A Study of Classification System Vocal Cord Disease Using Digital Image Processing

S.Harini
K.Ramakrishnan College Of Technology,Trichy,India

Abstract- *This paper presents the study of vocal cords disease classification using digital image processing. Stroboscopy and laryngoscopy are common tools used by doctors to observe the state of the vocal cords directly. The vocal cords disease can be characterized by characters changing the shape of glottis contour on the vocal cords. There are six classifications of vocal cords, normal, paralysis, nodule, papilloma, cyst,and granuloma which are present in this paper. Before the classification process, extraction image vocal cords to get the characteristics or information of objects in the image. More than half of patients presenting with hoarseness show benign vocal fold changes. The clinician should be familiar with the anatomy, physiology and functional aspects of voice disorders and also the modern diagnostic and therapeutic possibilities in order to ensure an optimal and patient specific management. This paper presents the study of vocal cords disease classification using digital image processing. - parametric technique. Proposes an Artificial neural network for Vocal card disease classification. Before the classification process, extraction image vocal cords to get the characteristics or information of objects in the image.*

I. INTRODUCTION

Nowadays, medical diagnosis and treatment in the vocal cords disease only depend on subjective analysis of images taken by a laryngoscopy device. Laryngoscopy be used by medical personnel to see the condition of the patient';s
vocal cords. The vocal cords are very influential in the production by the human voice that is used to communicate. Conditions of the vocal cords can experience temporary or permanent physiological disorders that result in changes in the sound produced. So if vocal cords disease can interfere with daily activities in communication. The results of the analysis of the diagnosis of each doctor can vary depending on the expertise and experience of the doctor who performs the laryngoscopy procedure. In general, examines the function and biomechanism of vocal cords and laryngoscopy examining the anatomy of the vocal cords. The laryngoscopy is only at the stage of observing the vocal cords directly displayed on the PC screen without any image storage. Based on the background, this study aims to help doctors detect and evaluate vocal cords disease by using image-based processing technology. The system can be a reference for doctors in terms of diagnosis of abnormalities in the vocal cords. In a previous study in 2016 regarding the detection of vocal cord disease with the Moore Neighbor Tracing method of image processing technology obtained an accuracy value of 85.83% of 120 data tested and can classify abnormalities in vocal cords, Paralysis, Papilloma, Granuloma, Nodules / Cyst. The study obtained the glottis contour of the vocal cords with the Moore Neighbor Tracing method. Before getting the glottis contour of the vocal cords, an image is processed into a binary image with black as the glottis area and white as the background. The coordinates of the resulting glottis contour form the basis of the parameter classification of abnormalities in the vocal cords. In this research, image processing is still done manually by users such as rotation, fitting, and multiple grayscale processes.With optimize glottis contour extraction algorithm this study can automatically obtain the glottis area and it is possible to obtain shape measurement to extract contour that can be analysed and classified. To obtain a fast, robust, and automatic vocal cords extraction the local feature description are been applied to get more characteristic of vocal cord. The SURF algorithm process can quickly detect key point descriptors because SURF combines two processes, namely integral blob detection image so that it will speed up computing in the process.

II. METHODS

A.VOCAL CORDS:

The vocal cords, also known as vocal folds or voice reeds are folds of tissue in the throat that are key in creating sounds through vocalization. The size of vocal cords affects the pitch of voice. Open when breathing and vibrating for speech or singing, the folds are controlled via the recurrent laryngeal branch of the vagus nerve. They are composed of twin infoldings of mucous membrane stretched horizontally, from back to front, across the larynx. They vibrate, modulating the flow of air being expelled from the lungs during phonation. One of two small bands of muscle within the larynx. These muscles vibrate to produce the voice. The vocal cords form a "V" inside the larynx, a 2-inch-long, tube-shaped organ in the neck.

B.Vocal Cords Disease::

Vocal cord disease can be caused by various factors. Several types of diseases, namely nodules, cysts, granulomas, papilloma, and paralysis can change the shape of the glottis contour of the vocal cords. The glottis contour is the edge of the inner vocal cords.

C.Structure:

The vocal cords are composed of twin infoldings of 3 distinct tissue. The the outer layer is squamous, non-keratinizing epithelium. Below this is the superficial layer of the lamina propria, a gel like layer, which allows the vocal fold to vibrate and produce sound. The vocalis and thyroarytenoid muscles make up the deepest portion. These vocal folds are covered with a mucous membrane and are stretched horizontally, from back to front, across the larynx. The vocal folds are located within the larynx at the top of the trachea. They are attached posteriorly to the arytenoid cartilages, and anteriorly to the thyroid cartilage. They are part of the glottis which includes the rima glottidis. Their outer edges are attached to muscle in the larynx while their inner edges, or margins, are free forming the opening called the rima glottidis. They are constructed from epithelium, but they have a few muscle fibres in them, namely the vocalis muscle which tightens the front part of the ligament near to the thyroid cartilage. They are flat triangular bands and are pearly white in color. Above both sides of the glottis are the two vestibular folds or false vocal folds which have a small sac between them. Situated above the larynx, the epiglottis acts as a flap which closes off the trachea during the act of swallowing to direct food into the esophagus. If food or liquid does enter the trachea and contacts the vocal folds it causes a cough reflex to expel the matter in order to prevent pulmonary aspiration.

D. Chan-Vese Algorithm:

The Chan-Vese Algorithm. Segmentation is the process of partitioning a digital image into multiple segments (sets of pixels). ... This model is based on the Mumford-Shah functional for segmentation, and is used widely in the medical imaging field, especially for the segmentation of the brain, heart and trachea. The level set based Chan-Vese algorithm primarily uses region information for successive evolutions of active contours of concern towards the object of interest and, in the process, aims to minimize the fitness energy functional associated with. Orthodox gradient descent methods have been popular in solving such optimization problems but they suffer from the lacuna of getting stuck in local minima and often demand a prohibited time to converge. This work presents a Chan-Vese model with a modified gradient descent search procedure, called the Delta-Bar-Delta learning algorithm, which helps to achieve reduced sensitivity for local minima and can achieve increased convergence rate. Simulation results show that the proposed search algorithm in conjunction with the Chan-Vese model outperforms traditional gradient descent and recently proposed other adaptation algorithms in this context.

E. K-Nearest Neighbor:

K nearest neighbors is a simple algorithm that stores all available cases and classifies new cases based on a similarity measure (e.g., distance functions). KNNhas been used in statistical estimation and pattern recognition already in the beginning of 1970;s as a non-parametric technique. In pattern recognition, the k-nearest neighbors algorithm (k-NN) is a non-parametric method used for classification and regression. In both cases, the input consists of the k closest training examples in the feature space. The output depends on whether k-NN is used for classification or regression: In k-NN classification, the output is a class membership. An object is classified by a plurality vote of its neighbors, with the object being assigned to the class most common among its k nearest neighbors (k is a positive integer, typically small). If k = 1, then the object is simply assigned to the class of that single nearest neighbor. In k-NN regression, the output is the property value for the object. This value is the average of the values of k nearest neighbors. k-NN is a type of instance-based learning, or lazy learning, where the function is only approximated locally and all computation is deferred until function evaluation. Both for classification and regression, a useful technique can be to assign weights to the contributions of the neighbors, so that the nearer neighbors contribute more to the average than the more distant ones. For example, a common weighting scheme consists in giving each neighbor a weight of 1/d, where d is the distance to the neighbor. The neighbors are taken from a set of objects for which the class (for k-NN classification) or the object property value (for k-NN regression) is known. This can be thought of as the training set for the algorithm, though no explicit training step is required. A peculiarity of the k-NN algorithm is that it is sensitive to the local structure of the data.

F. Shape Measurement:

Surface shape is the overall geometry of the area of interest, where the area of interest can vary according to the application. For example, if you were measuring the topography of a micro-lens, “shape” might be described as a measurement of the lens’;s curvature, as compared to a manufacturing specification. Even if we understand the Convolution Neural Network theoretically, quite of us still get confused about its input and output shapes while fitting the

data to the network. This guide will help you understand the Input and Output shapes for the Convolution Neural Network. Let's see how the input shape looks like. The input data to CNN will look like the following picture. We are assuming that our data is a collection of images.

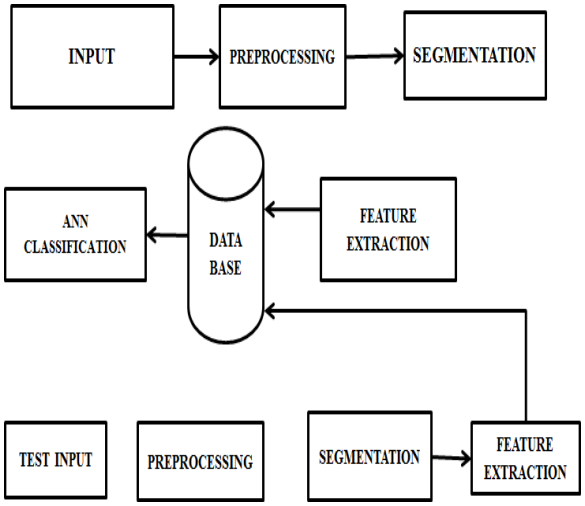
G.Input Shape:

You always have to give a 4D array as input to the CNN. So input data has a shape of (batch size, height, width, depth), where the first dimension represents the batch size of the image and other three dimensions represent dimensions of the image which are height, width and depth. For some of you who are wondering what is the depth of the image, it's nothing but the number of colour channel. For example, RGB image would have a depth of 3 and the greyscale image would have a depth of 1.

ORIGINAL IMAGE SHAPE

RESIZED IMAGE SHAPE

III. SYSTEM DESIGN



BLOCK DIAGRAM

RANGES:

*rgb2gray converts RGB values to gray scale values by forming a weighted sum of the R, G, and B components:0.2989 * R + 0.5870 * G + 0.1140 * B*

*0.299 * R + 0.587 * G + 0.114 * B*

The image is resize into the [0 255] range is taken for resize the image pixel. Historgram provide the number of pixel representation of image.

RED CHANNEL IMAGE

GREEN CHANNEL IMAGE

GREEN CHANNEL MEDIAN FILTER IMAGE

TECHNIQUES:

1. PREPROCESSING:

*It is also known as an intensity, gray scale, or gray level image. Array of class uint8, uint16, int16, single, or double whose pixel values specify intensity value*s.

Image formation using sensor and other image acquisition equipment denote the brightness or intensity I of the light of an image as two dimensional continuous function F(x, y) where (x, y) denotes the spatial coordinates when only the brightness of light is considered. Sometimes three-dimensional spatial coordinate are used. Image involving only intensity are called gray scale images

2. BACKGROUND IMAGE ANALYSIS:

Background subtraction, also known as Foreground Detection, is a technique in the fields of image processing and computer vision wherein an image's foreground is extracted for further processing (object recognition etc.). K-means (FCM) algorithm is a clustering algorithm developed by Dunn, and later on improved. It is useful when the required number of clusters are pre-determined; thus, the algorithm tries to put each of the data points to one of the clusters.

3. K means Segmentation:

K-means (FCM) algorithm is a clustering algorithm developed by Dunn, and later on improved. It is useful when the required number of clusters are predetermined; thus, the algorithm tries to put each of the data points to one of the clusters.

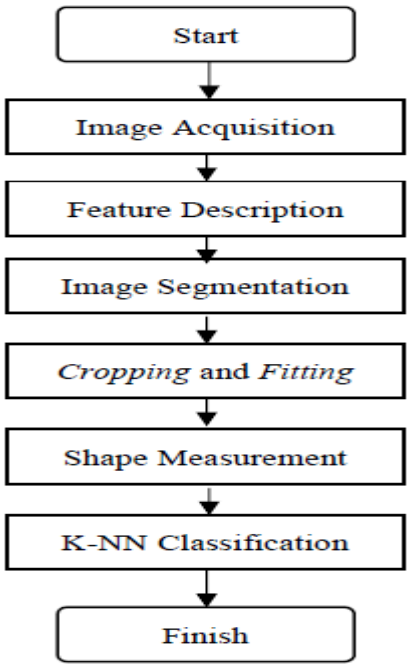
4. Feature extraction :

Detection and localization take place at two different levels. First, a side-view image of the crop row is taken by the cameras on the sensor rig. The color image of the camera mounted next to the camera is pixel-wise registered to the image. As a result, the color data match its corresponding 3D measurements from the camera.

5. Classification:

Ann classifier is used for classification. The dataset classification uses the most effective tool called back propagation neural network. The usage of Back Propagation Neural Network (BPNN) for classifying the images The neural network is given the target outputs on to which it should map its inputs, i.e. it is given in paired data of input and output.

IV. FLOW DIAGRAM OF PROPOSED SYSTEM



A. Image Acquisition:

Image acquisition in this study is the process of capturing analog images to digital images. This step is called the initial processing stage. Vocal cord images obtained from previous research and medical personnel documentation using laryngoscopy.

B. Image Segmentation :

The next step is image segmentation using the Chan-Vese algorithm. The segmentation process aims to obtain a binary image in black is the image background and white is the contour of the glottis area.

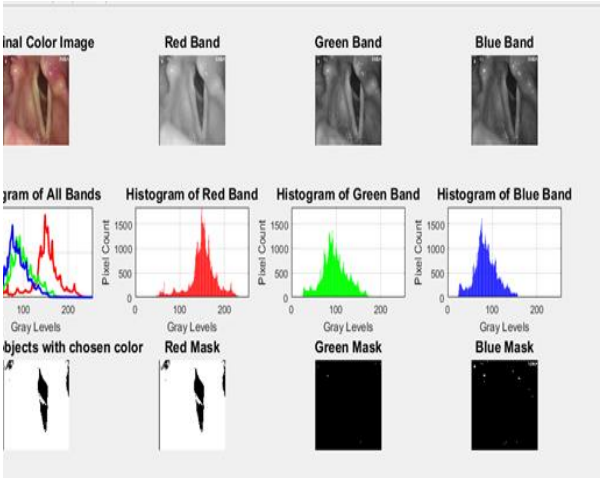
C.Cropping and Fitting:

Image cropping and fitting processes are needed for feature extraction processes. With the cropping process, the contour of the vocal cords of different sizes in each image will be the same and the process of fitting the glottic contour which was not in the center of the frame will be centered by shifting the midpoint of the object to the midpoint of the frame. This process is useful so that the image has the same size and shape reference.

D.E Shape Measurement:

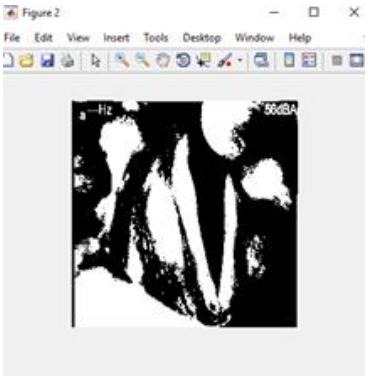
The next step is feature extraction from the binary contour glottis area. Feature extraction can obtain parameter values as a reference for determining the vocal cord disease. The glottis contour area is measured by the area so that a value with a certain range will be obtained. The range of values will be a parameter in classifying vocal cord images.

IV. RESULT



HISTOGRAM AND MASK OF DIFFERENT BAND IMAGES

The output of the CNN is also a 4D array. Where batch size would be the same as input batch size but the other 3 dimensions of the image might change depending upon the values of filter, kernel size and padding we use.



V. CONCLUSION

The classification system in vocal cord disease can produce 96.7% accuracy with 2 image data that affect classification errors, cyst and papiloma. The system can recognize the condition of the vocal cords with nodule, normal, granuloma dan paralysis conditions that produce 100% similarity with training data. The image processing process in this system is automatic so that it is easier to use compared to the image processing process in previous studies with moore neighbor. The processing process such as cropping and fitting is automatic and it detects the glottis contour of the vocal cords without any initialization from the user. Extraction parameters consist of two, feature description parameters and the glottis contour shape measurement parameters The classification system for the diagnosis of vocal cord conditions is the research stage of developing the application of disease diagnosis on the vocal cords using digital image processing.

VI. ACKNOWLEDGMENT

The author wish to acknowledgment the Dustira Hospital Cima which kindly lent material for this study.

REFERENCES

[1] B. G. L. PUBIYANGGA, “IDENTIFIKASI KONDISI PITA SUARA UNTUK DETEKSI KELAINAN PITA SUARA DENGAN METODE MOORE NEIGHBOR TRACING.” Universitas Telkom, 2016.
[2] R. Munir, “Bab 11_Citra Biner,” p. 169, 2004.

[3] T. F. Chan and L. A. Vese, "Active contours without edges," IEEE Trans. Image Process., vol. 10, no. 2, pp. 266–277, 2001.
[4] A. M. Zorrilla and B. G. Zapirain, "Vocal folds paralysis study using a pre-processing stage of Gabor filtering and Chan-Vese segmentation," 2010 IEEE Int. Symp. Signal Process. Inf. Technol. ISSPIT 2010, pp. 360–365, 2011.
[5] S. Z. Karakozoglou, N. Henrich, C. D'Alessandro, and Y. Stylianou, "Automatic glottal segmentation using local-based active contours and application to glottovibrography," Speech Commun., vol. 54, no. 5, pp. 641–654, 2012.
[6] H. Bay, A. Ess, T. Tuytelaars, and L. Vangool, "Speeded-Up Robust Features (SURF) (Cited by: 2272)," Comput. Vis. Image Underst., vol. 110, no. 3, pp. 346–359, 2008.
[7] G. Amato and F. Falchi, "kNN based image classification relying on local feature similarity," no. January, p. 101,2010
[8] T. F. Chan and L. A., Vese,Active contours without edges.,2001..
[9] A.Méndez Zorrilla,"Vocal folds paralysis study using a pre-processing stage of gabor filtering and chan-vese segmentation".,no vol.10,2017.
[10] Laura Verde., Voice Disorder Identification by Using Machine Learning Techniques"pp.590-599.,2017.