Pathological Vocal Folds Features Extraction Using a Modified Active Contour Segmentation

A. Méndez Zorrilla[1], Noha El-Zehiry[2], B. García Zapirain[1], Adel Elmaghraby[2]
1- University of Deusto. DeustoTech-LIFE Research Unit. Bilbao. Spain.
Email: amaia.mendez@deusto.es, mbgarciazapi@deusto.es
2- University of Louisville. Louisville. USA.
Email: adel@louisville.edu

Received: September 2010 Revised: November 2010 Accepted: December 2010

ABSTRACT:
This paper presents the study of vocal videostroboscopic videos to detect morphological pathologies using an active contour segmentation and objective measurements. The ad-hoc designed active contour algorithm permits to obtain a robust and fast segmentation using vocal folds images in RGB format. In this work, we have employed connected component analysis as a post-processing tool. After the segmentation process the image is analyzed and the pathology can be localized automatically and we can extract some features of each fold (such as the size of the polyp or cyst, the glottal space, the position…). Experimental results demonstrate that the proposed method is effective. Our proposal segments correctly the 95% of database test videos and it shows a great advance in design**.** The objective measurements complete a new method to diagnose vocal folds pathologies.

KEYWORDS: Segmentation, Active Contour Model, Objective Measurements, Videoendoscopy, Benign Pathologies

1. INTRODUCTION

Nowadays, the importance of an otorhinolaryngology specialist's diagnosis in the study of vocal folds pathologies is very high. The diagnosis and medical treatment depend solely on the subjective analysis of images captured by the fiberscope. This project is focused on the design and programming of an automated system to detect and evaluate vocal folds morphological pathologies.

The National Institute on Deafness and Other Communication Disorders (NIDCD) reports that in EEUU, 7.5 million people have diseases or disorders of the voice, which supports research into the nature, causes, diagnosis, and prevention of voice disorders [1].

The accepted methods to capture vocal folds videos are digital videoendoscopy and high speed recordings [7]. 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 othorhinolaryngologists and voice specialists. 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.

The most common pathologies have been analyzed in this study and they are the following:

- Nodules. This disorder prevents the vocal folds from meeting in the midline and thus produces an hourglass deformity on closure resulting in a raspy, breathy voice. Nodules are most common in children and females (see figure 1).

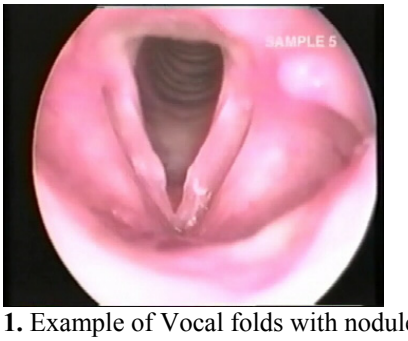

Fig 1. Example of Vocal folds with nodules

- Polyps. This pathology is a benign lesion of the larynx, occurring mostly in adult males that are usually located on the phonating margin of the vocal folds and prevent the vocal folds from meeting in the midline.
- Cysts. A cyst is a firm mass of organic material contained within a membrane. The cysts can be located near the surface of the

vocal fold or deeper, near the ligament of the vocal fold.

- Paralysis. The paralysis occurs when only one side is paralyzed in the paramedian position or has a very limited movement. It is more common than bilateral involvement.

Other problem of this kind of images usually is the quality of the images due to the different luminosity, zoom and color along the recording.

With the developed algorithm, the pathologies which affect vocal fold morphology (for example, nodules, polyps, cysts…) have been detected, and it is possible to obtain relative measurements regarding their size.

To obtain a fast, robust and automatic vocal folds segmentation (several algorithms are been applied to solve this problem [6]) it is worked in RGB format and it is applied an ad-hoc active contour model algorithm.

After the segmentation, the algorithm provides quantitative measurements related to the morphology of the cord (applied in benign and previously described pathologies). So, the doctor could support his diagnosis with the system results.

The main goal pursued by this research project is to develop a software application to aid the automated diagnosis of vocal folds pathologies studied by otorhinolaryngologist specialists.

This aim will be achieved through several specific objectives that are enumerated as follows:

- To create a database with records of endoscopic images of both healthy and pathological vocal cords.
- To establish a normalization and pre-processing of the images in the database.
- To design the algorithm using digital image processing techniques to segment the glottal space
- To define objective measurements to evaluate the benign morphological pathologies.

This paper is divided in the following principal sections: Section 2 describes the methods and proposed system, Section 3 presents the authors' results and Section 4 talks about the conclusions and future work.

2. METHODS AND PROPOSED SYSTEM

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

Fig. 2. Block Diagram

2.1. Segmentation Mathematical Formulation

The segmentation approach that we are proposing is devised in a graph cut framework and hence, we present a brief review of the graph terminology that will be used throughout the paper. A graph G={V,E} consists of a set of vertices v ∈ V and a set of edges e ∈ E ⊆ V×V. An edge eij is the edge incident to the vertices vi and vj. In our formulation, every edge eij is assigned a weight (cost) wij. The weights are used to reflect the segmentation energy function and to penalize the terminal points of the edge being assigned different labels in the segmentation scheme. An edge cut EC is a set of edges whose removal separates the graph in two disjoint sets S1 and S2 such that S1∪ S2= V. This may be represented by an indicator function x defined as:

$$x = \begin{cases} 1 \text{ if } v \in S1 \\ 0, \text{ otherwise} \end{cases}$$

The mapping between the image, the graph and the binary labeling is established by assigning a binary variable xp for each pixel p and each binary variable will be represented by a vertex vp in the graph G.

Our segmentation approach will formulate the pixel-labeling problem as a curve evolution problem in which an active contour evolves to capture the object of interest. Several segmentation approaches [3-5] have utilized this idea but our approach stands out by using graph cut optimization to solve the evolution problem, which makes our algorithm fast and robust.

The formulation of the segmentation problem is described as follows. Let u(x, y, i) be the intensity value of the pixel p=(x,y) in the ith channel. Each image consists of three channels R, G and B. For every pixel p, a binary variable xp will be assigned such that

$$x_p = \begin{cases} 1 \text{ if p is inside C} \\ 0, \text{ otherwise} \end{cases}$$

Let C be the contour that separates the two classes. Let C1=(c11, c21, c31) represent a vector of the mean intensity value inside the contour C for the three channels and alternatively, let C2=(c21, c22, c32) be the mean intensity value outside the contour in the three channels. Notice that if the evolving contour coincides with the boundary of the object of interest, the intensity variations around the means inside and outside the contour will be minimal. Alternatively, the boundary of the object of interest can be obtained by minimizing the following energy function:

$$F(x_1, x_2, \ldots, x_n, C_1, C_2) = \sum_{p\in\Omega} \frac{1}{3}\sum_{i=1}^{3} \lambda_{i1}|u(x,y,i)-c_{i1}|^2 x_p + \sum_{p\in\Omega} \frac{1}{3}\sum_{i=1}^{3} \lambda_{i2}|u(x,y,i)-c_{i2}|^2 (1-x_p) + \mu \text{ length (C)} \quad (1)$$

Where λ_{i1}, λ_{i2} and μ are constants that control

the effect of each term.

The pixel q is a neighboring pixel of p as depicted in figure 3. And wpq is the weight that penalizes the length of the contour and is given as [4]:

$$wpq = \frac{\delta^2 \Delta\theta}{|\, e_{pq} \,|} \tag{2}$$

The first two terms in the energy function represent the intensity variations around the means and the last term is a regularizer that minimizes the length of the contour. The values of C1 and C2 are updated in terms of xp s follows.

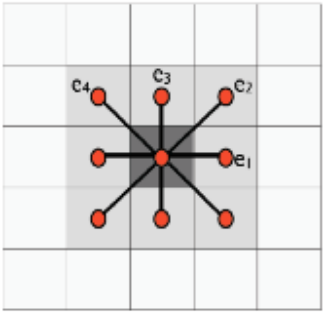

Fig.3. Neighborhood system of size 8. $\Delta\theta = \pi/4$.

$$C_{i1} = \frac{\sum_{p \in \Omega} u(x,y,i) x_p}{\sum_{p \in \Omega} x_p} \tag{3}$$

and

$$C_{i2} = \frac{\sum_{p \in \Omega} u(x,y,i)(1 - x_p)}{\sum_{p \in \Omega} (1 - x_p)} \tag{4}$$

To optimize an energy function using graph cuts, we have to guarantee that this function is graph representable and that the min cut of the graph will corresponding to the global optimum of the energy function. Fortunately, the feasibility of optimizing energy functions have been investigated by Kolmogorov and Zabih in [8] and by Friedman [9]. Their contributions led to an important result that is: Sumbodular functions are graph representable and can be globally optimized using graph cuts in polynomial time. A function E(x1, x2, …x n) is submodular if it can be expressed as a sum of unary functions or pairwise functions Eij (xi, xj) such that each pairwise interaction satisfies the submodularity constraint given as follows: Submodular functions are the set functions satisfying the constraint:

$$E(0,0) + E(1,1) \le E(0,1) + E(1,0). \tag{5}$$

A detailed discussion about submodularity can be found in [8].

According to the submodularity definition in (5). Our proposed energy function is submodular because the data terms are unary terms and the length term has a positive value if and only if the contour passes through the edge, i.e. the terminal vertices of the egde have different labeling and zero otherwise. So, E(0,1) and E(1,0) are positive while E(0,0) and E(1,1) are zero. Therefore, the graph cut optimization of the proposed function yields the global optimum in a polynomial time. The next section describes the details of the algorithm and the graph construction.

2.2. Step by Step algorithm

The algorithm can be summarized as follows:

1. Initialize a contour C in the input image.
2. For every pixel p=(x,y) in the image u assign a binary variable x_p such that x_p =1 inside C and zero outside C.
3. Calculate C_{i1} and C_{i2} using equations (3) and (4).
4. Construct a graph G with n+2 vertices where n is the number of pixels in the image u, i.e., every pixel p in the image has a corresponding vertex v_p in the graph and we add two auxiliary vertices S and T that will represent the labeling later.
5. Construct the graph edges as follows:
   a. Terminal Links: For every pixel p Calculate

      $E1 = \frac{1}{3}\sum_{i=1}^{3} \lambda_{i1} |\, u(x,y,i) - c_{i1} \,|^2$ and

      $E2 = \frac{1}{3}\sum_{i=1}^{3} \lambda_{i2} |\, u(x,y,i) - c_{i2} \,|^2$, if E1>E2 , add an edge e_{vpS} with weight $\frac{1}{3}\sum_{i=1}^{3} \lambda_{i1} |\, u(x,y,i) - c_{i1} \,|^2$, otherwise add an edge e_{vpT} with weight $\frac{1}{3}\sum_{i=1}^{3} \lambda_{i2} |\, u(x,y,i) - c_{i2} \,|^2$

   b. Neighboring Links: for every pixels p add neighboring links in an eight neighborhood system to represent the length of the contour as described in [4,10]

6. Find the minimum ST cut in the graph G. The cut partitions the graph into two sets S1 and S2 such that S ∈ S1 and T ∈ S2.

7. Update the values of the values of the binary variables as follows: if $v_p \in S1$, assigns a value 1 to x_p and if $v_p \in S2$, assigns a value 0 to x_p.

8. Repeat steps 3-7 until the energy is minimized.

2.3. Connected Component Analysis

The region of interest shares the same intensity of some areas in the background. Fortunately, the region of interest is far apart from these regions. In this paper, we have employed connected component analysis as a post-processing tool. This step highlights the different connected components in the foreground class and assigns a label for each component. Thereby, the ROI is simply identified as the central component of the foreground.

2.4 Features Measurements

Once the image is segmented, the method to extract some morphological characteristics and to localize the pathology (if it exists) is described in following steps:

1. At first, it is necessary to work on only with glottal area edge to obtain conclusions about it. The first step is to extract the lines that represent the left and right cords out of the image, in order to study them independently.

The objective is to detect if some pathology exists and the process consists of other two different steps that try to find out the points at which the pathology (if it exists) begins and ends, respectively:

2. To calculate the point where the pathology (nodule, polyp…) starts, using second derivative. Maximum and minimum value point are calculated and middle ground between them that is considered the point where the pathology starts, if it is before the lower half of the vocal fold, otherwise the points with less absolute value is utilized. It must be taken into account that both lines (the vocal folds) do not follow any well-known function, but when the cords have any pathology the discontinuities can be found out by this method.
3. The other algorithm finds out the ending point of the pathology by the intersection of the extracted line with its regression line. The regression line is calculated by the linear method (6):

$$b = \frac{1/n \cdot \sum (x_i - x_m) \cdot (y_i - y_m)}{1/n \cdot (x_i - x_m)^2} \qquad (6)$$

$$and \quad a = y_m - (b \cdot x_m)$$

Being b the gradient of the regression line and a the value of the function at zero. In (6), xi and yi represent the value of the original line on each point. Also, xm, ym are the mean value of all the points on each axe. By knowing the regression line, it can be extracted the ending point of the pathology with the intersections, calculated by the changing signs method.

In this study, the effciency of segmentation algorithm is also measured. This calculus is made per sequence knowing how many frames are incorrectly segmented.

$$Segmentation\ Error\ (\%) = \left(\frac{\text{Incorrectly segmented frames}}{\text{Total Sequence frames}}\right) * 100 \qquad (7)$$

3. RESULTS

In this research, we have used a commercial video and image database "Laryngeal Videostroboscopic Images (Dr. Wendy LeBorgne; Plural Publishing)" where sometimes the videos are captured by several devices too. The selected recordings contain sequences with polyps, nodules or cyst. The algorithms are been tested in 6 sequences, with 18 frames each (recordings 1 to 4 with pathological vocal folds and 5 and 6 recordings of healthy vocal folds).

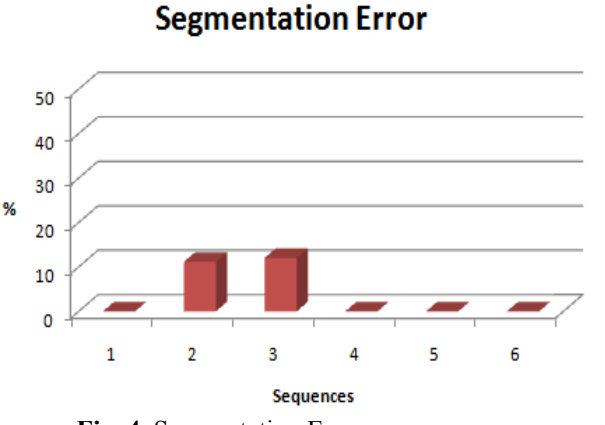

Fig. 4. Segmentation Error per sequence

In this study we can find partial results, the segmentation ones, and they can be seen in figure 4.

Figure 4 shows the percentage of frames segmented incorrectly per sequence and they are less than 10% in all cases (more graphic results in figure 6).

After the segmentation block (see figure 1), we calculate the area of glottal space and the area of the pathology, if it is exists.

Table 1 collects the data corresponding to the images of figure 6. In this table, glottal space area and the area of the pathology (nodules, polyp or cyst) are measured. In table 1, we can also see some errors: in

image 2 and image 5. The value "NO" means that no pathology has been found in the fold.

Table 1. Example of isolated frames measurements.

Sequence	Glottal Space Area (pix)	Left Fold Path. Area (pix)	Right Fold Path. Area (pix)
1- Nodules	4698	85.5	162
2- Nodules	**10151**	**21.5**	**NO**
3- Polyps	8906	91.5	NO
4- Polyps	40.99	-1	130
5- Cysts	**1564**	**NO**	**NO**

The doctor's diagnosis for image 2 is nodules and the measurement algorithm only detects abnormalities in left fold. This problem can be due to the size of the right nodule or the position of the cord in this frame (see figure 6). The same problem occurs in image 5.

Table 2. Complete sequence results of vocal folds with polyps.

Img.	GS Area (pix)	Pathology Area of Right VF(pix)	Pathology of Left VF Area (pix)
1	8368	98.5	NO
2	8810	105.5	NO
3	8707	100	NO
4	8795	99	NO
5	9015	90	NO
6	8737	104.5	NO
7	8644	115	NO
8	8906	91.5	NO
9	8755	118	NO
10	8776	90.5	NO
11	8743	116	NO
12	8682	112	NO
13	8625	88	NO
14	8794	86	NO
15	8679	95.5	NO
16	8425	102	NO
17	7814	81	NO
18	**4368**	**NO**	**NO**

Because of that, the authors analyze all the frames of the image sequence (usually 18 frames) because to give a diagnosis studying only a frame is not enough, due to the vibration deformations.

Table 2 shows the results of measurements corresponding to a vocal folds recording with polyps, and in table 3, we can find the results of a healthy vocal folds sequence (see figure 5). In both tables, we can see the glottal space measurement of each frame (first columnn) and the measurement of the right/left vocal fold pathology, if it is exists.

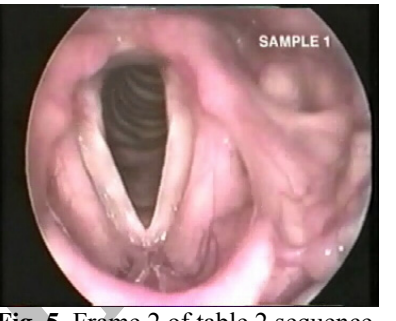

Fig. 5. Frame 2 of table 2 sequence.

Table 2 shows also an algorithm error in frame 18 of the sequence, because it does not detect the polyp in this frame. This problem is not significant to the final diagnosis due to the authors study all the frames and they make the average of the results. In this case, we detect the polyp in 17 frames, only in the last one the algorithm make a mistake.

Table 3. Complete sequence results of healthy vocal folds

Img.	GS Area (pix)	Pathology Area of Right VF(pix)	Pathology Area of Left VF (pix)
1	2698	NO	NO
2	4434	NO	NO
3	6133	NO	NO
4	6360	NO	NO
5	7040	NO	NO
6	7642	NO	NO
7	8802	NO	NO
8	8972	NO	NO
9	9085	NO	NO
10	8184	NO	NO
11	9451	NO	NO
12	9275	NO	NO
13	8033	NO	NO
14	9326	NO	NO
15	8844	NO	NO
16	8447	NO	NO
17	7786	NO	NO
18	7180	NO	NO

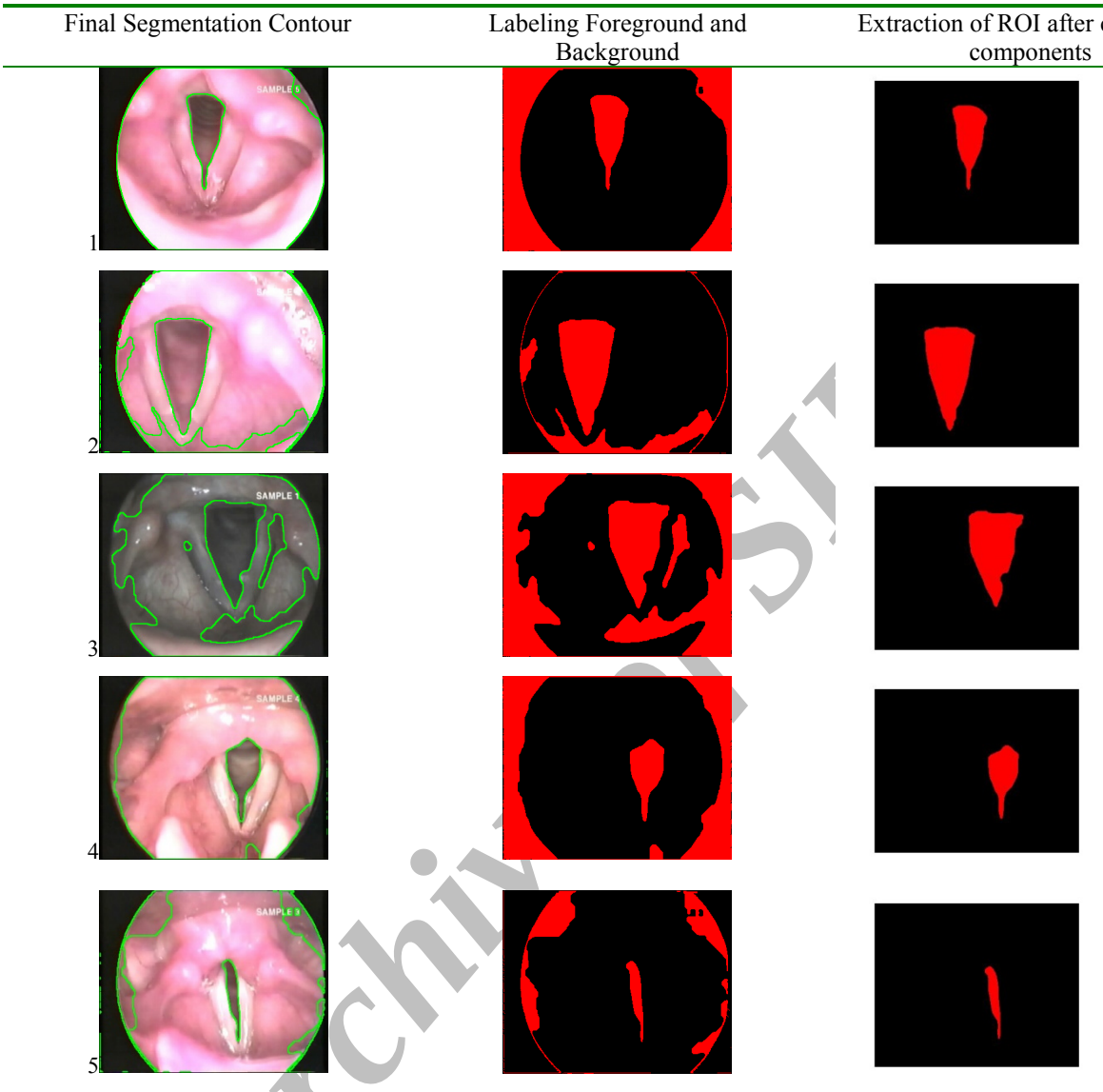

Fig. 6. Segmentation Results

These measurements (table 2 and table 3, areas in pixels) are relative, as these kind of images do not have enough information to obtain real measurements in centimetres.

4. CONCLUSIONS AND FUTURE WORK

This study is part of a complete design of an automated diagnosis application with which it will be possible to detect some benign morphological pathologies of the vocal folds (nodules, polyps, cysts and paralysis have been studied).

The study has been carried out using 106 laryngeal images. The methodology followed has reported some errors in static images inside the sequence (see image 18 in table 2), but this problem do not affect the final diagnosis due to the analysis of all the images in the sequence.

On one hand, the segmentation is the first step to objectify vocal folds characteristics using image processing techniques, specifically active contour models in this case. In the literature we can find other segmentation techiniques related with this subject, but the authors of this paper try to minimize the user interaction [11][12].

On the other hand, the is the authors' area of interest is the automatic diagnosis of vocal folds pathologies and because of that, it is very important to try to objectify the features of different pathologies.

These hopeful results open the possibility to make a broader study using more sequences and increasing the number of vocal folds pathologies.

The future work is being to be centred in the

optimization of segmentation and measurement algorithms.

5. ACKNOWLEDGMENT

The authors wish to acknowledge University of Deusto for lending their infrastructures and materials. This work was partially financed by Dep. of Education, Universities and Research from the Basque Country Government.

REFERENCES

[1] "**Taking care of your voice**". http://www.nidcd.nih.gov.

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

[3] J.S. Chang, E.Y. Kim, S. Hyun Park. "**Lip Contour Extraction Using Level Set Curve Evolution with Shape Constraint**", *Human-Computer Interaction*, Part III, LNCS 4552, pp. 583-588, (2007)

[4] Y. Boykov and V. Kolmogorov. "**Computing geodesics and minimal surfaces via graph cuts**", *International Conference for Computer Vision*, ICCV, (2003)

[5] L. D. Cohen. "**On active contour models and balloons**", *Computer Vision, Graphics, and Image Processing*, Vol. 53, No. 2, pp. 211–218, (1991)

[6] A. Méndez, B. García, I. Ruiz, I. Iturricha. "**Glottal Area Segmentation without Initialization using Gabor Filters**", *proc. ISSPIT*, pp. 18-22. Sarajevo, (2008)

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

[8] Vladimir Kolmogorov, Ramin Zabih. "**What Energy Functions Can Be Minimizedvia Graph Cuts?**", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 26 , Issue 2. pp. 147 - 159 (January 2004)

[9] H.L. Friedman. "**Lewis-Randall to McMillan-Mayer conversion for the thermodynamic excess functions of solutions. Part I. Partial free energy coefficients***". Journal of Solution Chemistry*, Vol. 1, No. 5, 387-412. (1972)

[10] Noha El-Zehiry, Steve Xu., Prasanna Sahoo, Adel Elmaghraby, "**Graph cut optimization for the mumford-shah model***"*, *Proceedings of the Seventh IASTED International Conference visualization, imaging and image processing*, pp.182–187. Springer, Heidelberg (2007).

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

[12] E. M. Ismaili Alaoui, A. Mendez, E. Ibn-Elhaj, B. Garcia. "**Keyframes detection and analysis in vocal folds recordings using Hierarchical Motion Techniques and Texture Information**". *Proc. of ICIP 2009*. El Cairo, Egypt. (2009)