Arytenoid Cartilage Feature Point Detection Using Laryngeal 3D CT Images in Parkinson's Disease

Nandakishor Desai[1], Aravinda S. Rao[1], Paari Palaniswami[2], Dominic Thyagarajan[2]
and Marimuthu Palaniswami[1]

Abstract— **Parkinson's disease is a neurodegenerative disorder that results in progressive degeneration of nerve cells. It is generally associated with the deficiency of dopamine, a neurotransmitter involved in motor control of humans and thus affects the motor system. This results in abnormal vocal fold movements in majority of the Parkinson's patients. Analysis of vocal fold abnormalities may provide useful information to assess the progress of Parkinson's disease. This is accomplished by measuring the distance between the arytenoid cartilages during phonation. In order to automate this process of identifying arytenoid cartilages from CT images, in this work, a rule-based approach is proposed to detect the arytenoid cartilage feature points on either side of the airway. The proposed technique detects feature points by localizing the anterior commissure and analyzing airway boundary pixels to select the optimal feature point based on detected pixels. The proposed approach achieved 83.33% accuracy in estimating clinically-relevant feature points, making the approach suitable for automated feature point detection. To the best of our knowledge, this is the first such approach to detect arytenoid cartilage feature points using laryngeal 3D CT images.**

I. INTRODUCTION

Parkinson's disease is a neurodegenerative disorder that results in progressive degeneration of nerve cells. It is generally associated with the death of neurons in substantia nigra that results in deficiency of dopamine, affecting the motor system. An estimated 10 million people suffer from Parkinson's disease in the world [1] and a significant proportion of 70-89% of the Parkinson's patients suffer vocal fold abnormalities [2]–[4]. The vocal folds are a pair of muscle bands covered by mucous membrane that appear horizontally across the larynx and they come together (adduction) during phonation and move apart (abduction) during respiration. Some of the existing commonly known techniques to assess the vocal fold dynamics are laryngeal electromyography, laryngeal endoscopy, and laryngeal stroboscopy [5]. However, most of them are invasive, causing significant patient discomfort, fail to capture the 3D movements of the vocal folds and are subjective.

With rapid advancements in medical imaging technology, Computed Tomography (CT) has become a reliable and non-invasive modality to image the human organs. CT scanners

[1]N. Desai, A. S. Rao, and M. Palaniswami are with the Department of Electrical and Electronic Engineering, The University of Melbourne, VIC - 3010, Australia. nandakishord@student.unimelb.edu.au, {aravinda.rao, palani}@unimelb.edu.au

[2]P. Palaniswami and D. Thyagarajan are with the Monash Medical Centre, 246 Clayton Rd, Clayton, VIC - 3168, Australia. pmpal4@student.monash.edu, dominic.thyagarajan@monash.edu.au

with high temporal resolution capture the rapid fluctuations in vocal fold movements during phonation. Distance between arytenoid cartilages on these CT images has been used as a feature to assess the hypokinetic nature of vocal fold movement in cases of Parkinson's disease [6]. However, manual identification of arytenoids as in [6], is a highly subjective and observer dependent task, making it error prone. Therefore, an automated technique is required to analyze CT images to provide reliable and consistent scores across subjects and observers. However, the development of such an automated technique involves challenges due to movement of multiple structures in the frame, as phonation is a dynamic activity involving the passage of air, which creates different appearances on CT images. In addition, CT images are of low resolution compared with other imaging modalities such as MRI, making it more difficult to distinguish cartilages.

The work in [7] is the first known approach that aims to develop an automated technique to extract vocal fold planes from neck 3D CT images. It is assumed that there is no coronal tilt during image acquisition and hence the vertebral column is approximately orthogonal to the plane of vocal folds. The anterior intersection point of vocal folds is determined and the orthogonal plane to the vertebral column passing through the Anterior Commissure (AC) is identified as the plane of vocal folds. However, it is possible that the vertebral columns are not always orthogonal to the vocal folds due to the anatomical differences or due to a coronal tilt in the data. In addition, it does not provide any additional information apart from determining the location of AC. This forms the motivation for our work to develop a rule-based approach to detect clinically-relevant arytenoid cartilage feature points.

II. PROPOSED APPROACH

We define feature points to be the most anterior points on the arytenoid cartilages converging towards the airway. This helps to determine points of interest from the arytenoids. Distance between them is generally used to assess the vocal fold abnormalities during phonation [6]. Steps involved in the detection of feature points are outlined in Fig. 1 and are described in the following subsections.

A. Dataset

Subjects were recruited from movement disorder clinic at the Monash Medical Center, Australia [6]. All the subjects were in age group of 50-90 years and some of them had Parkinson's disease for less than 6 years, whilst others did
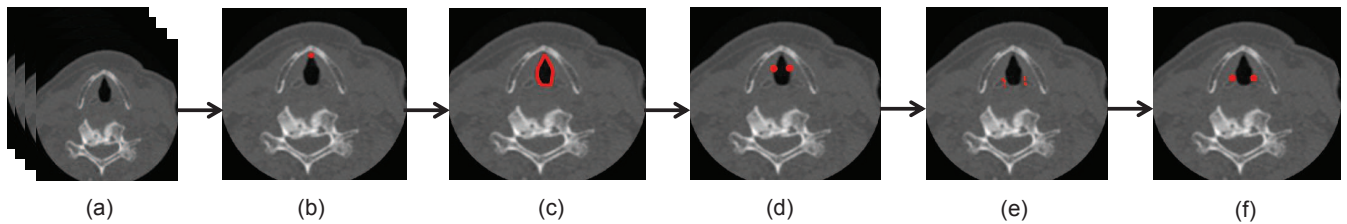
Fig. 1: Flow diagram of the proposed work for the detection of feature points: (a) axial plane of the 3D CT Data, (b) localization of anterior commissure on the axial slice, (c) detecting the airway boundary, (d) cutoff pixels to perform filtering, (e) detecting potential feature points, (f) final feature points after post processing.

not have any neurological or laryngeal disorders. All the subjects were asked to make five short and fast /i/ phonations and using a 320-CT scanner, the entire neck was imaged. Resulting images were converted into NIfTI-1 images with 512×512 pixels size and 12-bit graylevels with Right, Anterior, Superior (RAS) orientation. Subjects with blurred and distorted images were excluded.

B. Anterior Commissure (AC) Localization

AC is defined to be the junction of the two vocal folds in the anterior portion of the larynx. Locating AC from 3D CT volumes of the subjects forms the basic initialization step. Location of AC acts as a starting point to scan for arytenoid cartilages converging towards the airway on either side, which helps in detecting the feature points. We use the approach developed in [7] to locate AC. Mid-sagittal plane image of the scan data is extracted, which is used to compute a distance profile of the glottis from the posterior border of the neck. AC is defined to be the farthest point of the glottis from the posterior border of the neck. Laryngeal anatomical constraints ensure that the pairs of arytenoids cartilages are present close to the axial slice of AC in the caudal and cranial directions. Once the AC is located, axial slices on either side of the AC are scanned to locate the arytenoid cartilages to detect the feature points.

C. Airway Region Extraction

The selected axial slice image (as described in II-B) is first binarized by thresholding the original gray scale image. This is achieved by applying thresholds in a selected window of grayscale values. The lower and upper thresholds for the current dataset were -1200HU and -700HU, respectively. These thresholds were determined by using Otsu's multi-thresholding technique [8]. After binarization, 8-neighbor connected component analysis is carried out to determine the object regions in the image. The following anatomical constraints were used to formulate rules to detect airways:

- airway region cannot be located too close to the boundary of the image
- width and height of the airway region cannot be more than 50% of the width and height of the image
- width and height of the airway region cannot be less than 2% of the width and height of the airway region

Airway boundary is computed using the Moore-Neighbor Boundary Trace algorithm [9]. Since the arytenoids are always bounded by the thyroid cartilage on either side of the airway, image containing the airway region that is bounded by the thyroid cartilages on either side is extracted for further processing.

D. Preprocessing

Boundary pixels of the airway as shown in Fig. 2, are divided into two halves which come in contact with the two arytenoid cartilages on the left and right portions of the airway. Arytenoid cartilages are posteriorly attached to the vocal folds. Their movement during the phonation is constrained to the posterior parts of the vocal folds. This constraint is used to determine the cutoff locations (refer to Fig. 2c) beyond which arytenoids do not move. Hence, the pixels after the cutoff locations are filtered out. If we consider the upper half of the airway (Fig. 2a), cutoff pixel p is the pixel that satisfies the equation:

$$y(next(p)) - y(p) > 0, \qquad (1)$$

where $y(p)$ is the y coordinate of pixel p and $next(p)$ is the pixel after p, when the airway boundary is scanned in clockwise direction. If the lower half of the airway is considered, p should satisfy $y(next(p)) - y(p) < 0$. Boundary pixels after p are not considered for further processing.

E. Feature Point Detection

Feature point detection is carried out in the following two steps:

1) Step.1: Arytenoid cartilages are hyaline cartilages making them as distinct feature regions. Therefore, arytenoid cartilages appear bright on the CT images, resulting in bright pixels. As a result, distances of the airway boundary pixels from the first bright pixels in y direction, away from the airway are determined. A bright pixel is empirically defined to be a pixel with an intensity of more than 100HU for the aforementioned dataset. Let d be a function that computes City Block distance between any two pixels. Let p_i be the ith boundary pixel and B_i, corresponding bright pixel, then p_i is governed by:

$$p_i = \begin{cases} \text{potential feature point} & d(p_i, B_i) > T \\ \text{discarded} & d(p_i, B_i) \leq T, \end{cases} \qquad (2)$$

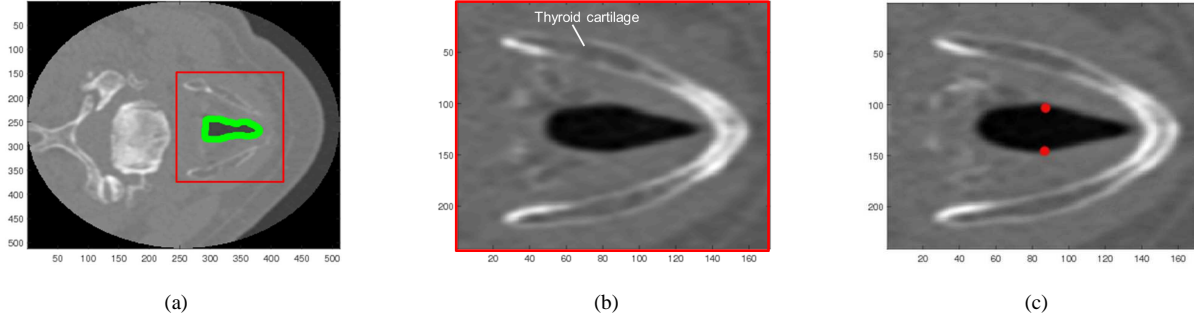where threshold T is empirically determined.

Fig. 2: Figure shows (a) detected airway outline on an axial slice, (b) image bounded by the thyroid cartilages for subsequent processing, (c) cutoff pixels identified by clinicians to perform filtering.

2) Step.2: Arytenoids contain air inside the cartilages, represented by negative intensities on the CT image. This characteristic is used as a rule in this step. The regions of the negative intensities are identified by calculating histogram of intensities. Let B^+ and B^- be the bins with positive and negative intensity edges, and let $N(B)$ be the number of pixels in a particular bin B. Then the ith boundary pixel p_i is governed by:

$$p_i = \begin{cases} \text{potential feature point} & N(B_i^+) < N(B_i^-) \\ \text{discarded} & N(B_i^+) \geq N(B_i^-) \end{cases} \quad (3)$$

F. Optimal Feature Point Detection

Pixels remaining after the previous steps are further clustered to form different clusters. Any two consecutive pixels that are separated by more than two pixels in either the row or column directions are assigned to different clusters. Means of distances of all the points of a cluster with respect to both references points (Fig. 3a) are summed, which is used as the final feature point. Cluster that has the minimum value is considered to best represent the edge of the arytenoid converging towards the airway. Centroid of this cluster gives the required feature point. It represents the anterior point on the arytenoid converging towards the airway and is a useful feature to get information about the movements of the arytenoids that support the vocal folds. The feature point detection can be carried out from the multiple axial slices, in which the arytenoids appear and can be interpreted by the clinicians.

III. RESULTS

The proposed approach is demonstrated on the CT data of 12 subjects, from the dataset described earlier. Ground truth was generated by marking the feature points manually. Euclidean distance measure was used to compute the error between estimated values and the ground truth. The implementations were performed in MATLAB 8.4 using the image processing toolbox on Windows 7 (64-bit system) equipped with an Intel® i7-4790 CPU running at 3.60 GHz.

Table I shows the estimation feature point coordinates, ground truths and estimated errors for 12 subjects using the proposed rule-based approach. Our interactions with
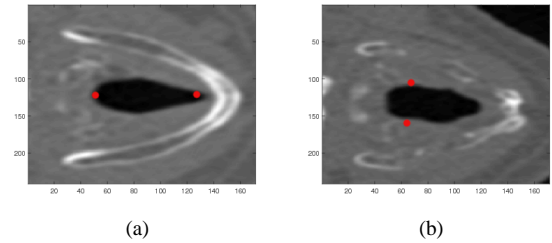

Fig. 3: Figure shows (a) reference pixels for clustering and (b) interest points of arytenoid cartilages.

clinicians resulted in an agreement to have an error tolerance of 15 pixels. From Table I, we see that for estimating lower arytenoid feature points, the proposed approach produced a maximum error of 21.54 pixels for Subject 2, whereas the maximum error for upper arytenoid feature points was 21.8 pixels for Subject 6. Only two subjects (Subject 2 and 6) had error over 15 pixel tolerance for estimating the lower arytenoid feature points. On the other hand, only one subject (Subject 6) had error more than 15 pixels in estimating the upper arytenoid feature point. Therefore, the estimation accuracies are 83.34% and 91.67% for lower and upper arytenoid feature points, respectively.
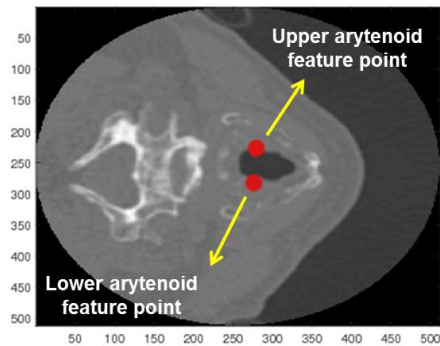
Fig. 4 shows the feature points detected using the proposed algorithm for subjects 9 (Fig. 4a) and 12 (Fig. 4b), respectively. From Fig. 4, it can be observed that despite different airway width and cartilage appearances across the two images, the proposed approach was still able to detect the feature points reliably. This automated approach helps to understand the movements of the vocal folds and identify the pathology due to the vocal fold disorders. Future directions include to account for variations in the anatomies of the arytenoids across multiple subjects by accounting for the patient movements during phonation, and also to handle movement artifacts.

IV. CONCLUSION

Analysis of vocal fold abnormalities is useful to assess the progress of Parkinson's disease. In the recent days, CT image

TABLE I: Comparison of estimated feature point coordinates with ground truth. The lower and upper coordinates indicate the feature points detected on arytenoid cartilages (either side of the airway).

Subject ID	Lower Arytenoid Feature Points			Upper Arytenoid Feature Points		
	Estimated [x, y, z]	Ground truth [x, y, z]	Error (pixels)	Estimated [x, y, z]	Ground truth [x, y, z]	Error (pixels)
1	293,317,76	290,310,76	7.61	234,323,76	236,314,76	9.21
2	274,324,92	266,304,92	21.54	222,320,92	225,306,92	14.3
3	265,280,52	263,279,52	2.23	230,293,52	229,287,52	6.08
4	276,392,63	274,389,63	3.60	244,390,63	243,390,63	1.00
5	274,333,69	274,330,69	3.0	232,332,69	230,329,69	3.6
6	266,300,42	284,296,42	18.43	286,302,42	265,296,42	21.8
7	266,311,53	264,312,53	2.23	233,311,53	231,312,53	2.23
8	266,314,56	264,309,56	5.38	238,314,56	236,313,56	2.23
9	280,275,30	277,266,30	9.48	226,278,30	225,269,30	9.05
10	300,247,38	294,240,38	9.21	249,253,38	246,248,38	5.83
11	284,339,52	283,333,52	6.08	249,332,52	247,328,52	4.47
12	267,315,66	265,312,66	3.60	242,316,66	242,314,66	2.0

(a)
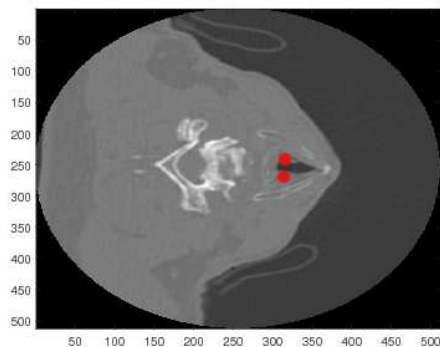
(b)

Fig. 4: Feature points detected using the proposed algorithm for (a) subject 9 (also shows the upper and lower arytenoid feature points) and (b) subject 12 marked on the axial slice image of the larynx.

scanning has become a reliable and non-invasive modality to image the human organs. Manual identification of arytenoids is a highly subjective and observer dependent task, leading to errors. Therefore, an automated technique is required to analyze CT images to provide reliable and consistent scores across subjects and observers. To this extent, a rule-based automatic arytenoid cartilage interest point detection is proposed. The proposed approach provided 83.33% accuracy in estimating clinically-relevant feature points. To the best of our knowledge, this is the first such automated approach to detect arytenoid cartilage feature points using laryngeal 3D CT images.

REFERENCES

[1] Parkinson's Disease Foundation, "Statistics on parkinson's," http://www.pdf.org/en/parkinson_statistics, 2017.
[2] R. J Holmes, J. M Oates, D. J Phyland, and A. J Hughes, "Voice characteristics in the progression of parkinson's disease," *International Journal of Language & Communication Disorders*, vol. 35, no. 3, pp. 407–418, 2000.
[3] J. A. Logemann, H. B. Fisher, B. Boshes, and E. R. Blonsky, "Frequency and cooccurrence of vocal tract dysfunctions in the speech of a large sample of parkinson patients," *Journal of Speech and Hearing Disorders*, vol. 43, no. 1, pp. 47–57, 1978.
[4] W. J. Mutch, A. Strudwick, S. K. Roy, and A. W. Downie, "Parkinson's disease: disability, review, and management." *Br Med J (Clin Res Ed)*, vol. 293, no. 6548, pp. 675–677, 1986.
[5] A. P. Zarzur, I. S. Duarte, G. d. N. H. Gonçalves, and M. A. U. R. Martins, "Laryngeal electromyography and acoustic voice analysis in parkinson's disease: a comparative study," *Brazilian journal of Otorhinolaryngology*, vol. 76, no. 1, pp. 40–43, 2010.
[6] L. Dumbrava, Lau.k, D.Phyland, P.Finlay, R.Beare, P.Bardin, P. Stuckey, P.Kempster, and D.Thyagarajn, "Vocal cords are hypokinetic in parkinson's disease," Departments of Neuroscience, Respiratory Medicine, Radiology Monash Medical Centre, Clayton, Victoria, Australia, Tech. Rep., 2014.
[7] S. Hewavitharanage, J. Gubbi, D. Thyagarajan, K. Lau, and M. Palaniswami, "Estimation of vocal fold plane in 3d ct images for diagnosis of vocal fold abnormalities," in *2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. IEEE, Conference Proceedings, pp. 3105–3108.
[8] N. Otsu, "A threshold selection method from gray-level histograms," *Automatica*, vol. 11, no. 285-296, pp. 23–27, 1975.
[9] P. R. Reddy, V. Amarnadh, and M. Bhaskar, "Evaluation of stopping criterion in contour tracing algorithms," *International Journal of Computer Science and Information Technologies*, vol. 3, no. 3, pp. 3888–3894, 2012.