment of Patients with Associated Lesions to Gastric Adenocarcinoma. Faculdade de Medicina, Universidade do Porto, PhD Thesis (2005)
2. Selvan, S., Ramakrishnan, S.: SVD-based modeling for texture classification using wavelets transformation. IEEE Trans. Image Process. 16(11), 2688–2696 (2007)
3. Harris, C., Stephens, M.J.: A combined corner and edge detector. In: Avley Vision Conference, pp. 147–152 (1988)
4. Rosten, E., et al.: A machine learning approach to corner detection. IEEE Trans. Pattern Analysis and Machine Intelligence 32(1), 105–119 (2009)
5. Lowe, D. G.: Object recognition from local scale-invariant features. In: International Conference on Computer Vision, Greece. Published by IEEE Signal Processing Society
6. Mikolajczyk, K., Schmid, C.: Scale and affine invariant interest point detectors. International Journal of Computer Vision, 63–86 (2004)
7. Lopez, A.M., Lumbreras, F., Serrat, J., Villanueva, J.: Evaluation of methods for Ridge and Valley detection. IEEE Trans. Pattern Analysis and Machine Intelligence 21(4), 327–335 (1999)
8. Herbrich, R.: Learning Kernel Classifiers: Theory and Algorithms. MIT Press, Cambridge (2001)