Vocal Folds Disorders Detection and Classification in Endoscopic Narrow-Band Images

Sara Moccia*†, Elena De Momi*, Giuseppe Baselli*, Leonardo S. Mattos†
*Department of Electronics, Information and Bioengineering
Politecnico di Milano, Milano (MI), Italy
Email: sara.moccia@iit.it
†Department of Advanced Robotics
Istituto Italiano di Tecnologia, Genova (GE), Italy

***Abstract*—The diagnosis of vocal folds (VF) diseases is error-prone due to the large variety of diseases that can affect them. VF lesions can be divided in nodular, e.g. nodules, polyps and cysts, and diffuse, e.g. hyperplastic laryngitis and carcinoma. By endoscopic examination, the clinician traditionally evaluates the presence of macroscopic formations and mucosal vessels alteration. Endoscopic narrow-band imaging (NBI) has recently started to be employed since it provides enhanced vessels contrast as compared to classical white-light endoscopy. This work presents a preliminary study on the development of an automatic diagnostic tool based on the assessment of vocal cords symmetry in NBI images. The objective is to identify possible protruding mass lesions on which subsequent vessels analysis may be performed. The method proposed here is based on the segmentation of the glottal area (GA) from the endoscopic images, based on which the right and the left portions of the vocal folds are detected and analyzed for the detection of protruding areas. The obtained information is then used to classify the VF edges as healthy or pathological. Results from the analysis of 22 endoscopic NBI images demonstrated that the proposed algorithm is robust and effective, providing a 100% success rate in the classification of VF edges as healthy or pathological. Such results support the investment in further research to expand and improve the algorithm presented here, potentially with the addition of vessels analysis to determine the pathological classification of detected protruding areas.**

***Index Terms*—Laryngoscopy; narrow band imaging; lesion detection; segmentation; automatic diagnostic tool.**

I. Introduction

Vocal folds (VF) disorders can be divided in benign and malignant lesions. Differences between the two classes are related both to VF macroscopic appearance and to blood vessels alterations. From the macroscopic point of view, lesions can be divided in nodular, e.g. nodules, polyps and cysts, which have smooth, regular surface and distinct margin surrounded by healthy tissue, and diffuse, e.g. hyperplastic laryngitis and carcinoma, which are irregular, rough and usually surrounded by inflamed tissue. Statistically, half of laryngeal cancers involve the vocal folds and, histologically, more than 90% of vocal folds cancers are squamous cell carcinoma [1]. Furthermore, it is now widely accepted that tumor growth and metastases are angiogenesis-dependent [2]. In particular, mucosal malignancies development is accompanied by the formation of intraepithelial papillary capillary loops (IPCL) on the mucosal surface, whose extension, expansion, irregularities in caliber and arrangement change as the tumor grows. In contrast, in benign lesions vessels run parallel to each other and to the mucosal surface [3].

The diagnosis of VF disorders is commonly performed by the endoscopic examination of larynx structural defects. Although endoscopic examination performed by a clinician represents the clinical gold standard, it has some drawbacks. It requires an expert to obtain the diagnosis and thus it is expensive in termst of time and costs. It is also characterized by low reliability across operators, sessions and sites and sometimes it is error-prone because of the variety of VF anomalies [4]. Hence, a computer-assisted scheme could considerably help in increasing the efficiency in managing patients, leading to potentially enormous benefits resulting from an improved detection of early tumors.

There have been very few attempts to create systems for automated analysis of the larynx. In [5] multiple feature sets were exploited to characterize both white-light endoscopic images and voice signals with the addition of patients' questionnaire. Features selection and classifier design were combined into the same learning process based on genetic search. In [6] the categorization of VF disease was treated as a pattern recognition task, based on color, texture and geometrical features. A pattern classifier then analyzed these feature-based VF representations. Despite the good results achieved, the main limit of these approaches is related to the need of a large ground truth set to train the classifier. Classification of laryngeal disorder based on shape and vascular defects of VF was achieved in white-light endoscopy in [7]. First a region of interest (ROI) containing the whole VF was identified using histogram of oriented gradients (HOG). Then the evaluation of mass lesions was performed analyzing VF edges. Separately, vessels segmentation was performed using region-based algorithms. Turkmen et al. reached good results (sensitivity of 86%, 94%, 80% 73% and 76% for healthy, polyp, nodule, laryngitis and sulcus vocalis) but kept out malignant vocal fold disorders as papillomata and carcinoma, thus the most frequent and life-threatening pathologies. Additionally, they did not take advantage of the VF edges investigation to identify possible protruding masses, performing vessels extraction in a wider area in respect to the protruding lesion, with a

higher computational cost and a higher risk of obtaining false negative.

In the last few years, narrow band imaging (NBI) endoscopy has been introduced as an inspection tool able to contrast vascular pattern better than the classical white-light endoscopy [8]. The NBI system consists of filters that narrow the classical endoscopy white-light illumination source as to obtain two bands in frequency in range of blue (400-430 nm) and green (525-555 nm). The blue filter is designed to correspond to the hemoglobin absorption peak, enhancing IPCL, while the green light, less scattered, penetrates deeper and thus contrasts submucosal plexus. Changes of IPCL are visible in NBI images as brown dots, typically not shown in presence of benign formations [9]. In [10] a promising work related to vessel segmentation in NBI endoscopy (NBI) images has been presented, where the classification between healthy and cancer tissues is performed by a statistical analysis of blood vessels characteristics, as thickness, tortuosity and shape.

The long-term goal of this work is to develop an automatic decision support method for VF disorders analysis in NBI images, which is based on both the investigation of macroscopic changes, to differentiate healthy and pathological tissue, and vessels modification, as to discern between benign and malignant lesions, taking advantage of NBI technique benefits.

This paper is concerned with the VF symmetry analysis with the intent of classifying healthy and pathological VF edges. The proposed method includes images pre-processing, to remove specular reflections, glottal segmentations, glottal axis detection and glottal left (LP) and right (RP) portion comparison to identify possible mismatching areas consequently identified as protruding lesions.

II. Methodology

A. Pre-processing

Due to the strong illumination of the endoscope and the wet and smooth surface of laryngeal tissue, specular reflections are often present in endoscopic images (Fig. 1a). Specular reflections can negatively affect the perceived image quality. Furthermore, they may be a significant source of error, especially for algorithms that make use of the gradient information in an image. Image pre-processing is therefore necessary to obtain reliable results from the automatic segmentation of VF structures. In this work, specular reflections were removed using sequentially bilateral filter (BF) [11] and Canny edge detector (CED) [12].

Using BF, every pixel intensity value is replaced by a weighted average of intensity values from nearby pixels. The weights, usually based on a Gaussian distribution, do not depend only on the Euclidean distance between the pixels, but also on the color intensity differences. BF is thus an appropriate tool since specular reflections are almost homogeneous internally while are characterized by a high color contrasts in respect to surrounding tissues. In this work, adopted neighborhood mask size was 15x15 pixels and both color space and coordinate space BF parameters were fixed to 215. The subsequent adoption of CED lead to the detection of specular reflections edges. To avoid eliminating regions that may be meaningful for further processing steps, too large detected specular regions, according to an empirical threshold of 200 pixels, were not removed. The color behind each specular reflection was replaced by neighbor pixels mean value (Fig. 1b).

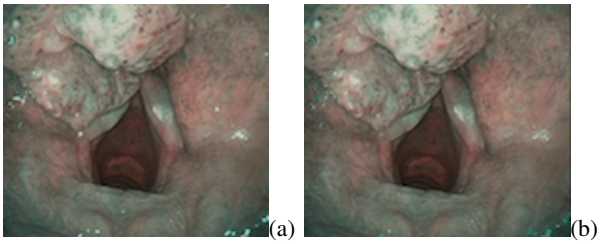
(a) (b)

Fig. 1. Vocal folds narrow-band images before (a) and after (b) specular reflections removal. To avoid eliminating meaningful regions, specular reflections larger than an empirical threshold of 200 pixels are not removed

B. Glottal Area Segmentation

The second step in the proposed method is the identification of the glottal area (GA), i.e., the area between VF edges. The process starts with the detection of the image region that contains the vocal cord tip. Here, this was performed using the histogram of oriented gradients (HOG). In [7] a similar approach was carried out for white-light endoscopy images to detect the region surrounding VF.

In this work, a VF reference dataset was assembled with images extracted from endoscopic videos of larynx in NBI mode. HOG descriptors of the images in dataset were obtained and later used to find the best match with HOG descriptors of new endoscopic images. Furthermore, since NBI VF endoscopy is usually performed nearby edges and the vocal folds are not always entirely visible, triangular images containing the vocal cord tip were used as the reference set.

During the image analysis, the HOG descriptor obtained from scrolling windows is compared with descriptors of reference images. A 1-D HOG with [-1,0,1] gradient filter and linear gradient voting into 9 orientation bins [0-180] was used. The best match was chosen according to the least Euclidean distance. An example of detected vocal folds tip-containing area is shown in figure 2a.

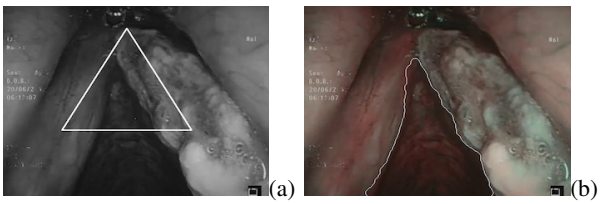
(a) (b)

Fig. 2. Vocal folds tip-containing area (a). Glottal area segmented with the proposed method (b).

Once the tip-containing region is obtained, a search is started to locate a 50x50 pixels area presenting the lowest intensity in that region. The center of this area is used as a

seed point to initialize a region-growing algorithm [13], which is carried out after having converted to gray and equalized the original image, as to guarantee a higher contrast between vocal cords edges (brighter) and GA (darker). In figure 2b, the achieved GA segmentation is shown on the original RGB image.

C. Mass Protruding Area Detection

A GA symmetry analysis is performed in order to find possible protruding masses. For this goal, the glottal axis is used to divide the segmented glottal area in two parts. To realize this, VF edges are approximated with a lower number of points, according to the Douglas-Peucker algorithm [14]. The glottal base is detected as the longest sequence of points aligned to the least inclined line. The glottal axis is then correspondingly traced. To avoid false negative associated to symmetrical lesions nearby VF tip, a line is tracked from the end of the base to the external upper vocal cord edge point, as to reproduce the healthy triangular shape (Fig. 3a). This approximation does not have effect on asymmetrical lesions, even if interesting the VF tip (Fig. 3b).

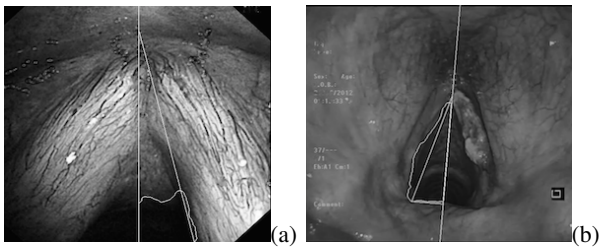
(a) (b)

Fig. 3. Triangular approximation of the biggest area identified after the symmetry analysis of glottal area to prevent false negative associated to symmetrical lesions in the upper part of vocal folds (a). Triangular approximation does not affect asymmetrical lesions, even if they interest vocal cords tip (b).

Finally, the mismatching part of the LP and RP detected areas is identified as protruding region (Fig. 4).

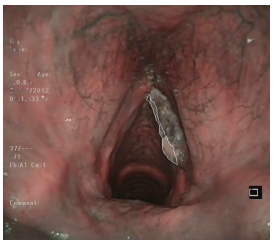

Fig. 4. Protruding mass portion detected with the proposed method.

III. Evaluation

A dataset of 22 VF NBI images was analyzed, 5 were healthy vocal folds and 17 pathological. Considered anomalies were carcinoma, keratosis and dysplasia. Images were provided by the "Clinica Otorinolaringoiatrica", San Martino Hospital, Genoa, Italy.

To evaluate the performance of the proposed method, a ground-truth reference set was built by an expert by manual segmentation of GA, LP and RP. Dice similarity coefficient (DSC) and volume ratio (VR) were computed to quantify the morphological similarity between the automatically detected and the manually traced GA, LP an RP areas:

$$DSC = \frac{2 * (A \cap B)}{A \cup B} \tag{1}$$

$$VR = \frac{A}{B} \tag{2}$$

where A and B are respectively the area obtained with the proposed method and the area obtained with the manual segmentation performed by an expert. Lilliefors test was performed to evaluate the normality distribution of analyzed samples. The results showed that the distributions were not normal and thus a Wilcoxon test (significance level = 0.05) was implemented to evaluate significant differences between manually and automatically traced GA, LP and RP.

IV. Results

The Wilcoxon test showed that there were not significant differences between the manual segmentation performed by an expert and the segmentation obtained with the proposed method.

The DSCs obtained for GA, LP and RP are shown in Fig. 5. Median values obtained were 0.93 for GA, 0.90 for LP and 0.90 for RP.

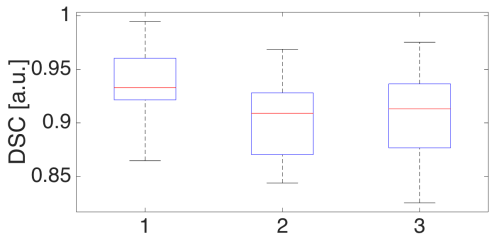

Fig. 5. Dice similarity coefficient between the manual segmentation and the results obtained with the proposed method. 1: total glottal area. 2: left glottal area portion. 3: right glottal area portion.

VRs median values obtained were 1.03 for GA, 1.10 for LP and 1.08 for RP, as shown in Fig. 6.

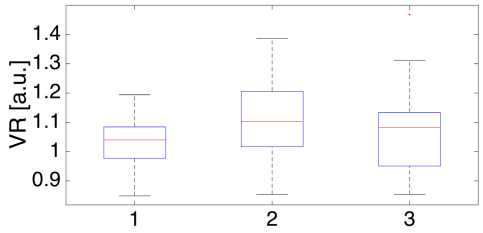

Fig. 6. Volume ratio between the segmentation achieved with the proposed method and the manual segmentation. 1: total glottal area. 2: left glottal area portion. 3: right glottal area portion.

All analyzed images were correctly classified in healthy and pathological VF edges: the proposed algorithm achieved

a 100% success rate in the detection of abnormalities in the VF images in the evaluation dataset.

The computational cost was high, over 5s per image in the computer used (processor: 2,4 GHz Intel core I5, Memory: 4 GB 1333 MHz). This was mainly due to the research of the VF tip-containing area. Anyway, this does not represent a limit due to the off-line application of the proposed method. Moreover, the code could be optimized to strongly reduce the computational cost.

V. Conclusion and future works

In this paper an automatic method for VF protruding mass individuation in VF NBI images has been presented. The asymmetry of VF edges was evaluated following the segmentation of the GA. The left and right glottal areas, detected based on the VF axis, were compared and the mismatching region was detected as protruding mass. This allowed the detection of abnormalities in the VF and its classification as health or pathological.

The performance of the proposed method was evaluated using DSC and VR metrics to compare automatically segmented and manually traced results for GA, LP and RP. All images in the evaluation dataset were correctly classified in healthy and pathological VF edges.

The main limit of the proposed method is related to the detection of small lesions that do not protrude in the glottal area. This limit could be overcome by performing further analysis in the region surrounding GA also in the case of healthy-classified VF edges, focusing the research in a smaller area in respect to the entire image.

The robustness of the obtained results supports the investment in further research to expand and improve the algorithm presented here. Considering the large number of medical studies related to the relationship between vessels features and VF diseases and the advantages of NBI technique, detection of suspicious areas and pathological classification could be performed including both vessels analysis and an automatic decision system, with the aim of developing a robust clinical support tool able to detect early lesions that may pass unnoticed to the human eye.

Acknowledgment

The authors would like to thank Prof. Giorgio Peretti and Dr. Luca Guastini from the San Martino Hospital (University of Genoa) for enlightening lessons about laryngeal diseases and for the endoscopic videos used in this research.

References

[1] P. Schultz, "Vocal fold cancer," *European annals of otorhinolaryngology, head and neck diseases*, vol. 128, no. 6, pp. 301–308, 2011.

[2] E. Ichihara, K. Kiura, and M. Tanimoto, "Targeting angiogenesis in cancer therapy," *Acta Med Okayama*, vol. 65, no. 6, pp. 353–362, 2011.

[3] S. Fujii, M. Yamazaki, M. Muto, and A. Ochiai, "Microvascular irregularities are associated with composition of squamous epithelial lesions and correlate with subepithelial invasion of superficial-type pharyngeal squamous cell carcinoma," *Histopathology*, vol. 56, no. 4, pp. 510–522, 2010.

[4] N. G. De Biase and P. A. de Lima Pontes, "Blood vessels of vocal folds: a videolaryngoscopic study," *Archives of Otolaryngology–Head & Neck Surgery*, vol. 134, no. 7, pp. 720–724, 2008.

[5] A. Verikas, A. Gelzinis, M. Bacauskiene, M. Hållander, V. Uloza, and M. Kaseta, "Combining image, voice, and the patient?s questionnaire data to categorize laryngeal disorders," *Artificial intelligence in medicine*, vol. 49, no. 1, pp. 43–50, 2010.

[6] A. Verikas, A. Gelzinis, M. Bacauskiene, and V. Uloza, "Towards a computer-aided diagnosis system for vocal cord diseases," *Artificial Intelligence in Medicine*, vol. 36, no. 1, pp. 71–84, 2006.

[7] H. I. Turkmen, M. E. Karsligil, and I. Kocak, "Classification of laryngeal disorders based on shape and vascular defects of vocal folds," *Computers in biology and medicine*, vol. 62, pp. 76–85, 2015.

[8] C. Piazza, O. Dessouky, G. Peretti, D. Cocco, L. De Benedetto, and P. Nicolai, "Narrow-band imaging: a new tool for evaluation of head and neck squamous cell carcinomas. review of the literature," *Acta otorhinolaryngologica italica*, vol. 28, no. 2, p. 49, 2008.

[9] P. Lukes, M. Zabrodsky, J. Plzak, M. Chovanec, J. Betka, E. Foltynova, and J. Betka, "Narrow band imaging endoscopic method for detection of head and neck cancer," *Endoscopy*, pp. 75–87, 2013.

[10] C. Barbalata and L. Mattos, "Laryngeal tumor detection and classification in endoscopic video," *IEEE journal of biomedical and health informatics*, 2014.

[11] C. Tomasi and R. Manduchi, "Bilateral filtering for gray and color images," in *Computer Vision, 1998. Sixth International Conference on*. IEEE, 1998, pp. 839–846.

[12] J. Canny, "A computational approach to edge detection," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, no. 6, pp. 679–698, 1986.

[13] A. Mehnert and P. Jackway, "An improved seeded region growing algorithm," *Pattern Recognition Letters*, vol. 18, no. 10, pp. 1065–1071, 1997.

[14] S.-T. Wu and M. R. G. Marquez, "A non-self-intersection douglas-peucker algorithm," in *Computer Graphics and Image Processing, 2003. SIBGRAPI 2003. XVI Brazilian Symposium on*. IEEE, 2003, pp. 60–66.