COMPUTER METHODS AND PROGRAMS IN BIOMEDICINE 112 (2013) 455–465

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

Using image processing technology and mathematical algorithm in the automatic selection of vocal cord opening and closing images from the larynx endoscopy video

Chung-Feng Jeffrey Kuo[a,*], *Yueng-Hsiang Chu*[b], *Po-Chun Wang*[a], *Chun-Yu Lai*[a], *Wen-Lin Chu*[a], *Yi-Shing Leu*[c], *Hsing-Won Wang*[d]

[a] *Graduate Institute of Automation and Control, National Taiwan University of Science and Technology, No. 43, Keelung Road, Sec. 4, Taipei 106, Taiwan, ROC*
[b] *Department of Otolaryngology-Head and Neck Surgery, Tri-Service General Hospital, National Defense Medical Center, No 325, Cheng-Gong Rd. Sec 2, Neihu District, Taipei, Taiwan, ROC*
[c] *Mackay Memorial Hospital, Mackay Medical College, Taipei 104, Taiwan, ROC*
[d] *Graduate Institute of Clinical Medicine and Department of Otolaryngology, College of Medicine, Taipei Medical University, Shuang Ho Hospital, No. 291, Jhong-Jheng Road, Jhonghe District, New Taipei City, Taiwan, ROC*

ARTICLE INFO

Article history:
Received 23 November 2012
Received in revised form
6 August 2013
Accepted 8 August 2013

Keywords:
Videolaryngoscope
Image processing
Vocal cord dynamic images
Automatic vocal cord image selection

ABSTRACT

The human larynx is an important organ for voice production and respiratory mechanisms. The vocal cord is approximated for voice production and open for breathing. The videolaryngoscope is widely used for vocal cord examination. At present, physicians usually diagnose vocal cord diseases by manually selecting the image of the vocal cord opening to the largest extent (abduction), thus maximally exposing the vocal cord lesion. On the other hand, the severity of diseases such as vocal palsy, atrophic vocal cord is largely dependent on the vocal cord closing to the smallest extent (adduction). Therefore, diseases can be assessed by the image of the vocal cord opening to the largest extent, and the seriousness of breathy voice is closely correlated to the gap between vocal cords when closing to the smallest extent. The aim of the study was to design an automatic vocal cord image selection system to improve the conventional selection process by physicians and enhance diagnosis efficiency. Also, due to the unwanted fuzzy images resulting from examination process caused by human factors as well as the non-vocal cord images, texture analysis is added in this study to measure image entropy to establish a screening and elimination system to effectively enhance the accuracy of selecting the image of the vocal cord closing to the smallest extent.

1. Introduction

The human larynx is an important organ for voice production and respiratory mechanisms. The vocal cord is approximated for voice production and open for breathing. The videolaryngoscope is widely used for vocal cord examination. At present, physicians usually diagnose and present vocal cord diseases by manually selecting the image of the vocal cord opening to the largest extent (abduction), thus maximally exposing the

* *Corresponding author*. Tel.: +886 918854105; fax: +886 227908821.
E-mail address: jeffreykuo@mail.ntust.edu.tw (C.-F.J. Kuo).
0169-2607/$ – see front matter
http://dx.doi.org/10.1016/j.cmpb.2013.08.005

vocal cord lesion. On the other hand, the severity of diseases such as vocal palsy, atrophic vocal cord is largely dependent on the vocal cord closing to the smallest extent (adduction). This is because the voice is typically produced by the air forcedly squeezed through the approximated vocal cords. Therefore, diseases can be assessed by the image of the vocal cord opening to the largest extent, and the seriousness of breathy voice is closely correlated to the gap between vocal cord when closing to the smallest extent. However, in practical situation, the video recording the larynx sometime can be blurred or interrupted due to uncontrolled movement by patient or strong gag reflex by the tongue base, making the frame by frame selection of vocal cord open and close image difficult and time-consuming. Thus, the aim of the study was to design an automatic vocal cord image selection system to improve the conventional selection process by physicians and enhance diagnosis efficiency.

This study presented an automatic vocal fold recognition system that can capture the vocal area in a videolaryngoscope video automatically and can analyze images of the maximum opening and minimum closing positions of the vocal folds for diagnosis. First, the endoscope video provided by the hospital is imported into the system for glottis detection, and the vocal area is strengthened by image processing technology, in order to detect the maximum opening and minimum closing positions of the vocal folds. When all images from the video are finished processing, the texture analysis and the established screening mechanism are used to automatically eliminate unwanted images. This system can pick out images of the maximum opening and minimum closing positions of vocal folds automatically for diagnosis and subsequent disease recognition processing.

2. Literature review

With the development of digital technology, many scholars have conducted the real-time analysis of vocal fold vibration behavior. It is generally believed that there are two reasons for hoarseness including poor vocal cord closure and vocal fold vibration irregularities. The observation of vocal cord vibration requires high speed camera to record the vocal cord vibration completely. Using three ways of scanning the vocal cord images, Chodara et al. [1] measured the vocal cord vibration frequency, amplitude, symmetry and the vertical–horizontal phase difference, analyzed the vibration characteristics of the normal vocal cord, the vocal cord with vocal nodule and the vocal cord with single-side vocal polyp. Through quantified parameters, the normal and pathological vocal cord can thus be diagnosed and objective quantified data can be provided for clinical application. Švec et al. [2] used the high speed digital technology to develop a system for the observation of vocal cord mucosa fluctuation process. The system can be used to observe the vocal cord amplitude, frequency, symmetry, opening–closing glottis cycles and the mucosa fluctuations, and furthermore process and quantify vocal cord physiological parameters for the automatic evaluation of vocal cord vibration and health state. Eysholdt et al. [3] used the high-speed video camera (HSC) to record by shooting the vocal cord vibration process. The method is known as the high-speed glottography (HGG). Coupled with the image processing technology, it can automatically recognize and select location to present the vibration curves on the right and left side of the vocal cord by kymogram. Finally, through computer simulation and modeling, the irregular properties of the vocal cord vibration curves can thus be discussed. Larsson et al. [4] mainly analyzed the signals generated by the sound making mechanism. By combining the image processing technology, it can automatically track and segment the glottis area and make statistics of the pixel area of the glottis area to draw the glottis area waveform (GAW). Then, through the sound wave signal diagram and kymogram, it can be seen that the vocal cord repeatedly closes and opens in making the sound. The amplitude of GAW can be reflected by the sound waveforms of sound base frequency from the air density waveforms and mucosal vibration details to analyze the movements of normal and pathological vocal cords in making the sound. Qin et al. [5] proposed a comprehensive record vocal cord vibration information system consisting of high speed digital image and Electroglottography (EGG). To improve the accuracy and reliability of the inverse parameter model, Zernike Moments operation and improved level set method are used to detect the glottis contour edges in the image. Through the vocal cord model simulation combined with genetic algorithm, the inverse parameter is calculated to analyze the change in the vocal cord vibration mechanism and distinguish the normal and pathological vocal cords in diagnosis. Zhang et al. [6] proposed to use the in vitro excision throat experimental device coupled with the high speed photography system to analyze the vocal cord vibration curves of the normal vocal cord and vocal scar from the perspective of spatial domain, and work out the entropy values on the right and left sides of the vocal cords to diagnose the normal and pathological vocal cords for the evaluation of throat pathological clinical diagnosis. Chen et al. [7] used the image processing technology in the automatic segmentation of the glottis area and made statistics of the area of the glottis by the image morphology and regional growth method to draw the GAW according to the relationship between area and image time. Méndez et al. [8] used the Gabor filter to filter the vocal cord images and segment the glottis images using the image processing technology, and diagnosed the vocal cord diseases by statistics of the pixel number of the glottis binary images. Zorrilla et al. [9] applied the Gabor filter coupled with the Chan–Vese theory to segment the glottis image and analyze the vocal cord pathological information using the image including the vocal cord opening angle, deviation and area in the discussion of vocal cord paralysis and the normal vocal cord. Voigt et al. [10] applied the Phonovibrography method to objectively analyze voices and classify using a collective of 45 normophonic and paralytic voices. Captured the laryngeal dynamics by specialized Phonovibrogram features and analyzed with different machine learning. Elidan and Elidan [11] applied a preprocessing frame-by-frame crude midpoint identification and the energy minimization method to automatically detect the full glottal boundary throughout the video. Verikas et al. [12] applied multi-scale retinex theory based color image enhancement to lessen the influence of variation of the image capturing condition and use fourier transform to categorize the images into three decision classes,