Barrett's Esophagus Identification Using Optimum-Path Forest

Luis A. Souza Jr., Luis C. S. Afonso
Federal University of São Carlos
Department of Computing
São Carlos, São Paulo, 13565-905
luis.souza@dc.ufscar.br, sugi.luis@dc.ufscar.br

Christoph Palm
Ostbayerische Technische Hochschule
Regensburg, Germany, 93053
christoph.palm@oth-regensburg.de

João P. Papa
São Paulo State University
Department of Computing
Bauru, São Paulo, 17033-360
papa@fc.unesp.br

***Abstract*—Computer-assisted analysis of endoscopic images can be helpful to the automatic diagnosis and classification of neoplastic lesions. Barrett's esophagus (BE) is a common type of reflux that is not straightforward to be detected by endoscopic surveillance, thus being way susceptible to erroneous diagnosis, which can cause cancer when not treated properly. In this work, we introduce the Optimum-Path Forest (OPF) classifier to the task of automatic identification of Barrett's esophagus, with promising results and outperforming the well-known Support Vector Machines (SVM) in the aforementioned context. We consider describing endoscopic images by means of feature extractors based on key point information, such as the Speeded up Robust Features (SURF) and Scale-Invariant Feature Transform (SIFT), for further designing a bag-of-visual-words that is used to feed both OPF and SVM classifiers. The best results were obtained by means of the OPF classifier for both feature extractors, with values lying on** 0.732 **(SURF) -** 0.735 **(SIFT) for sensitivity,** 0.782 **(SURF) -** 0.806 **(SIFT) for specificity, and** 0.738 **(SURF) -** 0.732 **(SIFT) for the accuracy.**

I. Introduction

The incidence of adenocarcinoma in patients affected by Barrett's esophagus (BE) has increased significantly in western populations, explained mainly by the obesity, a well-known risk factor [1]–[3]. As such, the expectation of this disease to rise up in the next years must be considered. Additionally, the bad prognosis for patients that have esophageal adenocarcinoma is related to its late diagnosis. However, the prognosis of the disease can be largely improved through early identification and surgical treatment, thus achieving very successful rates of handling the disease with 93% of the patients having complete remission after 10 years [2], [4], [5].

Developments in interventional therapies, such as photodynamic therapy, cryotherapy and radio-frequency ablation, showed promising results concerning the treatment of Barrett's esophagus. However, most of such methods are not able to describe properly the disease's level [6]–[8]. Once the identification of the BE is needed for the further evaluation of its region of interest by any computer or physician, some studies have focused on the definition of the area affected by the disease using the tissue properties in order to define a pattern to be followed.

Sonmmen et al. [9] presented a novel algorithm that computes local color and texture features based on the original and on the Gabor-filtered image for the automatic detection of early cancerous tissue in high definition endoscopic images. Appropriate filters based on spectral the characteristics of cancerous regions were designed for further feature extraction. A Support Vector Machine (SVM) classifier was used for classification purposes. Considering 7 patients evaluated, the experiments compared 32 annotations performed by the algorithm with the corresponding delineations provided by an gastroenterologist expert. With respect to 38 lesions indicated independenty by the gastroenterologist, the system detected correctly 36 of those lesions, with a recall as of 0.95 and a precision as of 0.75.

Kandemir et al. [10] performed a study for the diagnosis of Barrett's cancer from hematoxylin-eosin stained histopathological biopsy images using multiple instance learning (MIL) and SVM classifiers. In regard to the experiments, the tissue cores were partitioned into rectangular patches, and a feature vector was calculated based on a large set of cell-level and patch-level features for each patch. The tissue core was considered a bag (group of instances with a single group-level ground-truth label) and each patch an instance. The authors achieved an accuracy of around 82%, and 0.89 of AUC (area under the curve) using Bayesian logistic regression.

Rosenfeld et al. [11] aimed at studing how data mining can be applied to aid the diagnosis of patients with high-risk lesions and BE. As the patient's information is open to interpretation, the authors demonstrated that composite rules learned from multiple experts can be more accurate than that of one single expert. Such fact can be explained because even expert doctors may interpret endoscopy scans differently, which turns out to be way important to aggregate multiple opinions. Also, the authors demonstrated that decision trees can generate simple rules for dysplasia diagnosis. These rules can either be used to encapsulate the rules of the most accurate expert for training purposes or to help identifying diagnostic errors. The authors employed two decision models: one considering the expert decisions about dysplasia and non-dysplasia, and another without the expert decisions. The overall accuracies concerning the aforementioned models were around 79% (with the experts' decision) and 77% (without the experts' decision).

Souza Jr. et al. [12] conducted a study to test the feasibility of automatic adenocarcinoma classification in endoscopic images. A database composed of 100 expert-annotated endo-

scopic images were used for further feature extraction using Speeded up Robust Features (SURF) and classification with SVM using the leave-one patient out protocol for training and testing sets. The authors also considered two distinct protocols in the experimental section: using the full images and using only the adenocarcinoma regions annotated by the experts. The work obtained sensitivity as of 0.77 and specificity as of 0.82 considering the full images, and sensitivity as of 0.89 and specificity as of 0.95 considering the region-based approach.

Considering the growth of studies related to BE and adenocarcinoma automatic identification, we observed that a number of machine learning techniques have not yet been considered in this context to date. As an example, one can refer to the Optimum-Path Forest (OPF) classifier [13]–[15], which is a graph-based pattern recognition technique that models the problem of learning decision boundaries as a competition-reward process, where some key samples compete among themselves in order to conquer the remaining samples. The output of the competition-reward process is a graph partitioned into optimum-path trees, which stand for clusters of samples that share some content.

The Optimum-Path Forest classifier has been applied in a number of problems related to public health, including breast cancer [16]–[18] and laryngeal pathology detection [19], as well as human intestinal parasites identification [20], among others. However, as far as we are concerned, OPF has never been applied to the context of Barrett's esophagus identification so far, which turns out to be the main contribution of this paper. We showed OPF can outperform by far the well-known Support Vector Machines using two distinct feature extraction approaches and different kernels, thus showing to be a promising technique for BE automatic identification. The remainder of this paper is organized as follows. Section II and III present a brief background about Barrett's esophagus and the OPF classifier, respectively. Section IV discusses the methodology employed in this work, and Section V presents the experimental results. Finally, Section VI states conclusions and future works.

II. Barrett's Esophagus

The condition in which columnar cells replace the usual squamous cell in the mucosa of the esophagus is known as Barrett's esophagus. This condition is recognized as a complication of gastroesophageal reflux disease, and in some critical stages, it can progress and evolve into esophageal cancer [2], [21], [22].

Squamous cells similar to those of the skin or mouth compose the mucosa of the normal esophagus. The normal squamous mucosal surface appears whitish-pink in color, while the gastric mucosa appearance goes sharply from salmon pink to red, composed of columnar cells [2], [22]. A demarcation line, the squamocolumnar (SC) junction or "Z-line", represents the normal esophagogastric junction where the squamous mucosa of the esophagus and columnar mucosa of the stomach meet [21]. Barrett's mucosa may extend upward in a continuous pattern in which the entire circumference of the distal esophagus is covered by columnar mucosa. A distinction is drawn among patients with more than 3 cm of Barrett's esophagus ("long-segment Barrett's esophagus") and those with less than 3 cm of Barrett's esophagus ("short-segment Barrett's esophagus" [21]. The esophageal mucosa of patients with suspected BE is carefully examined for the presence of any visible lesions, which are then characterized by the Paris classification [23].

Multiple endoscopic image enhancement technologies, such as chromoendoscopy, electronic image enhancement (narrow band imaging, flexible spectral imaging color enhancement, i-Scan), confocal laser endomicroscopy, and optical coherence tomography, have been developed for BE diagnosis, which may enable endoscopists to conduct a more accurate endoscopic assessment of the dysplasia with *in vivo* characterization of esophageal histology. This ability could result in improvements regarding the detection of BE (screening), detection of dysplasia based on BE surveillance, characterizing abnormalities within BE (selecting lesions and delineating margins during endoscopic therapy), and detection of recurrent neoplasia in patients who have received endoscopic therapy (post-treatment surveillance) [22].

BE is often misdiagnosed during endoscopy, and this fact can be attributed to one out of two main reasons: (1) inability to differentiate columnar mucosa of the proximal stomach (cardia) from metaplastic epithelium in the distal esophagus, or (2) lack of goblet cells in biopsies obtained from columnar lined epithelium in the esophagus. Areas of dysplasia or early cancer in BE are sometimes not visible with standard white-light endoscopy. Hence, the Seattle biopsy protocol is recommended in which 4-quadrant biopsies are taken every 1 cm of the Barrett's mucosa. However, this biopsy protocol is prone to sampling error because only a small fraction of the entire BE mucosa is sampled (especially in patients with extensive disease) [22]. In addition, the biopsy protocol can be costly and time-consuming (because of the number of biopsies submitted to pathology). Therefore, endoscopists usually do not follow the recommended procedure for extensive biopsies, thus causing a significant rise in the risk of missed dysplasia or cancer [24].

III. Optimum-Path Forest

In this section, we explain the OPF working mechanism. Although we have different versions in the literature, we considered the first one proposed by papa et al. [13], [14], since it is parameterless and quite fast for both training and classification purposes.

Roughly speaking, the OPF classifier models the problem of pattern recognition as a graph partition in a given feature space. The nodes are represented by the feature vectors and the edges connect all pairs of them, defining a full connectedness graph. The partition of the graph is performed through a competition process among some key samples (prototypes), which offer optimum paths to the remaining nodes of the graph. Each prototype sample defines its own optimum-path

tree (OPT), and the collection of all OPTs defines an optimum-path forest, which gives the name to the classifier.

Let $\mathcal{Z} = \mathcal{Z}_1 \cup \mathcal{Z}_2$ be a dataset labeled with a function λ, in which $\mathcal{Z}_1$ and $\mathcal{Z}_2$ stand for the training and test sets, respectively, such that $\mathcal{Z}_1$ is used to train a given classifier and $\mathcal{Z}_2$ is used to assess its accuracy. Let $\mathcal{S} \subseteq \mathcal{Z}_1$ a set of prototype samples. Essentially, the OPF classifier creates a discrete optimal partition of the feature space such that any sample $\mathbf{s} \in \mathcal{Z}_2$ can be classified according to this partition.

The OPF algorithm may be used with any *smooth* path-cost function which can group samples with similar properties [25]. Papa et al. [13], [14] employed the path-cost function f_{max}, which is computed as follows:

$$\begin{aligned} f_{max}(\langle \mathbf{s} \rangle) &= \begin{cases} 0 & \text{if } \mathbf{s} \in \mathcal{S}, \\ +\infty & \text{otherwise} \end{cases} \\ f_{max}(\pi \cdot \langle \mathbf{s}, \mathbf{t} \rangle) &= \max\{f_{max}(\pi), d(\mathbf{s}, \mathbf{t})\}, \end{aligned} \quad (1)$$

in which $d(\mathbf{s}, \mathbf{t})$ denotes the distance between samples $\mathbf{s}$ and $\mathbf{t}$, and a path π is defined as a sequence of adjacent samples. As such, we have that $f_{max}(\pi)$ computes the maximum distance among adjacent samples in π, when π is not a trivial path.

The OPF algorithm assigns one optimum path $P^*(\mathbf{s})$ from $\mathcal{S}$ to every sample $\mathbf{s} \in \mathcal{Z}_1$, forming an optimum path forest P (a function with no cycles that assigns to each $s \in \mathcal{Z}_1 \backslash \mathcal{S}$ its predecessor $P(\mathbf{s})$ in $P^*(\mathbf{s})$ or a marker nil when $\mathbf{s} \in \mathcal{S}$. Let $R(\mathbf{s}) \in \mathcal{S}$ be the root of $P^*(\mathbf{s})$ that can be reached from $P(\mathbf{s})$. The OPF algorithm computes for each $\mathbf{s} \in \mathcal{Z}_1$, the cost $C(\mathbf{s})$ of $P^*(\mathbf{s})$, the label $L(\mathbf{s}) = \lambda(R(\mathbf{s}))$, and the predecessor $P(\mathbf{s})$.

A. Training

In the training phase, the OPF algorithm aims to find the set $\mathcal{S}^*$, that is the optimum set of prototypes, by minimizing the classification errors for every $\mathbf{s} \in \mathcal{Z}_1$ through the exploitation of the theoretical relation between minimum-spanning tree (MST) and optimum-path tree (OPT) for f_{max} [26]. The training essentially consists in finding $\mathcal{S}^*$ from $\mathcal{Z}_1$ and an OPF classifier rooted at $\mathcal{S}^*$.

By computing an MST, we obtain a connected acyclic graph whose nodes are all samples of $\mathcal{Z}_1$ and the arcs are undirected and weighted by the distances d between adjacent samples. The spanning tree is optimum in the sense that the sum of its arc weights is minimum as compared to any other spanning tree in the complete graph. In the MST, every pair of samples is connected by a single path which is optimum according to f_{max}. That is, the minimum-spanning tree contains one optimum-path tree for any selected root node.

The optimum prototypes are the closest elements of the MST with different labels in $\mathcal{Z}_1$ (i.e., elements that fall in the frontier of the classes). After finding prototypes, we run the competition process in order to build the optimum-path forest.

B. Testing

For any sample $\mathbf{t} \in \mathcal{Z}_2$, we consider all arcs connecting $\mathbf{t}$ with samples $\mathbf{s} \in \mathcal{Z}_1$, as though $\mathbf{t}$ were part of the training graph. Considering all possible paths from $\mathcal{S}^*$ to $\mathbf{t}$, we find the optimum path $P^*(\mathbf{t})$ from $\mathcal{S}^*$ and label $\mathbf{t}$ with the class $\lambda(R(\mathbf{t}))$ of its most strongly connected prototype $R(\mathbf{t}) \in \mathcal{S}^*$. This path can be identified incrementally by evaluating the optimum cost $C(\mathbf{t})$ as

$$C(\mathbf{t}) = \min\{\max\{C(\mathbf{s}), d(\mathbf{s}, \mathbf{t})\}\}, \ \forall \mathbf{s} \in \mathcal{Z}_1. \quad (2)$$

Let the node $\mathbf{s}^* \in \mathcal{Z}_1$ be the one that satisfies Equation 2 (i.e., the predecessor $P(\mathbf{t})$ in the optimum path $P^*(\mathbf{t})$). Given that $L(\mathbf{s}^*) = \lambda(R(\mathbf{t}))$, the classification simply assigns $L(\mathbf{s}^*)$ as the class of $\mathbf{t}$. An error occurs when $L(\mathbf{s}^*) \neq \lambda(\mathbf{t})$.

IV. Materials and Methods

This section describes the steps to develop a computerized system for the detection, delineation and characterization of endoscopic images obtained from individuals with clinically manifest tissue abnormalities in the esophagus. Based on a given set of endoscopic photographs (benchmark database), features are extracted (SURF- and SIFT-based) for the further classification using supervised learning (SVM and OPF). The next sections describe in more details the techniques used in this work.

A. Image Database

The set of images used as a benchmark database was provided at the MICCAI 2015 EndoVis Challenge [27], being composed of 100 endoscopic pictures of the lower esophagus captured from 39 individuals, 17 of them being diagnosed with early stage Barrett's, and 22 displaying signs of esophageal adenocarcinoma. From each proband, several endoscopic images were available, ranging from one to a maximum of eight. The database contained a total of 50 images displaying cancerous tissue areas, plus 50 images showing dysplasia without signs of cancer. Suspicious lesions observed in the cancerous images had been delineated individually by five endoscopy experts.

B. Speeded up Robust Features

The SURF algorithm [28] operates on integral images to detect dominant structures and their spatial orientation. To ensure scale and spatial invariance, SURF seeks for maxima of the determinant of the Hessian matrix, demarcating specific key points in the image, which are further explored in their local neighborhood. These sub-regions are evenly split into square patches while their wavelet responses in horizontal and vertical directions generate the elements of a high-dimensional feature vector of size 64.

C. Scale-Invariant Feature Transform

The SIFT algorithm [29] operates on image local regions aiming to calculate features that are invariant to image scaling and rotation, and partially invariant to change in illumination and 3D camera viewpoint. First, the algorithm seeks for the scale-space extrema detection evaluating all the image scales and regions using difference-of-Gaussian function to provide the potential image regions that are invariant to scale and

orientation. The key point localization is performed based on the candidate regions previously defined, and aims at measuring their stability. The final steps of SIFT algorithm are related to the definition of the orientation of each key point by means of the calculation of the gradient directions, and the key point descriptor calculation based on gradients measurement at the selected scale in the region around the key point [29]. These local descriptors are mapped into a global high-dimensional feature vector of size 128 that allows invariability for significant levels of local shape distortion and changes in the illumination.

D. Points of Interest

The computation of the points of interest (PoI) was performed with the SURF and SIFT algorithms using the OpenCV support package. The assessment of suitable IPs was based on two major approaches: one using SURF features, and other using SIFT features. Both approaches simulate "real life situations", thus lacking detailed information about tissue abnormalities. The analysis worked on the original full images. Two attributes were defined for the SVM and OPF training process: class '0' images (non-cancerous but with possible signs of early dysplasia), and class '1' images (exhibiting cancerous tissue regions). Figure 1 depicts some images of the database with delineations performed by five different physicians. Notice that some of the experts' demarcations exhibited substantial regional deviations, even for identical images.

E. Bag-of-Visual-Words

Bag-of-Visual-Words (BoVW) constitutes a robust representation approach in which each image is treated as a collection of regions. For this representation, the only information cared about is the appearance of each region. The objective when visual dictionaries are created is to learn, from a training set of examples, the generative model that selects the k most representative visual words for a given problem [30]. In order to fulfill that purpose, the original image's regions are mapped from the original space to a new representation encoded the visual words. Therefore, BoVW uses the PoIs from a set of reference images in order to generate a visual dictionary that is employed in the training and testing phases.

In order to select the most discriminative visual words, we used two approaches: k-means and a random selection. Once the visual dictionary is generated, a feature vector is created for each image by computing the frequency of each visual word in the image, and the feature vectors of all the images are now in the very same dimension [30]. We also considered dictionaries with three different sizes: 100, 500 and $1,000$, being such numbers chosen empirically. The pipeline adopted in this work is summarized in Figure 2.

F. Classification

In order to evaluate the robustness of the OPF classifier for automatic Barrett's esophagus identification, we also considered SVM with Radial Basis Function (SVM-RBF) and Linear (SVM-Linear) kernels with parameters optimized by cross-validation for comparison purposes.

In order to compose the dictionary of PoIs, we considered two techniques: (i) k-means and (ii) random. The former is a well-known unsupervised pattern recognition technique that aims at minimizing the distance between a given dataset sample and its nearest center, meanwhile the random approach simply chooses the visual words that will compose the dictionaries at random.

V. Experiments

In this section, we present the experimental results concerning the aforementioned methodology. In order to provide a more robust analysis of the results, we employed the Wilcoxon signed-rank statistical test [31] with confidence value as of 0.05 over 30 runs with randomly generated training and testing sets. Concerning the database partition, we used 70% of the data to compose the training set, and the remaining 30% for classification purposes.

The above training set percentage generated $14,411$ and $28,137$ PoIs with respect to the SURF and SIFT descriptors, respectively (on average). Additionally, the computational load concerning SURF extraction was around $13,77$ minutes, while SIFT algorithm took 5.95 minutes in an 8 GB memory and Intel core i5 - 2.30 GHz computer. We employed the OpenCV [32] implementation for both feature extraction techniques.

Table I presents the average results of sensitivity (SE), specificity (SP) and accuracy (AC) concerning the SURF features. We first considered the influence of each dictionary composition technique, i.e., k-means and random for BE identification. The values in bold stand for the most accurate ones (taking into account the statistical test) considering AC measure for different sizes of dictionaries. Clearly, one can observe that OPF obtained the best recognition rates in all situations, i.e., different dictionary generation technique and size.

The underlined values stand for the most accurate ones by analyzing the three distinct dictionary sizes and fixing the technique to generate it, i.e., if one considers k-means for dictionary generation technique, the best result was achieved by OPF with 500 words. With respect to k-means, the best result was obtained by OPF with 500 words as well. However, the statistical test did not point out that different dictionary sizes influence the results. Finally, the best overall result concerning different dictionary generation technique and size stands for OPF with k-means and 500 works with an accuracy as of 0.738 (highlighted with *). This is the best result one can summarize from Table I.

Table II presents the average results of sensitivity, specificity and accuracy concerning SIFT features. Once again, we first considered the influence of each dictionary composition technique for BE identification. One can observe that OPF obtained the best recognition rates in all situations once more, behaving similarly to SURF experiment.

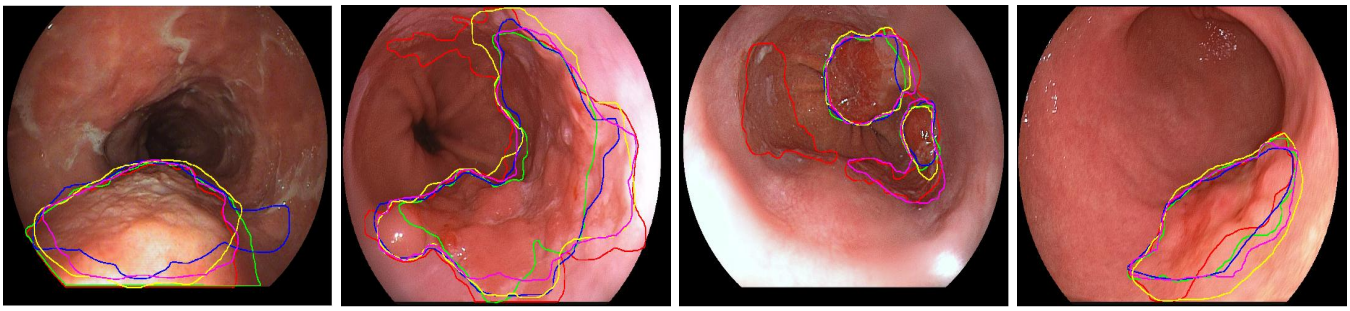

Fig. 1. Annotation by five different experts from four different cancerous images. Each colored delineation stands for a different expert.

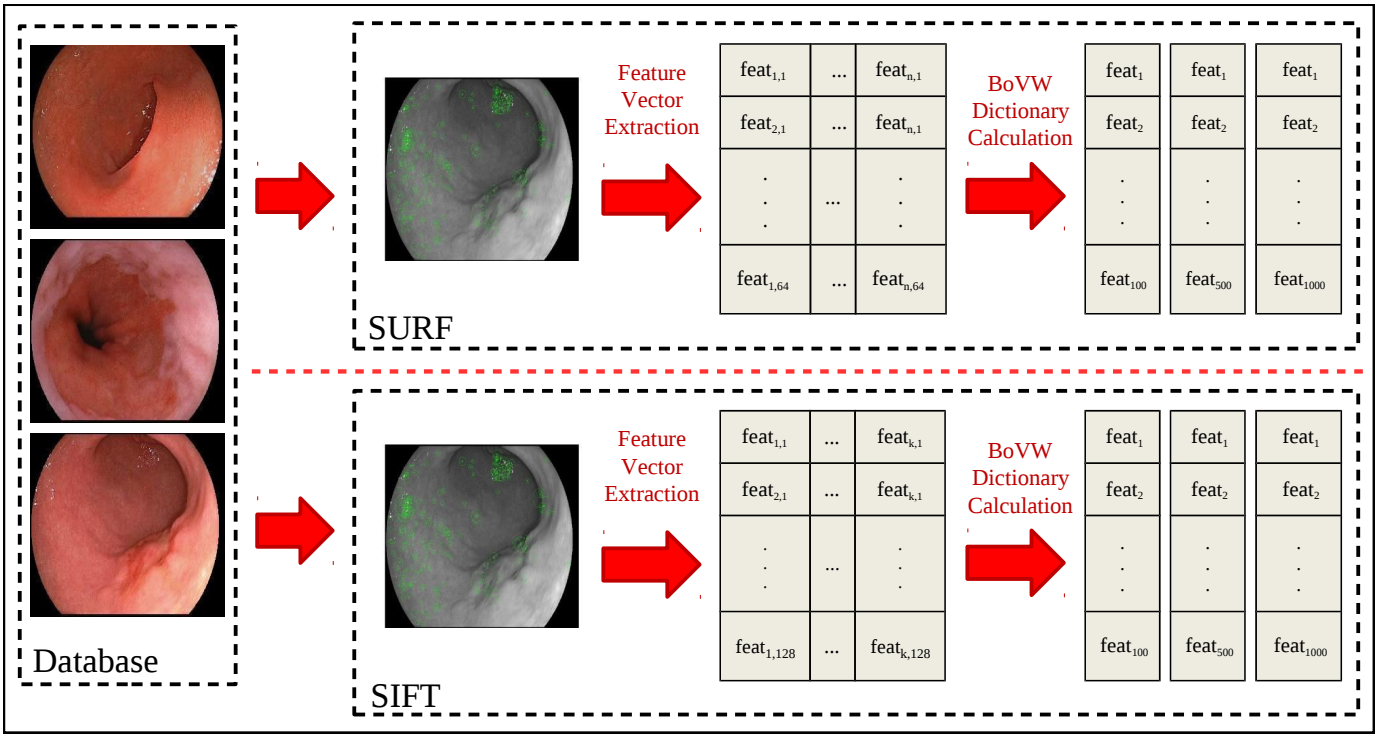

Fig. 2. Pipelined adopted in the paper for feature extraction concerning BE automatic identification.

TABLE I
SENSITIVITY, SPECIFICITY AND ACCURACY RESULTS USING SURF FEATURES.

Dictionary		100			500			1,000		
		SE	SP	AC	SE	SP	AC	SE	SP	AC
k-means	OPF	0.698	0.711	**0.700**	0.732	0.7823	***0.738**	0.714	0.777	**0.736**
	SVM-RBF	0.644	0.704	0.636	0.657	0.692	0.648	0.639	0.665	0.626
	SVM-Linear	0.614	0.672	0.620	0.582	0.593	0.586	0.608	0.627	0.628
random	OPF	0.688	0.739	**0.697**	0.692	0.718	**0.702**	0.664	0.695	**0.661**
	SVM-RBF	0.615	0.656	0.610	0.651	0.678	0.634	0.629	0.657	0.621
	SVM-Linear	0.548	0.590	0.517	0.591	0.631	0.576	0.586	0.627	0.565

If one considers k-means for dictionary generation technique, the best result was achieved by OPF with 500 words as well. With respect to k-means, the best result were obtained by OPF with 500 words as well. The statistical test did point out that different dictionary sizes can influence the results concerning k-means for visual words selection. Finally, the best overall result considering different dictionary generation technique and size stands for OPF with k-means and 500 works with an accuracy as of 0.723 (highlighted with *). This is the best result one can summarize from Table II. At the end, the best results from SURF and SIFT approaches (* values) did not show statistical similarity among themselves.

TABLE II
SENSITIVITY, SPECIFICITY AND ACCURACY RESULTS USING SIFT FEATURES.

Dictionary		100			500			1,000		
		SE	SP	AC	SE	SP	AC	SE	SP	AC
K-means	OPF	0.705	0.773	**0.683**	0.735	0.806	***0.723**	0.727	0.761	**0.714**
	SVM-RBF	0.645	0.658	0.641	0.672	0.706	0.655	0.642	0.686	0.641
	SVM-Linear	0.552	0.572	0.552	0.577	0.631	0.568	0.583	0.617	0.673
Random	OPF	0.644	0.663	**0.664**	0.673	0.713	**0.707**	0.642	0.662	**0.712**
	SVM-RBF	0.651	0.672	0.621	0.666	0.694	0.656	0.649	0.678	0.637
	SVM-Linear	0.545	0.566	0.532	0.548	0.572	0.545	0.544	0.553	0.527

Considering the parameters defined for the experiments, the k-means clustering for the BoVW calculation needs only one parameter corresponding to the number of clusters, and hence, the number of visual words for the dictionaries. As cited before, the k values (100, 500, 1,000) were defined arbitrarily, but the aim was to define diferent values to evaluate the impact of the PoI number in the dictionarie calculation. For the SVM-RBF, the $gamma$ and $cost$ parameters were optimized aiming for a more accurate and problem-adjusted classification. The $gamma$ parameter defines how far the influence of the training examples reach, changing the model's behaviour. The $cost$ parameter trades off the missclassification, making it low and allowing more errors (usually called soft margin), or making it high, forcing the SVM to make input data stricter and pottentially overfiting (usually called hard margin). In this work, for each feature extraction and classification approach, the values of $gamma$ and $cost$ were defined as the best combination providing the best accuracy result, falling in the ranges of $[6104e^{-5}, 0.0025]$ and $[1, 8]$, respectively.

At some extent, one can conclude that SURF and SIFT provided similar results considering OPF with k-means and 500 words. Interestingly, SIFT did not benefit from larger dictionaries as occurred with SURF. Actually, SIFT extracted considerably more PoIs than SURF, which may be a reasonable explanation for that behavior, since using more words for SIFT would be desired. As a matter of fact, our next work will be guided to consider such scenario as well.

VI. CONCLUSIONS AND FUTURE WORKS

CAD systems are a promising approach to evaluate various types and stages of dysplasia in patients suffering from Barrett's esophagus. In this paper, we introduced the Optimum-Path Forest classifier in the context of BE automatic identification by means of bag-of-visual words generated by SURF and SIFT techniques. The proposed approach was applied to full images annotated by five experts, providing a gold standard for the identification of malignant lesions. Since one is dealing with supervised classification problems, the proposed approach required a reference database comprising endoscopy images from patients with non-cancerous and cancerous linings in the esophagus.

An analysis of the sensitivity, specificity and accuracy results allows us to conclude the OPF classifier operating on the BoVW approach can provide much more efficient results than SVM with RBF and linear kernels for the problem of BE identification. The experiments can be extended using the annotated adenocarcinoma regions made by experts on the images for the BoVW dictionary calculation.

Considering the dictionary generation techniques, k-means showed as the best choice associated with OPF classifier. The SURF-based PoI extraction step required much more computational load than SIFT, which seems to be a bottleneck since the statistical test did not point out a clear difference between SURF and SIFT considering the effectiveness. Therefore, SIFT seems to be a better choice for feature description.

The SURF and SIFT techniques for the PoI extraction were selected considering the applicability and satisfatory results found in the medical image-description literature works [33]–[36], providing satisfatory generalization and rotation and translation invariance [28], [29]. SURF features were already used in [12] for the Barrett's esophagus analysis, but with a different dimension reduction approach. On the other hand, the SIFT technique was not yet applied for this problem. Therefore, combined with BoVW, SURF and SIFT could provide a new way for the Barrett's esophagus automatic diagnosis and analysis. These two feature extraction techniques provide the PoI of each image, however, the label of each image is based on the previous expert's annotations. In light of that, in the cancerous labeled images, PoI from non-cancerous regions are considered cancerous ones, once the descriptors of each image is composed of its PoI. The PoI label is basically an approximation, but it is intrinsic to the supervised classification in the full image approach, and there is the possibility of only visual words from non-cancerous regions be generated in some cancer labeled images. In this case, these visual words are labeled as well as cancerous, and can provide misclassification by inconsistent training.

In regard to future works, we intend to evaluate SIFT features under larger dictionaries, to employ OPF clustering [37] for BoVW selection, and to consider using deep learning right after histogram generation for each image as a post-processing feature learning approach. Also we consider to employ other pattern recognition techniques, such as k-Nearest Neighbors and Random Forest, as well as feature extraction techniques, such as fisher vectors and AKAZE features.

ACKNOWLEDGMENT

The authors are grateful to FAPESP grants #2014/12236-1 and #2016/19403-6, Capes, CNPq grant #306166/2014-3, as

well as Capes/Alexander von Humboldt grant #BEX 0581-16-0.

REFERENCES

[1] J. Lagergen, "Oesophageal cancer," *British Medical Journal*, vol. 341, no. c6280, 2010.

[2] J. Dent, "Barret's esophagus: a historical perspective, an update on core practicalities and predictions on future evolutions of management," *Journal of Gastroenterology and Hepatology*, vol. 26, pp. 11–30, 2011.

[3] C. Lepage, B. Rachet, and V. Jooste, "Continuing rapid increase in esophageal adenocarcinoma in england and wales," *The American Journal of Gastroenterology*, vol. 103, pp. 2694–2699, 2008.

[4] P. Sharma, J. J. G. H. M. Bergman, K. Goda, M. Kato, H. Messmann, and *et al*, "Development and validation of a classification system to identify high-grade dysplasia and esophageal adenocarcinoma in barretts' esophagus using narrow-band imaging," *Gastroenterology*, vol. 150, pp. 591–598, 2016.

[5] K. N. Phoa, R. E. Pouw, R. Bisschops, and *et al*, "Multimodality endoscopic eradication for neoplastic barret esophagus: results of an european multicentre study (euro-ii)," *Gut*, 2015.

[6] N. J. Shaheen, P. Sharma, B. F. Overhold, and *et al*, "Radiofrequency ablation in barret's esophagus with dysplasia," *The New England Journal of Medicine*, vol. 360, pp. 2277–2288, 2009.

[7] M. H. Johnson, J. A. Eastone, J. D. Horwhat, and *et al*, "Cryoablation of barret's esophagus: a pilot study," *Gastrointestinal Endoscopy*, vol. 62, pp. 842–848, 2005.

[8] B. F. Overholt, M. Panjehpour, and D. L. Halberg, "Photodynamic therapy for barret's esophagus with dysplasia and/or early stage carcinoma: long-term results," *Gastrointestinal Endoscopy*, vol. 58, pp. 183–188, 2003.

[9] F. V. D. Sommen, S. Zinger, W. L. Curvers, and *et al*, "Computer-aided detection of early neoplastic lesions in barrett's esophagus," *Endoscopy (online)*, 2016.

[10] M. Kandemir, A. Feuchtinger, A. Walch, and F. A. Hamprecht, "Digital pathology: Multiple instance learning can detecte barrett's cancer," in *IEEE 11th International Symposium on Biomedical Imaging*, 2014, pp. 1348–1351.

[11] A. Rosenfeld, V. Sehgal, D. G. Graham, M. R. Banks, R. J. Haidry, and L. B. Lovat, "Using data mining to help detect dysplasia: Extended abstract," in *IEEE International Conference on Software Science, Technology and Engineering*, 2014, pp. 65–66.

[12] L. A. Souza Jr, C. Hook, J. P. Papa, and C. Palm, *Barrett's Esophagus Analysis Using SURF Features*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2017, pp. 141–146.

[13] J. P. Papa, A. X. Falcão, and C. T. N. Suzuki, "Supervised pattern classification based on optimum-path forest," *International Journal of Imaging Systems and Technology*, vol. 19, no. 2, pp. 120–131, 2009.

[14] J. P. Papa, A. X. Falcão, V. H. C. Albuquerque, and J. M. R. S. Tavares, "Efficient supervised optimum-path forest classification for large datasets," *Pattern Recognition*, vol. 45, no. 1, pp. 512–520, 2012.

[15] J. P. Papa, S. E. N. Fernandes, and A. X. Falcão, "Optimum-path forest based on k-connectivity: Theory and applications," *Pattern Recognition Letters*, vol. 87, pp. 117–126, 2017.

[16] P. B. Ribeiro, K. A. P. Costa, J. P. Papa, and R. A. F. Romero, "Optimum-path forest applied for breast masses classification," in *IEEE 27th International Symposium on Computer-Based Medical Systems*, 2014, pp. 52–55.

[17] P. B. Ribeiro, L. A. Passos, L. A. Silva, K. A. P. Costa, J. P. Papa, and R. A. F. Romero, "Unsupervised breast masses classification through optimum-path forest," in *IEEE 28th International Symposium on Computer-Based Medical Systems*, 2015, pp. 238–243.

[18] P. B. Ribeiro, J. P. Papa, and R. A. F. Romero, "An ensemble-based approach for breast mass classification in mammography images," in *SPIE Medical Imaging*, 2017, pp. 101 342N–1–101 342N–8.

[19] J. P. Papa, A. A. Spadotto, A. X. Falcão, and J. C. Pereira, "Optimum path forest classifier applied to laryngeal pathology detection," in *15th International Conference on Systems, Signals and Image Processing*, 2008, pp. 249–252.

[20] C. T. N. Suzuki, J. F. Gomes, A. X. Falcão, J. P. Papa, and S. Hoshino-Shimizu, "Automatic segmentation and classification of human intestinal parasites from microscopy images," *IEEE Transactions on Biomedical Engineering*, vol. 60, no. 3, pp. 803–812, 2013.

[21] J. Hopkins, "Barrett's esophagus: Introduction," *Gastroenterology & Hepatology*, pp. 1–5, 2008.

[22] P. Sharma, J. Brill, M. Canto, D. DeMarco, B. Fennerty, N. Gupta, L. Laine, D. Lieberman, C. Lightdale, E. Montgomery, R. Odze, J. Tokar, and M. Kochman, "White paper aga: Advanced imaging in barretts esophagus," *Clinical Gastroenterology and Hepatology*, vol. 13, no. 13, pp. 2209 – 2218, 2015.

[23] M. Fujishiro, N. Yahagi, N. Kakushima, S. Kodashima, Y. Muraki, S. Ono, N. Yamamichi, A. Tateishi, Y. Shimizu, M. Oka, K. Ogura, T. Kawabe, M. Ichinose, and M. Omata, "Endoscopic submucosal dissection of esophageal squamous cell neoplasms," *Clinical Gastroenterology and Hepatology*, vol. 4, no. 6, pp. 688 – 694, 2006.

[24] J. A. Abrams, R. C. Kapel, G. M. Lindberg, M. H. Saboorian, R. M. Genta, A. I. Neugut, and C. J. Lightdale, "Adherence to biopsy guidelines for barrett's esophagus surveillance in the community setting in the united states," *Clinical Gastroenterology and Hepatology*, vol. 7, no. 7, pp. 736–742, 2009.

[25] A. Falcão, J. Stolfi, and R. Lotufo, "The image foresting transform theory, algorithms, and applications," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 26, no. 1, pp. 19–29, 2004.

[26] C. Allène, J.-Y. Audibert, M. Couprie, and R. Keriven, "Some links between extremum spanning forests, watersheds and min-cuts," *Image and Vision Computing*, vol. 28, no. 10, pp. 1460–1471, 2010, image Analysis and Mathematical Morphology.

[27] "Miccai 2015: 18^{th} international conference," 2015.

[28] H. Bay, A. Ess, T. Tuytelaars, and *et al*, "Speeded-up robust features (surf)," *Computer Vision and Image Understanding*, vol. 110, no. 3, pp. 346–359, 2008.

[29] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *International Journal of Computer Vision*, vol. 60, no. 2, pp. 91–110, Nov. 2004.

[30] L. C. S. Afonso, J. P. Papa, L. Papa, A. N. Marana, and A. Rocha, "Automatic visual dictionary generation through optimum-path forest clustering," in *19th IEEE International Conference on Image Processing*, Sept 2012, pp. 1897–1900.

[31] F. Wilcoxon, "Individual comparisons by ranking methods," *Biometrics Bulletin*, vol. 1, no. 6, pp. 80–83, 1945.

[32] Itseez, "Open source computer vision library," https://github.com/itseez/opencv, 2015.

[33] L. j. Zhi, S. m. Zhang, D. z. Zhao, H. Zhao, and S. k. Lin, "Medical image retrieval using sift feature," in *2009 2nd International Congress on Image and Signal Processing*, Oct 2009, pp. 1–4.

[34] A. Wojnar and A. M. G. Pinheiro, "Annotation of medical images using the surf descriptor," in *2012 9th IEEE International Symposium on Biomedical Imaging (ISBI)*, May 2012, pp. 130–133.

[35] S. Cui, H. Jiang, Z. Wang, and C. Shen, "Application of neural network based on sift local feature extraction in medical image classification," in *2017 2nd International Conference on Image, Vision and Computing (ICIVC)*, June 2017, pp. 92–97.

[36] S. Khan, S. P. Yong, and J. D. Deng, "Ensemble classification with modified sift descriptor for medical image modality," in *2015 International Conference on Image and Vision Computing New Zealand (IVCNZ)*, Nov 2015, pp. 1–6.

[37] L. M. Rocha, F. A. M. Cappabianco, and A. X. Falcão, "Data clustering as an optimum-path forest problem with applications in image analysis," *International Journal of Imaging Systems and Technology*, vol. 19, no. 2, pp. 50–68, 2009.