**International Journal of Electrical and Computer Engineering (IJECE)**
Vol.2, No.4, August 2012, pp. 463~472
ISSN: 2088-8708


# Automatic Segmentation of Glottal Space from Video Images Based on Mathematical Morphology and the hough Transform

**Davod Aghlmandi, Karim Faez**
Department of electronics / Amirkabir University (Tehran Polytechnic), Iran
Department of computer and electronics / Qazvin Islamic Azad University, Iran
Qzvin, Iran
e-mail: D_aghlmandi@qiau.ac.ir, Kfaez@aut.ac.ir

**Article Info**

***Article history:***

Received Jun 12th, 2012
Revised Aug 20th, 2012
Accepted Aug 26th, 2012

***Keyword:***

Glottal area
Mathematical Morphology
Video Stroboscopy
Vocal Folds

**ABSTRACT**

Vocal disorders directly arise from the physical shape of the vocal cords. Videostroboscopic imaging provides doctors with valuable information about the physical shape of the vocal cords and about the way these cords move. Segmentation of the glottal space is necessary in order to characterize morphological disorders of vocal folds. One of the main problems with the methods presented is their low level of accuracy. To solve this problem, an automatic method based on Mathematical Morphology edge detection and the Hough transformation is presented in this article to extract the glottal space from the videostroboscopic images presented. This method and two other popular algorithms, histogram and active contour, are performed on 10 sets of videostroboscopy data from excised larynx experiments to compare their performances in analyzing videostroboscopy images. The accuracy in computing glottal area of these methods are investigated. The results show that our proposed method provides the most accurate and efficient detection, and is applicable when processing low-resolution images. In this paper we used edge detection based on geometric morphology to detecting the edges of vocal cords. Then in the next step the edges that were extracted, using Hough transform change to some lines. After that through using proposed algorithm, we omit the extra lines and extract the glottis.



***Corresponding Author:***

Davod Aghlmandi,
Department of Electronics,
Amirkabir University,
Qzvin, Iran.
Email: D_aghlmandi@qiau.ac.ir

## 1. INTRODUCTION

It is necessary to study structure and function of the vocal folds in order to determinate normality of the voice. The scientific community uses two methods: high speed digital imaging [2] and low speed imaging with a stroboscopy [3] to evaluate and identify the movements of the vocal folds. A sample of stroboscopic images taken of the vocal folds at the stages of opening and closing is shown in Figure 1.

In examining the larynx using the videostroboscopy method, an endoscope is used which is connected to a camera and to a light source. The endoscope enters the mouth and is placed at the entrance of the larynx to start filming the structure and the operation of the larynx, and especially of the vocal cords. At present, evaluating the larynx in the videostroboscopy method is one of the most practical and most often used methods of examining the larynx because it has numerous advantages.

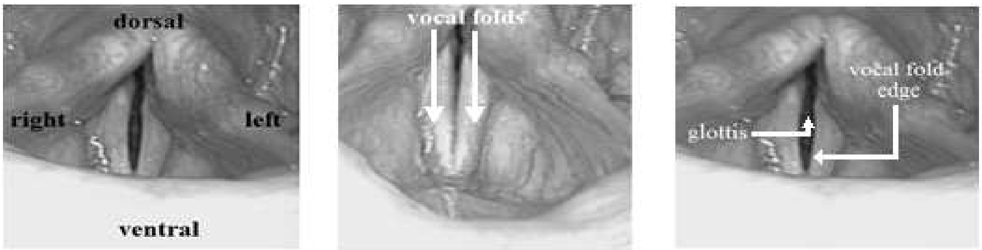


Figure 1. Videostroboscopic Images of the Vocal Folds at the Opening and Closing Stages of the Vocal Folds.

From the image processing point of view, the following basic problems are usually associated with the images taken:

- Sudden changes in the lighting intensity
- Movement of the patient at the moment of taking the image, which blurs the image

Therefore, given the problems cited above, segmentation of images without initialization is a difficult task.

In previous works in this area, segmentation algorithms based on region growing or thresholding [4]-[6] have been mostly used. The remaining approaches conducted on segmentation have been mostly based on edge detection algorithms and active contours models [7],[8]. Marendic et al.[9] proposed active contours with shape priors. However, this method is very time consuming and may not yield good results. Several automatic segmentation methods of the glottal space have also been previously presented which are based on region growing and active contours. The histogram algorithm is widely used for content-based image processing.[5],[10], A histogram is the graphical version of a table which shows what proportion of image intensity falls into each of several or many specified grayscale values. As the vocal fold image contains both high-intensity (surrounding tissue) and low-intensity (glottal gap) areas, then the vale in between the two can be treated as the threshold level. The threshold is determined by finding the first local minimum in the histogram, similarly to the method used by Wittenberg et al.[10] Based on the threshold derived from the histogram algorithm, we may separate the glottal gap from the surrounding tissues.

Due to its simplicity, the histogram algorithm has high efficiency in batch processing large amounts of image data. The histogram method, assuming that there is a fundamental difference between the lighting of pixels of the object and that of the pixels of the background, will give a good result. When the glottal is closing, or is completely closed, this algorithm will not be able to define a suitable threshold for the segmentation [14].

The active contour (snake) model introduced by Kass et al[15], is an automatic method for mage segmentation. The contour, or snake, is defined as a curve V(s) = (x(s),y(s)), s ∈ [0,1] that moves through the spatial domain of an image I(x,y) to minimize the energy function, [9]-[10], $E = \int 1/_2 \left(\alpha|vs|2 + \beta|vss|2\right) + Eext(v(s))$, where vs and vss represent the first and second derivatives of v with respect to s. Then the snake satisfies the following Euler quation, avss – bvssss - $\nabla$ VEext = 0, where α and β are the weights to control the snake's tension and rigidity, respectively. vssss denotes the fourth derivative of *v* with respect to s. We apply Eext from the image gray level f(x,y) that will take on smaller values at the boundaries. The final position of the contour will have a minimum value of energy E, so finding the object boundary becomes an energy minimization problem. After initializing a curve close to the object (or glottal gap) boundary, the snake, which is seen as an energy minimizing spline, starts deforming to fit the local minima so as to move toward the desired object boundary and finally settles on the object boundary. [9]-[10], [15]-[18].

Given the long time it needs to process the images, the active contours method also will not be suitable for cases where there is a large volume of images. Therefore, it is necessary to implement a new method to overcome the above – mentioned restrictions in segmenting video images of the larynx. Consequently, an automatic algorithm based on edge detection and the Hough transformation for extracting the glottal area from the video images is presented. This method consists of three stages of: preprocessing of the image, edge detection, and the Hough transformation.

This paper is divided in the following principal sections: section 2 describes the proposed algorithm, section 3 shows the obtained Results, and section 4 presents the the authors Conclusions.

## 2. THE PROPOSED METHOD

Block diagram of the proposed method is presented in Figure 2.

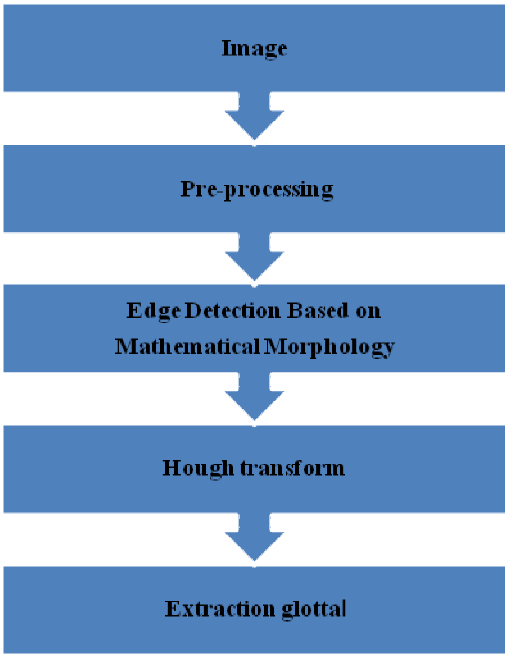


Figure 2. Block diagram of the proposed method.

The preprocessing stage was used to reduce noise and details of the image. In the next stage, Edge Detection Based on Mathematical Morphology was used to extract edges of the image. Then, the Hough transformation was used to find lines of the image. The lines of the glottal area were found and expanded by employing an innovative algorithm. Finally, the lines of the glottal area were extracted by using these lines. The details are presented in later sections.

### 2.1. Image Preprocessing

The images taken have poor quality due to the use of flash at the time of taking the images and also because of the limitations of the equipment. Therefore, to increase the accuracy of the segmentation, before doing this, an initial preprocessing was carried out on the image. The effect of the flash was marked by a circle in Figure 3a. The presence of these points will reduce the accuracy of the segmentation algorithm, which is based on edges. Therefore, non-linear mapping method was used to solve this problem.

$$I_{out}(x,y) = \begin{cases} 255 & for\ I(x,y) > Li \\ 255 \times \left(\frac{I(x,y)}{Li}\right) & for\ I(x,y) \le Li \end{cases} \quad (1)$$

$$L_i = \frac{1}{n(1.3)} \sum_{i=1}^{n} I(x_i, y)$$

Where I(x, y) denotes an stroboscopic image, L is the mean of lighting levels in each row of the image, n is the number of columns in the image and $\gamma$ is an image coefficient considered to be $\gamma$ = 1.8 in our implementation. As can be seen in Figure 3c and 3d, a large volume of unwanted information was omitted through the mapping relation 1.

### 2.2. Finding edges of the glottal

By omitting the noise lines, we use the edge detection algorithm with double thresholding to extract the edges of the glottal area [11].

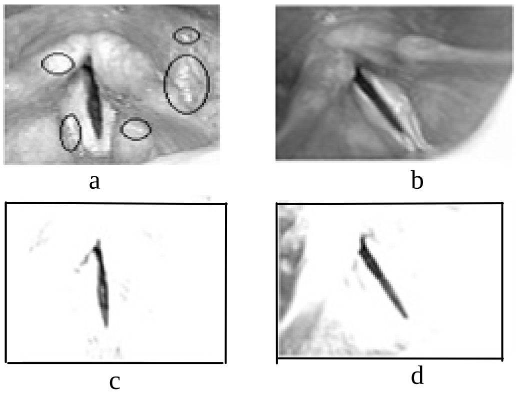


Figure 3. (a) and (b) are the original images, (c) and (d) show (a) and (b), respectively, preprocessed by relation 1. In image (a) the effect of the flash was marked by circle.

**2.2.1. Edge Detection**

Edge detection is one of the most fundamental stages in image segmentation. In edge detection based on differentiation, the differentiation process is directly used to detect edges. The first order differentiation operators, such as the Robert and Sobel operators, increase the extent to which the lighting in the images changes, and then, with the help of a threshold, they extract the edge points. Second order differentiation operators, such as the Laplacian operator, use their ability to pass through zeros to determine the edge points. The Prewitt and Canny edge detector algorithms also use the differentiation process, and certain algorithms as well, to detect the edges of the images; however, these edge detectors are not able to extract the edges in the dark parts of the image, in dimmed images, or in images which have noise.

Therefore, we use edge detection based on mathematical morphology in order to have strong edges.

**2.2.1. Base Morphological Operators**

Mathematical morphology is a mathematical theory which can be used to process and analyze images. According to this theory, images behave like sets, so that morphological transforms can be used to extract features of the images. A structural element is used in all morphological processing. This element has an important role in morphological processing; therefore its appropriate selection is often the most important part of a processing operation. The structural element usually moves across the image like a mask and controls the morphological processing.

Dilation and erosion are two of the base morphological operators and other morphological operators are often formed from a combination of these two. Assuming that I represents the two dimensional gray-scale image and B is a structural element , then the dilation of a gray- scale image I( x,y) with the structural element B(s,t) is determined as follows :

$$(I \oplus B)\ (x,y = \max \{I(x-s,y-t) + B(s,t)\} \quad (1)$$

The domain of $I \oplus B$ is the dilation of the domain of I by the domain of B.

The erosion of a gray-scale image I(x,y) with the structural element B(s,t) is specified as follows:

$$(I\ \Theta\ B)\ (x,y) = \min \{I(x+s,y+t) - b(s,t)\} \quad (2)$$

The domain of $I\ \Theta\ B$ is the erosion of the domain of I by the domain of B.

The opening of a gray -scale image I(x,y) with the structural element B(s,t) is specified as follows:

$$I\ o\ B = (I \Theta B) \oplus B \quad (3)$$

The opening operation usually smoothes the curves of the image, breaks narrow connections , and omits narrow protrusions .

The closing process of a gray surface image I(x,y) with the structural element B(s,t) is specified as follows:

$$I \bullet B = (I \oplus B)\, \Theta B \quad (4)$$

The closing operator also smoothes parts of the surrounding area, but, contrary to the opening operator, it usually merges narrow breaks, omits small holes, and fills the gaps surrounding the object. In this article, we have used the following method to detect the edges of glottal [12]:

$$M = (I \bullet B)\ o\ B \quad (5)$$

$$Eadg = (M \bullet B) \oplus B - (M \bullet B) \quad (6)$$

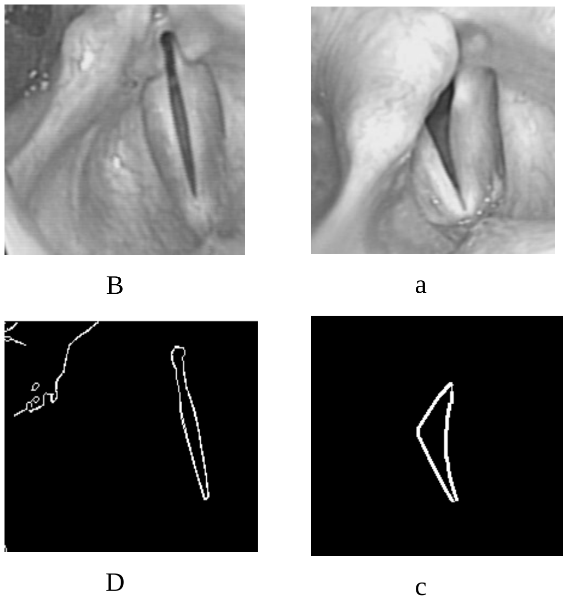


Figure 4. (a) and (b) the original image, (c) and (d) extracted edge.

Although the edge detection algorithm gives acceptable results in extracting the glottal area, there will be no guarantee that the extraction of the glottal area is correct since the quality of video images is low. For example, Figure (4–d), it can be seen that the extra lines are not produced by the edges of the glottal area .There is also this possibility that a part of the edge of the glottal will not be detected. Therefore, in later stages we will be using the Hough transformation in extracting the lines of the glottal in order to increase the accuracy and correctness of the process of extracting the glottal area.

### 2.3 Finding the initial lines

Hough proposed a method for identifying lines [12] in 1962 in which the line yi=axi+b is plotted for all the points in the space of the parameter (a,b) and the peak points (i.e., points in the parameter space through which many lines pass) are determined. Each of these points is an equivalent of the parameters of a line. To solve the problem of vertical lines, we consider the equation for the line as: $x.\cos\theta + y.\sin\theta = \rho$ and we search for peak points in the space of the parameters $\rho$ and $\theta$. In this method, the equation for each line is defined by the $\theta$ and $\rho$ specified for that line so that $\rho$ stands for the distance from the origin of coordinates to that line and $\theta$ is the angle between the line from the origin of coordinates which is perpendicular to the line under consideration and the positive direction of the x axis. At this stage, the edges extracted from the previous stage are transformed to a series of lines through the use of the Hough transformation.

#### 2.3.1. Extraction of the contours of the glottal area

In this stage, the purpose is to extract lines which separate the area of the glottal from the vocal folds (the edges of the vocal folds). To do this, we used an innovative method presented in this article. In this method, the feature of the image itself, which is the difference in the level of lighting in the glottal area and that of the vocal folds, was used to find the lines related to the glottal. In the images of the larynx, the level of lighting of the glottal area is much lower than that of the vocal folds. Therefore, assuming that the purpose was to find the lines of the edge of the vocal fold on the right, we will witness a sharp decline in the levels of

ISSN: 2088-8708

lighting when we move from the center of the line under consideration in the direction of the horizontal axis. This feature can be used to find the lines of right and left vocal

For ( i = 1 to number of lines)
F(x,y) = Mid($line_i$)
$R_{ave} = \frac{1}{n} \cdot \sum_{j=1}^{n=3} F(x, y+j)$
$L_{ave} = \frac{1}{n} \cdot \sum_{j=1}^{n=3} F(x, y-j)$
$Dif_r$ = Intensity(F(x,y)) – $R_{ave}$
$Dif_l$ = Intensity(F(x,y)) – $L_{ave}$
If ( ($Dif_r$ > T) and ($Dif_l$ <= 0) or ($Dif_r$ > T) and ($Dif_l$ <= 0) ) Then
Add $line_i$ to newlines{}
Else
Remove $line_i$
End of for

Figure 5. The pseudo - code for extracting the lines of the vocal folds.

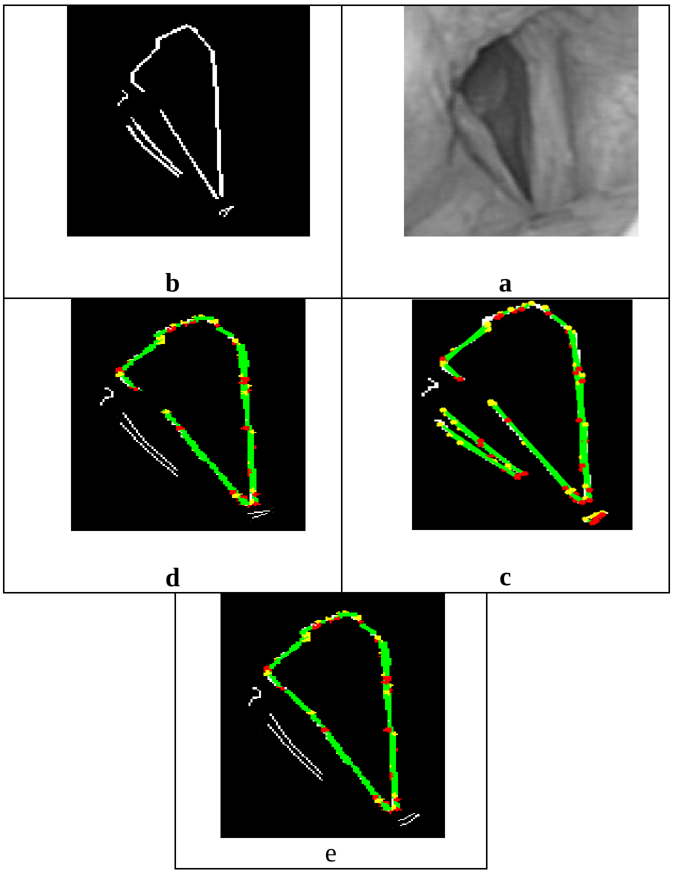


Figure 6. (a) the main image , (b) the edges extracted from the image (a) , (c) the production of lines related to the edges, the image (d) , detection of the lines related to boundaries of glottal by the pseudo code offered and the image (e) , the production of lines not produced due to whatever reason.

In Figure 5, F(x,y) is the center of the ith line in the coordinate axes, Lave and Rave define the means of the levels of the lighting of the three pixels to the right and left of F(x,y), Difr and Difl respectively. Results obtained from this pseudo-code are shown in Figure 6-b.

This method increases the accuracy and correctness of the extraction of the glottal area by omitting the unnecessary lines. In Figure 6-a, the main image is shown along with the unnecessary lines, which are omitted by the innovative algorithm mentioned above (Figure 6-b). Finally, in order to have connected lines, they were merged by considering the distance between them and by searching for lines which were least different from each other (Figure 6-c). It must be mentioned that the empty spaces between lines were connected to each other with suitable estimations.

## 3. EXPERIMENTAL RESULTS

Our proposed algorithm has been selected and tested on 1540 images taken from 93 video films. These video images were taken with a stroboscopic camera having a resolution of 640 × 480 pixels for each image at the Amir-Aalam Hospital in Tehran. Details of the gathered data set are presented in Table 1.

Table 1. Data (video images) analyzed by doctors based on type of the laryngeal mass.

| | polyp | nodule | Paresis | Healthy | Total |
|---|---|---|---|---|---|
| Female | 14 | 11 | 9 | 19 | 53 |
| Male | 7 | 4 | 7 | 11 | 40 |

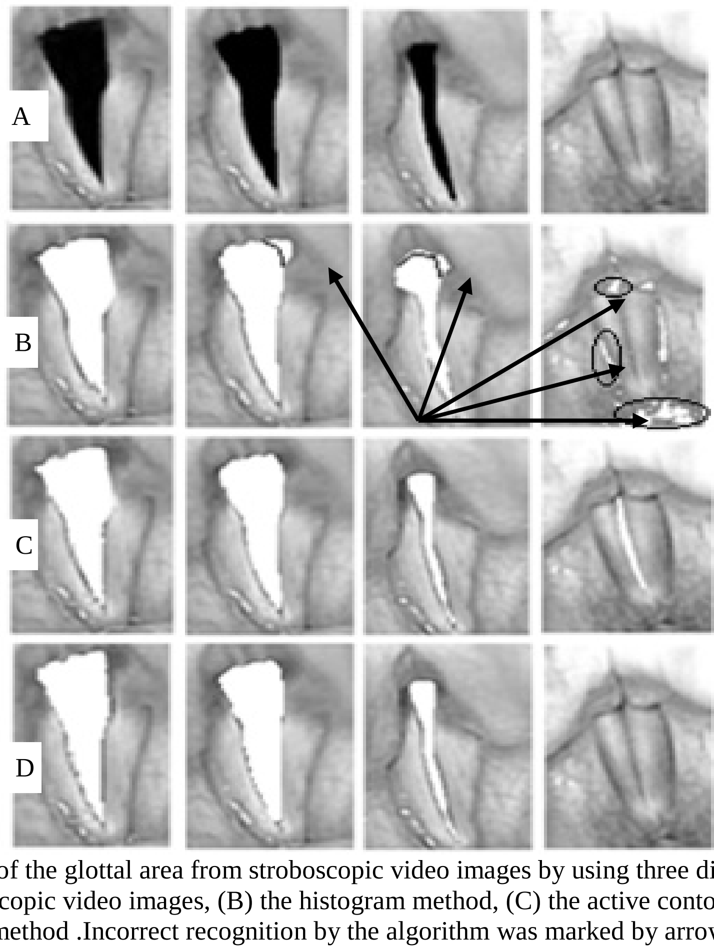


Figure 8. recognition of the glottal area from stroboscopic video images by using three different methods: (A) a sequence of stroboscopic video images, (B) the histogram method, (C) the active contours method, and (D) our automatic method .Incorrect recognition by the algorithm was marked by arrows and circles.

To show the efficiency of the proposed algorithm in extracting the glottal area, it was compared with the two popular methods of histogram and active contours. Figure 8-a shows four consecutive frames of the videostroboscopic images at the time when the vocal folds are close or open. Figure 8-b, 8-c and 8-d show the results obtained by using the histogram method, the active contours method, and our proposed method respectively. The space extracted by this algorithm was marked white. As it can be seen in Figure 8-b, the histogram method, with the assumption that there is a basic difference between the levels of lighting of the pixels of the object and that of the pixels of the background, will yield a suitable result. However, at the time the glottal is near to being closed, or it is completely closed, this algorithm cannot define a suitable threshold for the segmentation process. The active contours method has been designed for recognizing the contours of one and only one object (glottal) while the vocal folds may contain no objects (closing stage of the glottal) but contain several objects if there are nodules on the vocal folds (Figure 9).

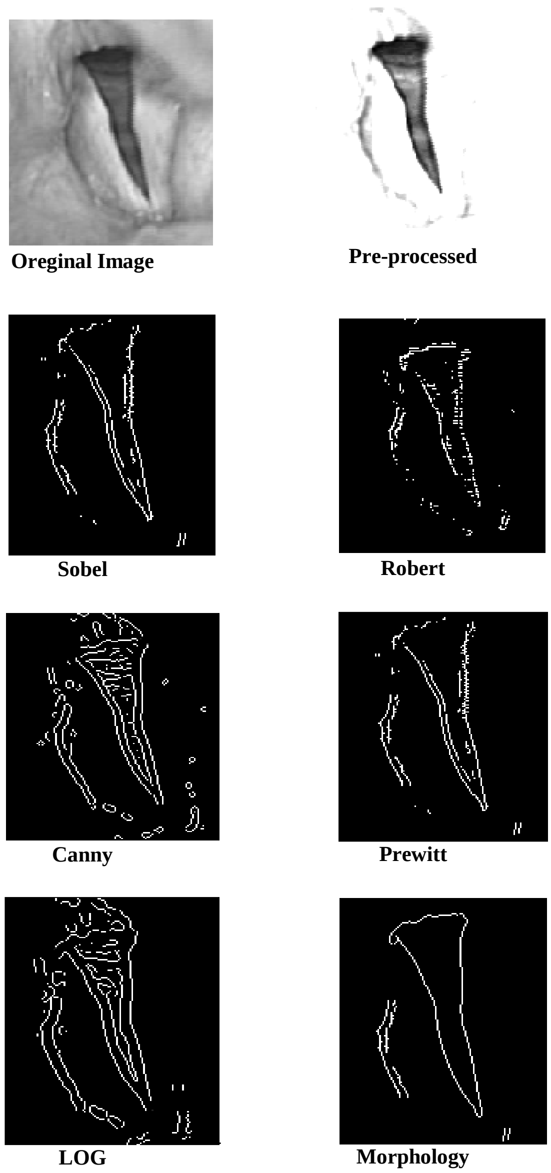


Figure.9. Results of detecting edges by using the Sobel, Robert, Prewitt, Canny, Laplacian , and the morphological methods.

Furthermore, active contours are very sensitive to initial areas and to parameters, and if these parameters are not correct, this method will be faced with difficulties. In comparing these two methods, extracting the glottal area with our proposed method enjoys very high accuracy and correctness. Automatic methods of recognizing the glottal area by using video images should enjoy high computational accuracy. In this article, we have used the following error function to assess the accuracy of the algorithm for extracting the glottal area [13]:

$$\Delta = \frac{\sqrt{\frac{1}{T}\int_0^T \left[a(t) - as(t)\right]^2 dt}}{\sqrt{\frac{1}{T}\int_0^T as(t)^2\, dt}} \quad (7)$$

Where T is the number of frames; as (t), the extracted space computed by using the manual methods; and a (t), the extracted space computed by using histogram algorithms and our proposed method. This experiment was conducted on 10 films taken of 10 different cases, with 20 frames taken at random of each larynx when it was closing and when it was opening. The mean errors for the 10 cases in the histogram, active contours, and our proposed method were 35.3%, 12.6%, and 6.7%, respectively.

Given the above results, the histogram method had the most errors. This method is useful when there is a basic difference between the lighting of the glottal area and other spaces. This difference does not exist when the glottal is closing, and this factor causes an increase in errors. Active contours algorithms have been designed to recognize contours of a single object (glottal) in an image. In the closing and opening phrase of the glottal, when there are no objects (when the glottal is closed) and states when there will be several objects (for example when there are nodules on the vocal folds), This state and sensitivity to the initial area will increase the number of errors. Our proposed method, compared to the histogram and the active contours methods will yield much more suitable results as it does not need parameter adjustments nor does it require high quality images.

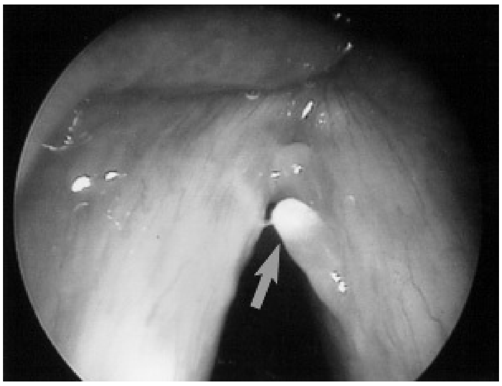

Figure.10. An example of a polyp in a larynx.

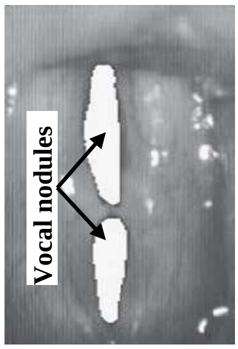


Figure. 11. Vocal folds with nodules.

## 4. CONCLUSIONS

In this article, we have offered an automatic method for extracting the glottal area. In this method, the edges are produced by the morphological operators and are turned into a number of lines with the help of the hough transform and then, from among these lines, those related to the glottal area are found and extracted by an innovative algorithm presented in this article. Our method, which is proposed in this article, is compared with the histogram and active contours methods regarding accuracy of extracting the glottal area. The results obtained in our method are much better than those of other methods. In particular, this method is capable of extracting vocal fold vibrations at low image resolutions. The proposed automatic image-processing algorithm may provide a valuable tools for analyzing thousand of videostroboscopy images and offer a valuable biomedical application for the clinical diagnosis of laryngeal pathologies.

## REFERENCES


[1] S.Goldman, J. Hargrave, ’’Stress, Anxiety, Somatic Complaints, and Voice Use in Women with Vocal Nodules’’. *American Journal of Speech-Language Pathology,* Vol. 5, pp. 44-54 February 1996.

[2] D. Deliyski, et al, ‘’ Clinical Implementation of Laryngeal High-Speed Videoendoscopy’’ Challenges and Evolution. *Folia Phoniatr Logop*; 60: 33-44, 2008.

[3] B.Poburka, ’’A New Stroboscopy Rating Form’’ *Jornal of voice,* Vol. 13, No. 3, pp. 403-413.1999.

[4] J. Lohscheller, H. Toy, ’’clinically evaluated procedure for the reconstraction of vocal fold vibrations from endoscopic digital high-speed videos’’ *Jornal Med. Im. Anal.* Vol 11, pp. 400-413, 2007.
[5] Y. Yan, X. Chan, ’’automatic tracing of vocal fold motion from high-speed digital images. *IEEE trans. Biomed. Eng*, vol 53, no. 7, pp. 1394-1400, 2006.
[6] D. Voigt, M. DÖllinger, ‘’ Automatic diagnosis of vocal fold paresis by employing phonovibrogram features and machine learning methods’’ *computer methods and programs in biomedicine*, COMM-3010; No. of Pages 14, 2010.
[7] I. School, A. Sovakar, ’’Motion analysis of vocal folds using adaptive snakes’’, *IEEE trans. Biomed. Eng*, vol 44, no 8, pp. 1221-1228, 1997.
[8] H. Larsson, ’’ vocal fold vibrations high speed imaging, kymography and acoustic analysis’’ *Laryngoscope*, Vol 110, no 12, pp. 2117-2122, 2000.
[9] B. Marendic, N.Galatsanos *’’A new active contour algorithm for tracking vocal fold’’ proc.IEEE int. Conf. On image processing*, pp.397-400, 2001.
[10] Wittenberg T, Moser M, Tigges M, Eysholdt U. Recording, processing, and analysis of digital high-speed sequences in glottography. Mach Vis Applicat. 1995;8:399–404
[11] J. Canny, ’’A computational approach *IEEE Pattern Anal Machine Intell, no* 8, pp. 679–698, 1986.
[12] J. A. R. Artolazabal J. Illingworth” *LIGHT: Local Invariant Generalized Hough Transform” Proc. of 18th International Conference on Pattern Recognition (ICPR 2006),*Vol 3, pp. 304-307, 2006.
[13] Y. Zhang, E. Bieging ‘’Efficient and Effective Extraction of Vocal Fold Vibratory Patterns from High-Speed Digital Imaging’’ *Journal of Voice*, Vol. 24, No. 1, pp. 21-29, 2010.
[14] J. Lohscheller, M. Dollinger ‘’Quantitative investigation of the vibratory pattern of the substitute voice generator’’ IEEE *Trans Biomed Eng*, no 51, pp. 1394–1400, 2004.
[15] Kass M, Witkin A, Terzopoulos D. Snakes: active contour models. Int J Comput Vision. 1988;1:321–331.
[16] Xu J, Chutatape O, Chew P. Automated optic disc boundary detection by modified active contour model. IEEE Trans Biomed Eng. 2007;54:473–482.
[17] Shan ZY, Ji Q, Gajjar A, Reddick WE. A knowledge-guided active contour method of segmentation of cerebella on MR images of pediatric patients with medulloblastoma. J Mag Res Imag. 2005;21:1–11.
[18] Yezzi A, Kichenassamy S, Kumar A, Olver R, Tannenbaum A. A geometric snake model for segmentation of medical imagery. IEEE Trans Med Image. 2006;16:199–209.

## BIOGRAPHIES OF AUTHORS

**Davood Aghlmandi** was born in Ardebil, Iran in 1980. He received the B.S. degree in computer engineering from Azad university of Lahijan, Lahijan, Iran, in 2007 and the M.Sc. degree in Artificial intelligence from Azad University of Qazvin, Qazvin, Iran in 2011.
His research interests include Biometrics Recognition and authentication, Pattern Recognition, Image Processing, Neural Networks.

**Karim Faez** Was born in Semnan, Iran. He received his BSc. degree in Electrical Engineering from Tehran Polytechnic University as the first rank in June 1973, and his MSc. and Ph.D. degrees in Computer Science from University of California at Los Angeles (UCLA) in 1977 and 1980 respectively.
Professor Faez was with Iran Telecommunication Research Center (1981-1983) before Joining Amirkabir University of Technology (Tehran Polytechnic) in Iran in March 1983, where he holds the rank of Professor in the Electrical Engineering Department. He was the founder of the Computer Engineering Department of Amirkabir University in 1989 and he has served as the first chairman during April 1989-Sept. 1992.
Professor Faez was the chairman of planning committee for Computer Engineering and Computer Science of Ministry of Science, Research and Technology (during 1988-1996).
His research interests are in Biometrics Recognition and authentication, Pattern Recognition, Image Processing, Neural Networks, Signal Processing, Farsi Handwritten Processing, Earthquake Signal Processing, Fault Tolerance System Design, Computer Networks, and Hardware Design.
Dr. Faez coauthored a book in Logic Circuits published by Amirkabir University Press. He also coauthored a chapter in the book: *Recent Advances in Simulated Evolution and Learning*, Advances in Natural Computation, Vol. 2, Aug.2004,World Scientific. He published about 300 articles in the above area. He is a member of IEEE, IEICE, and ACM, a member of Editorial Committee of Journal of Iranian Association of Electrical and Electronics Engineers, and International Journal of Communication Engineering. Emails: kfaez@aut.ac.ir, kfaez@ieee.org, kfaez@m.ieice.org.