31st Annual International Conference of the IEEE EMBS
Minneapolis, Minnesota, USA, September 2-6, 2009

Glottal Space Segmentation from Motion Estimation and Gabor Filtering

A. Méndez, E.M. Ismaili Alaoui, B. García, E. Ibn-Elhaj, I. Ruiz

Abstract— **Obtaining the glottal space segmentation is essential to characterize morphological disorders of vocal folds. In this study, the tested images are been acquired by direct optical inspection of the glottis using an endoscope and most of them are very poor quality. The application of motion estimation is very useful to segment the vocal folds endoscopic videos without user interaction. This approach involves three process steps: 1) Wiener motion estimator - to shift the measurement the next frame regarding to the current frame, and look for similarities between them. The best matching will accurate a shift equal to the displacement vector of the object; 2) Segmentation using motion estimation results and applying Gabor filtering; 3) Experimental results to demonstrate that the proposed method is effective. Our proposal works correctly with 95% of database test videos and it shows a great advance in design, and in the nearby future, a complete method to diagnose vocal folds pathologies.**

I. Introduction

Communication ability of human beings can be extremely influenced by voice disorders. When misfeeling in the larynx or changes in the voice pitch appear, it could be important to go to the specialist's office to examine the vocal folds movements, their colors and textures [1]. Applying videoendoscopy the examiner is able to visualize a virtual vibration of the high frequency motion of the vocal folds during phonation. In clinical routine, online findings rely on the visible evaluation of the video sequences and the doctors' experience.

The scientific community accepted method is to capture vocal folds videos by means of digital videoendoscopy and high speed recordings. But the problem is that the specialist delivers a diagnosis in a subjective way and it depends on his experience in this area.

The study of glottal space in a video sequence can be very useful and decisive to obtain an accurate diagnosis.

It is known that videostroboscopic imaging systems are not suited to derive correct information about abnormal vocal folds vibrations [2], but the method is interesting to detect pathologies related to the morphology. Other problem usually is the quality of the images due to the different luminosity, zoom and colour along the recording. Previously mentioned parameters depend on the specialist experience, the camera and the patient, and all of them are worse than in high speed recordings [4]. So, the characteristics of the images have to be unified.

In this approach, only low speed recordings illuminated with a stroboscopic light are going to be used (between 25 and 50 frames per second). We employ this kind of images due to their extended use among the otolaryngologists and voice specialists in Spain.

The present study proposes an approach for robust motion estimation between two successive image frames from a medical sequence. The method is based on generalized cross-correlation methods, where the phase of the Fourier components is used for motion parameter estimation. This method uses finite impulse response (FIR) filter, to sharpen the cross correlation maximum, thereby improving the accuracy of identification of the peak. For robust motion estimation between two successive image frames it is found that the WIENER estimator is particularly suited to this purpose. It is very common to find in the literature works related to analyze medical imaging taking into account the information from motion estimation [3].

There exist several problems to perform correctly the segmentation of the glottal space. Reviewing the literature we can find solutions related to active contour models [5], Watershed Transform [6] but all of them find problems to analyze the images without user interaction. The authors are going to combine motion estimation technique with the analysis of textures using a Gabor Filter Bank to obtain an optimum segmentation of glottal space.

II. Objectives

The main objective is to segment vocal folds videoendoscopic images without user interaction taking into account motion estimation information. The specific goals of this study are the followings:

- To analyze the video sequences from a commercial database: "Laryngeal Video stroboscopic Images (Dr. Wendy LeBorgne; Plural Publishing)".
- To detect the interest area motion (vocal folds and glottal space), using Wiener Estimator in the frequency domain..
- To obtain the glottal space segmentation based on the motion information
- To provide a report containing objective information about the image to the doctor.

Manuscript received April 23, 2009. This work was supported in part by the Spanish Department of Science and Technology under Grant TEC2006-12887-C02-02.

A. Méndez, B. García and I. Ruiz are with the University of Deusto, Bilbao, Spain (corresponding author e-mail: amendez@eside.deusto.es).

E.M. Ismaili Alaoui is with Faculty of Sciences University Mohamed V Rabat Agdal Morocco & with National Institute of Post and Telecommunications Rabat-Morocco.

E. Ibn-Elhaj is with the National Institute of Post and Telecommunications Rabat-Morocco.

III. PROPOSED SEGMENTATION SYSTEM

In this section, we present the main steps of the developed project. As it can be shown in the block diagram of figure 1, there are different stages where the video sequence is processed but we focus the research on blocks 1 and 2.

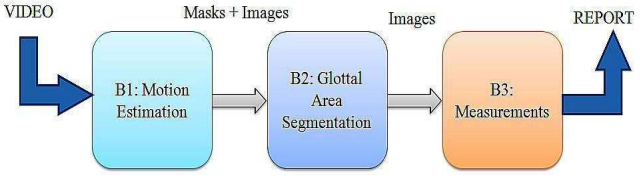


Fig. 1. Block diagram of the algorithm

A. Block 1: Motion Estimation

A1. Problem Formulation

The problems of motion estimation can be stated as follows:"Given an image sequence, compute a representation of the motion field that best aligns pixels in one frame of the sequence with those in the next". This is formulated as [7]:

$$g_{k-1}(x,y) = f_{k-1}(x,y) + n_{k-1}(x,y) \quad (1)$$

$$g_k(x,y) = f_{k-1}(x - d_k x, y - d_k y) + n_k(x,y) \quad (2)$$

where $g_k(x,y)$ and $g_{k-1}(x,y)$ are observed image intensities at instant k and $k-1$ respectively; $f_{k-1}(x,y)$ is noise-free frame; $n_k(x,y)$ and $n_{k-1}(x,y)$ are assumed to be spatially and temporally stationary, zero-mean image Gaussian noise sequences and $(d_k x, d_k y)$ is the displacement vector of the object during the time interval $[k-1,k]$.

If $n_k(x,y)$ and $n_{k-1}(x,y)$ are uncorrelated, the cross-correlation between $g_{k-1}(x,y)$ and $g_k(x,y)$. That is:

$$R_{g_k g_{k-1}}(r,l) = R_{f_k f_{k-1}}(r,l) \otimes \delta(r - d_k x, l - d_k y) \quad (3)$$

Where (r,l) are frames pixel coordinates respectively and $\otimes$ denotes convolution.

For non-identical frames several peaks can be simultaneously present. Therefore, the true cross-correlation is given by:

$$R_{g_k g_{k-1}}(r,l) = R_{f_k f_{k-1}}(r,l) \otimes \sum_i \delta(r - d_{k,i} x, l - d_{k,i} y \quad (4)$$

Selecting the maximum from the correlation surface in those cases does not provide the best estimate. A solution to this problem is to sharpen the true cross-correlation by using WIENER estimator.

A2.Problem Resolution

To accentuate the peak in the cross-correlation function associated with the motion vector, the input image frames can be pre-filtered. This operation needs the time and frequency domain. In the time domain, the image frames are filtered prior to displacement, multiplication, and integration, while in the frequency domain, a window or weighting function is applied to the cross spectral density (CSD) function, $S_{g_k g_{k-1}}(f_1, f_2)$, prior to performing the inverse Fourier transform. Thus the cross-correlation is given by [8]:

$$R^{(w)}{}_{g_k g_{k-1}}(r,l) = F^{-1}[\Psi_w(f_1,f_2) S_{g_k g_{k-1}}(f_1,f_2)] \quad (5)$$

Where the subscript w is to distinguish the WIENER and the weighting function for the WIENER estimator $\Psi_w(f_1,f_2)$ is defined by [8]:

$$\Psi_w(f_1,f_2) = (\gamma_{g_k g_{k-1}}(f_1,f_2))^2 \quad (6)$$

Note: $(\gamma_{g_k g_{k-1}}(f_1,f_2))^2$ is the ordinary coherence function between $g_k(x,y)$ and $g_{k-1}(x,y)$.

Thus, the WIENER estimator adjusts the CSD according to the value of the coherence and takes account of effect of noise in the estimation procedure, which will probably be more beneficial to estimate the motion vector. This method, therefore, has the desirable effect of suppressing those frequency regions where the coherence is poor.

The WIENER estimator $R_{g_k g_{k-1}}(r,l)$ is given by:

$$R_{g_k g_{k-1}}(r,l) = F^{-1}[S_{g_k g_{k-1}}(f_1,f_2) \gamma^2_{g_k g_{k-1}}(f_1,f_2)] \quad (7)$$

Where $\hat{S}_{g_k g_{k-1}}(f_1,f_2)$ is an estimation of the CSD.

The WIENER estimator may be realized by choosing (r,l) that maximizes (7) with proper weighting $\Psi_w(f_1,f_2)$ and proper estimate $S_{g_k g_{k-1}}(f_1,f_2)$.

B. Block 2: Glottal Space Segmentation

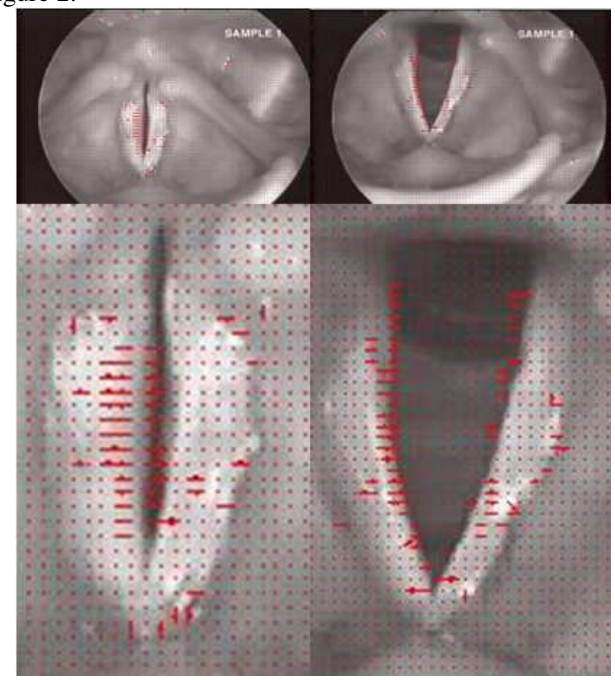
B1. Problem Formulation

In the introduction, some of the studied database recordings problems have been presented. One of them is the blurring caused by the stroboscope. But, the worst effect is the movement of the vocals folds during the recording process that makes difficult to apply the segmentation algorithms. This is caused because it is hard to know where the folds are going to be placed in the image. Knowing the information given by the motion estimation block, it can be applied a robust algorithm to estimate where the cords are. Then, the segmentation is made easily.

But when the movement between two frames is not meaningful we have to solve the segmentation with techniques related to image texture. To solve it, the authors apply Gabor Filters [9] which are bandpass filters used in image processing for feature extraction and texture analysis. These filters will provide the capacity necessary to highlight an image's features as regards a certain orientation and frequency.

B.2 Problem Resolution

The B2 module (figure 1) implements the segmentation of the glottal space. The input of this module can be seen in figure 2.



a) b)

Fig. 2. First block results and the zoom to observe better the motion vectors. a) Motion estimation for frame 1 and 2 of the sequence 4. b) Motion estimation for frame 67 and 68 of the sequence 4.

First of all, the image is reviewed automatically to know where the movement vectors are. Known this information, only this part of the original image is studied to avoid the oversegmentation problem.

After that, gradient image of greyscale frame is calculated. The method used to apply the gradient has been Sobel Operator [6].

Before Gabor filtering technique, a basic morphological transformation is applied. It is Erosion, following the definition "A pixel will have the value 1 in processing image if this pixel and all set of its neighbours are worth 1 in the original image".

Next step is to apply the Gabor Filter Bank defined in the equation 8 to highlight the characteristics of the image to be studied.

$$\Psi(f,\theta,x,y)=\exp\left(i(f_x x+f_y y)-\frac{f^2(x^2+y^2)}{2\sigma^2}\right) \quad (8)$$

Where:

$$f_x=f\cos\theta,\quad f_y=f\sin\theta \qquad i=\sqrt{-1} \quad (9)$$

V. Conclusion

In this paper, the WIENER motion estimation of vocal folds vibrations and the segmentation algorithm of the glottal space are presented. On one hand, the results demonstrate that the WIENER estimator produces a smooth displacement vector field with a very accurate measure of object motion. On the other hand, the WIENER estimator minimizes the influence of the noise and simplifies the identification of the

x and ***y*** are the coordinates of the image's pixels in the range (-x/2, x/2) y (-y/2, y/2). The pass-band filter's central frequency is ***f***, the spatial orientation θ, and the parameter σ determines the filter band's width.

Taking care not to let filter frequencies surpass π, the following values are determined for f_k:

$$fk=\frac{\pi}{\sqrt{2}^k} \quad k=2,3\ and\ 4 \quad (10)$$

And the orientations used are:

$$\theta_t=\frac{\pi t}{4},\ t=0,\ldots 3 \quad (11)$$

In the near future, block 3 will give the report to the specialist. This report will contain objective measurements (such as: glottal space area, symmetry, where is the pathology, the size of the pathology it exists…) obtained after the segmentation process.

IV. Results

The proposed algorithm has been tested with 201 images from 4 different videos. All the images have a resolution of 360x288 pixels and all of them belong to a commercial video and image database "Laryngeal Videostroboscopic Images (Dr. Wendy LeBorgne; Plural Publishing).

The ability of the WIENER estimator to estimate accurately the displacement vector field is demonstrated in figure 2. The motion vectors estimated between the frames 1 and 2 from the sequence 4 are shown in figure 2a and between frames 67 and 68 in figure 2b. The motion results achieve a high precision and a small measurement error. Because of the noise resistant property of the WIENER estimator, it produces reliable estimates.

Motion vectors delimit the region of the image to be studied. Because of that, the oversegmentation problem has disappeared due to the area of interest study.

After applying the designed algorithm (explained in section 3) over stored images, some results are shown in figure 3.

The algorithm works correctly in 95% of studied images comparing with a manual segmentation. This approach fails in the following cases:

- The vocal folds are close. The motion estimation does not give enough information and it is necessary to apply the segmentation process to the entire image. Then we can find some artefacts in the final result.
- There is not too movement between consecutive images. In this case, we use non-consecutive images to segment the glottal space.

dominant peak on the correlation surface. The movement vectors definition has been a reference to improve the segmentation process and to make it more efficient, so motion estimation information permits to delimit the part of the image to process.

The segmentation is the first step to objectify vocal folds characteristics using image processing techniques. This is the authors' area of interest, as they have been working in this area during last 3 years.

REFERENCES

[1] R. Cogwell Anderson, M. A. Rusch, S. Pitt, S. Stacy & K. A. Franke. "Observed Similarities in Four Adolescents with Paradoxical Vocal Fold Disorder". *The Internet Journal of Pulmonary Medicine*. 2005 Volume 5 Number 1

[2] Peter S. Popolo, Ingo R. Titze. "Qualification of a Quantitative Laryngeal Imaging System Using Videostroboscopy and Videokymography". *Ann Otol Rhinol Laryngol* 2008; 117:404-412.

[3] Y. Yan, X. Chen, D. Bless. "Automatic tracing of vocal-fold motion from high-speed digital images". *Biomedical Engineering, IEEE Transactions on*. Volume 53, Issue 7, Page(s):1394 – 1400. July 2006.

[4] Y. Yan, K. Ahmad, M. Kunduk, D. Bless. "Analysis of Vocal-fold Vibrations from High-Speed Laryngeal Images Using a Hilbert Transform-Based Methodology". *Journal of Voice*, Volume 19, Issue 2, Pages 161-175. 2008

[5] Allin, S. Galeotti, J. Stetten, G. Dailey, S.H. Enhanced snake based segmentation of vocal folds. *In proc. ISBI 2004*. 812- 815 Vol. 1.

[6] Victor Osma-Ruiz , Juan I. Godino-Llorente, Nicolás Sáenz-Lechón, Rubén Fraile. "Segmentation of the glottal space from laryngeal images using the watershed transform". *Computerized Medical Imaging and Graphics* 32 (2008) 193–201.

[7] E.M. Ismaili Alaoui, E. Ibn-Elhaj, "Estimation of Displacement Vector Field from Noisy Data using Maximum Likelihood Estimator", *In Proceedings IEEE ICECS07*, Pages: 1380 – 1383. Marrakech, Morocco.2007

[8] E.M. Ismaili Alaoui, E. Ibn-Elhaj, "Estimation of Motion Fields from Noisy Image Squences using Generalized Cross- Correlation Methods". *In Proceedings IEEE ICSPC07*, Pages. 1271-1274. Dubai, UAE.2007

[9] B. Bianconi, A. Fernández. "Evaluation of the effects of Gabor filter parameters on texture classification". *Pattern Recognition*. Volume 40, Issue 12, Pages 3325-3335. December 2007.

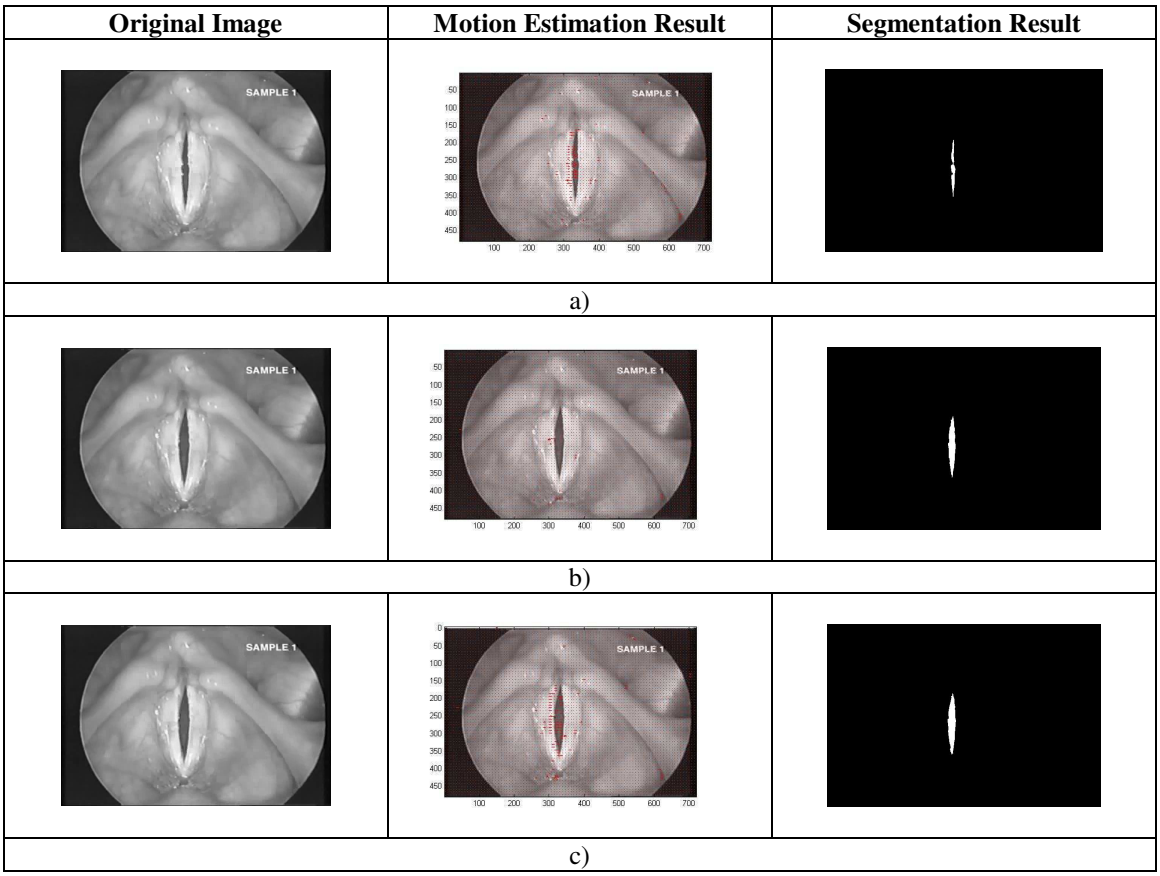


Fig. 3. Segmentation Results. a) Original Image, motion estimation for frame 3 and 4 of the sequence 2, segmentation result b) Idem for frame 4 and 5 of the sequence 2. c) Idem for frame 5 and 6 of the sequence 2