ELSEVIER

journal homepage: www.intl.elsevierhealth.com/journals/cmpb

Multiple feature sets based categorization of laryngeal images

A. Verikas[a,b,], A. Gelzinis[a], D. Valincius[a], M. Bacauskiene[a], V. Uloza[c]*

[a] *Department of Applied Electronics, Kaunas University of Technology, Studentu 50, LT-51368 Kaunas, Lithuania*
[b] *Intelligent Systems Laboratory, Halmstad University, Box 823, S-301 18 Halmstad, Sweden*
[c] *Department of Otolaryngology, Kaunas University of Medicine, LT-50009 Kaunas, Lithuania*

ARTICLE INFO

Article history:
Received 17 February 2006
Received in revised form
17 September 2006
Accepted 9 November 2006

Keywords:
Laryngeal image
Colour
Texture
Co-occurrence matrix
Support vector machine
Fourier transform

ABSTRACT

This paper is concerned with an automated analysis of laryngeal images aiming to categorize the images into three decision classes, namely *healthy*, *nodular*, and *diffuse*. The problem is treated as an image analysis and classification task. Aiming to obtain a comprehensive description of laryngeal images, multiple feature sets exploiting information on image colour, texture, geometry, image intensity gradient direction, and frequency content are extracted. A separate support vector machine (SVM) is used to categorize features of each type into the decision classes. The final image categorization is then obtained based on the decisions provided by a committee of support vector machines. Bearing in mind a high similarity of the decision classes, the correct classification rate of over 94% obtained when testing the system on 785 laryngeal images recorded at the Department of Otolaryngology, Kaunas University of Medicine is rather promising.

1. Introduction

The diagnostic procedure of laryngeal diseases in clinical practice is rather complex and based on evaluation of patient's complaints, history, and data of instrumental and histological examination. During the last 2 decades clinicians and researches have developed a variety of imaging techniques for examination of the larynx and objective measurements of voice quality [1,2]. Evaluation of larynx has improved significantly with the establishment of the computed tomography (CT) and magnetic resonance imaging (MRI), as the technologies provide insights into the endoscopically blind areas and reveal depth of tumour infiltration as well as cartilage and bone marrow invasion. The technologies may be beneficial in staging larynx carcinoma and planning the most appropriate surgical procedure [3–6]. Ultrasonography is useful in cases of larger laryngeal lesions and may have some role in screening unilateral vocal fold pathologies. At the same time, further fine-tuning of the technique may be necessary [7,8].

However, visualization of the larynx, by performing indirect video laryngostroboscopy and direct micro-laryngoscopy remains one of the most important diagnostic procedures, especially when making a primary diagnosis and then planning other more sophisticated diagnostic actions and the treatment. A physician evaluates colour, shape, geometry, contrast, irregularity, and roughness of the visual appearance of vocal folds. This type of examination is rather subjective and to a great extent depends on physician's experience.

* *Corresponding author at*: Intelligent Systems Laboratory, Halmstad University, Box 823, S-301 18 Halmstad, Sweden. Tel.: +46 35 167140; fax: +46 35 216724.
E-mail addresses: antanas.verikas@ide.hh.se (A. Verikas), adas.gelzinis@ktu.lt (A. Gelzinis), donatas.valincius@ktu.lt (D. Valincius), marija.bacauskiene@ktu.lt (M. Bacauskiene), uloza@kmu.lt (V. Uloza).

doi:10.1016/j.cmpb.2006.11.002

Availability of objective measures of these features would be very helpful for assuring objective analysis of laryngeal images and creating systematic databases for education, research, and health care purposes.

In addition to the data obtained from one particular patient, experience plays also a very important role in the decision making process. However, a physician interpreting the data from a particular patient may have a limited knowledge and experience in the analysis of the data. Moreover, to exploit the experience, a tedious and time consuming analysis of a large amount of data (images for example) may be required sometimes. In such a situation, a decision support system for an automated analysis and interpretation of medical data is of great value. Therefore, physicians would very much appreciate having a system able to automatically categorize the laryngeal images into several decision classes corresponding to different diseases. Such a system would enable a physician to effectively exploit a priori information available from a database of laryngeal images in screening laryngeal diseases.

Due to a large variety of appearance of vocal folds, the categorization task is sometimes difficult even for a trained physician [9,10]. Fig. 1 presents three examples of laryngeal images. The image placed on the right-hand side of the figure comes from the *diseased* class, while the other two are taken from the *healthy* vocal folds. In this case, the only discriminative feature is the slightly convex vocal fold edges in the upper part of the image coming from the *diseased* class.

Attempts made to develop computer-aided systems for analyzing vocal fold images are few. In Ref. [11], a system for automated categorization of manually marked suspect lesions into *healthy* and *diseased* classes is presented. The categorization is based on textural features extracted from co-occurrence matrices computed from manually marked areas of vocal fold images. A correct classification rate of 81.4% was reported when testing the system on a very small set of 35 images. In our previous studies [12,13], a committee of multilayer perceptrons employed for categorizing vocal fold images into three decision classes correctly classified over 90% of test set images.

In this study, aiming to obtain a comprehensive description of laryngeal images, we extract multiple feature sets coming from different processing approaches. Image colour distribution, distribution of the image intensity gradient direction, parameters characterizing the geometry of edges of vocal folds, distribution of the spectrum of the Fourier transform of the colour image complex representation, and parameters calculated from multiple co-occurrence matrices are the feature types used to describe laryngeal images. By extracting different types of features, we aim to exploit various image domains that could prove useful for image categorization into three decision classes. Description of the classes will be given shortly. Aiming to better exploit information contained in extracted features, the kernel principal components [14,15] are used for classification instead of the original features. A separate support vector machine (SVM) is used to categorize kernel principal components computed from features of each type into the decision classes. The final image categorization is then obtained based on the decisions provided by a committee of support vector machines.

The variety of feature types used to characterize laryngeal images, the KPCA based preprocessing, and the SVM committee based classification are the main differences between the technique utilized in this work and the studies presented in Refs. [11–13]. In Ref. [11], ordinary co-occurrence matrix based features have been utilized. The Fourier power spectrum, image intensity gradient, edge geometry, and the co-occurrence matrix based features are the new feature types utilized in this study if compared to those exploited in Refs. [12,13]. The processing approach applied allowed to increase the classification accuracy of laryngeal images.

The remainder of the paper is organized as follows. In the next section, we briefly describe the data used. Section 3 outlines the analysis techniques employed. Section 4 presents the results of the experimental investigations. Finally, conclusions of the work are given in Section 5.

2. Images used

A set of 785 laryngeal images recorded at the Department of Otolaryngology, Kaunas University of Medicine during the period from October 2002 to December 2003 is used in this study. The images were acquired during routine direct micro-laryngoscopy employing the Moller-Wedel Universa 300 surgical microscope. The 3-CCD Elmo colour video camera of 768×576 pixels was used to record the images. To lessen the influence of variation of the image capturing conditions on image appearance, we apply the multi-scale *retinex* theory-based colour image enhancement [16,17]. Details on how the enhancement has been applied can be found in Ref. [12].

A rather common, clinically discriminative group of laryngeal diseases was chosen for the analysis, i.e. mass lesions of vocal folds. In the study group, diagnosis of the mass lesions of the vocal folds – *the gold standard* – was based on routine clinical examination of the patient, including patient's complaints, history, voice assessment, indirect and direct micro-laryngoscopy, and finalized (confirmed) by the histological examination.

Mass lesions of vocal folds could be categorized into six classes, namely *polypus*, *papillomata*, *carcinoma*, *cysts*, *kerato-*

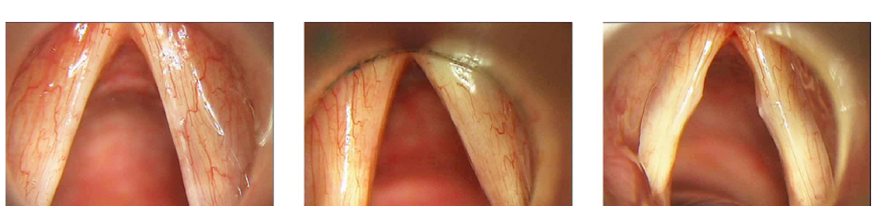

Fig. 1 – Three examples of laryngeal images.