A Review of Image Segmentation Techniques for Tracking the Velum

Anish Sana
Department of Computer Science
College of Science & Technology
East Carolina University
Greenville, NC – 27858
sanaa14@students.ecu.edu

Jamie L. Perry
Department of Communication Sciences & Disorders
College of Allied Health Sciences
East Carolina University
Greenville, NC – 27834
perryja@ecu.edu

Nasseh Tabrizi
Department of Computer Science
College of Science & Technology
East Carolina University
Greenville, NC - 27858
tabrizim@ecu.edu

***Abstract*— Several studies have used 2D and 3D modeling to visualize the velum. Very few attempts have been made to track the velum and plot its movement against time. Image segmentation has been used widely for various purposes. However, its proficiency in tracking the velum is questionable at the moment. Two image segmentation methods, EdgeTrak and the Hidden Markov Model, are reviewed in this report. EdgeTrak, a software developed at VIMS Lab, has been proven to track the surface of a human tongue during speech production. An attempt was made to similarly track the velum during speech production using EdgeTrak but the results were disappointing. Also, synchronized audio mapping using the Hidden Markov Model was only partially successful. This paper describes the challenges image segmentation faces with regards to tracking the velum.**

***Index Terms*—Image segmentation, tracking, velum, soft palate, machine learning**

Type – Short Paper

I. Introduction

Image segmentation can be used to detect objects where the goal is to cluster pixels into salient image regions using methods such as thresholding, clustering, histogram based, edge detection and stereovision based, among others [1].

Edge detection is one such method where image segmentation is performed based on the discontinuity in images that are spliced when an abrupt change in intensity occurs in the edges of an image [2]. Several edge detection techniques are currently in existence but a novel method called EdgeTrak was specifically developed by VIMS Labs to track a human tongue [3]. The method was successful in tracking the human tongue due to which, a study was conducted to enquire whether it can be used to track muscles other than the tongue, specifically the velum. This proved to be a failure for which the reasons are detailed in this report.

Additionally, another study is reviewed in the report where image segmentation was combined with a machine learning technique, Hidden Markov Model (HMM) to track the velum [4]. In an HMM, the state at some time encapsulates all information about the process in order to predict the future of that process [5]. Using this technique, the study tried to map movement in MRIs (magnetic resonance images) with its corresponding audio. A success rate of 81% was achieved in predicting velar movement. However, this is insufficient from a clinical standpoint as the purpose of tracking the velum is to obtain information on its movement with regards to speech that can be used to treat velopharyngeal inadequacy (VPI).

II. Background

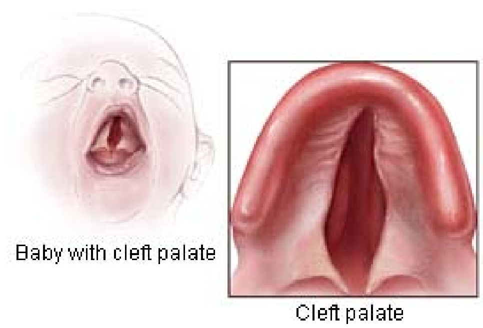

Fig. 1. Cleft Palate *[6]*

Cleft Palate (Fig. *1*) is one of the most commonly occurring birth defects. It occurs during embryonic development where a fissure is formed in the midline of the palate due to failure of the two sides to fuse [7].

Normal velopharyngeal anatomy consists of several muscles, which includes the levator veli palatini and the velum. These muscles are of particular interest as they aid in velopharyngeal closure, which is essential for speech production and swallowing. Although there are other muscles present in the velopharyngeal system, computational modelling has shown that the increase in LVP (Levator Veli Palatini) cross-sectional area and increase in extra velar length causes a closure force increase of more than 10% (Fig. *2*), due to which movement of the LVP and the velum needs to be studied [8].

Velopharyngeal closure is achieved by retraction and elevation of the velum due to contraction of the LVP. In children with a cleft palate, the LVP is attached onto the lateral and posterior aspect of the hard palate (Fig. *3*), which leads to several complications such as feeding, hearing and speech, among others [9]. Even with corrective surgery to restore anatomy, patients are sometimes unable to gain full speech due to velopharyngeal inadequacy which is characterized by hypernasality and sometimes require secondary surgery [10]. Tracking the velum and understanding

978-1-4673-9795-7/15 $31.00 © 2015 IEEE
DOI 10.1109/CSCI.2015.99

its movement in relation to speech can help speech pathologists customize speech training for patients after undergoing corrective surgery.

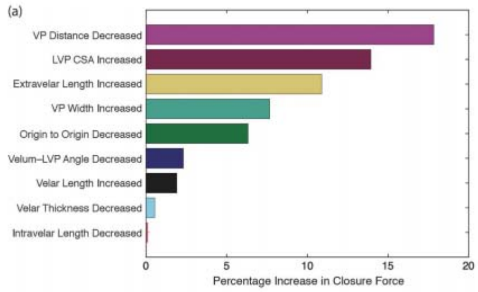

Fig. 2. Effect of LVP on closure force *[8]*

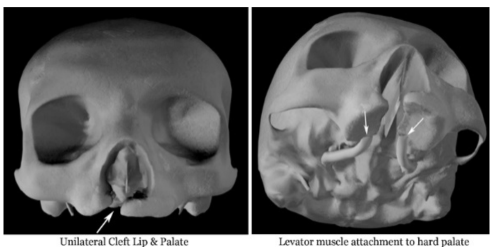

Fig. 3. Normal anatomy vs anatomy with a cleft palate *[9]*

III. EdgeTrak

EdgeTrak is a software developed to automatically track the surface of a human tongue in a sequence of ultrasound images which is a challenge due to noise and unrelated high contrast edges in ultrasound images. Instead of using only the gradient of images as the image force, EdgeTrak uses edge gradient and intensity information in local regions around snake elements. One of the advantages is that EdgeTrak can be used with open contours and track partial tongue surfaces whereas others can only be applied to closed contours. Also, any unnecessary edges are discarded. The software was successfully able to track the surface of a human tongue in ultrasound images (Fig. *4*) [11].

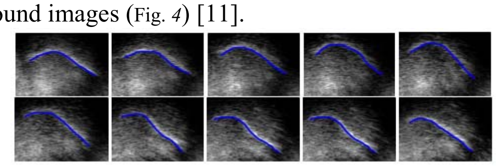

Fig. 4. EdgeTrak tracking the surface of a human *[11]*

A. Method

Given the success of EdgeTrak in tracking the human tongue surface, it was hypothesized that it could be used to similarly track the velum. An 8 second video of a child with normal anatomy uttering the phrase "pick up the pup" was used (Fig. *5*). This video was originally compiled from a sequence of MRIs. The video was split into 250 sequence of images using the software Blender (Fig. *6*) [12].These images were then cropped to the region of interest using MATLAB. The cropped images were loaded onto EdgeTrak and snake initialization was done on the velum (Fig. *7*). EdgeTrak was then allowed to automatically track the velum through all the images in the uploaded image sequence.

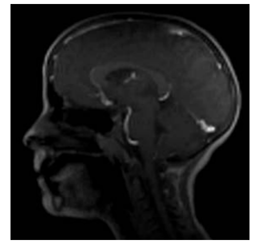

Fig. 5. Image from the video of a child saying "pick up the pup"

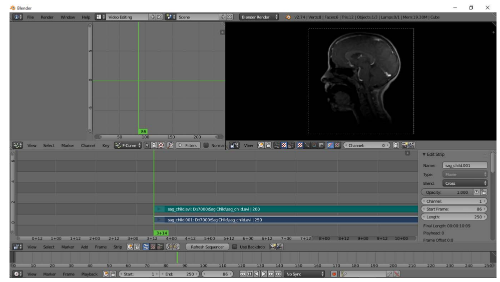

Fig. 6. The software Blender generating image sequences from videos

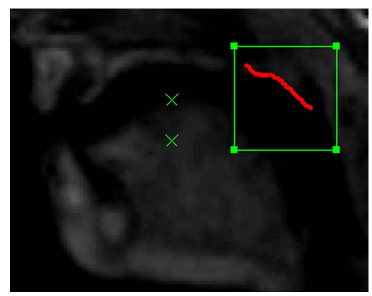

Fig. 7. Snake initialization on EdgeTrak

B. Discussion

Image quality is a major concern while using EdgeTrak. In sequence 1 (Fig. *8*), the velum is clearly visible in all instances due to which the software was able to efficiently track it. EdgeTrak relies on image intensity to track its objective. One of the problems with using image segmentation to track the velum has been that the velum and the PPW (posterior pharyngeal wall) have the same intensity. As it can be seen in sequence 2 (Fig. *9*), the snake gets attracted to the PPW and gets lost, due to the PPW intensity being similar to that of the velum. As mentioned previously, image quality is important for the software to track the velum. In sequence 3 (Fig. *10*), the velum disappears, due to which the size of the snake shrinks. Another major concern of EdgeTrak is consistency. Sequence 4 (Fig. *11*) shows the snake tracking the velum for the same sequence of images used in sequence 3, but noticeably getting smaller with each successive image and later regaining its size. This is clearly undesirable as results cannot be duplicated. One reason for this inconsistency is that EdgeTrak requires a human to manually initialize the snake and then tracks its objective based on initialized image intensity. A small error during initialization could cause the snake to show variability in its tracking. However, this is very unlikely to avoid as it is difficult for the human eye to select the same two pixels on an image.

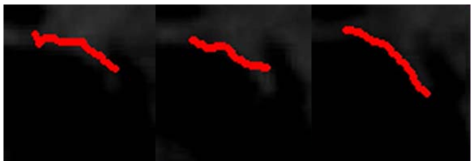

Fig. 8. Snake tracks the velum successfully

Fig. 9. Snake gets attracted to the PPW due to similar intensity

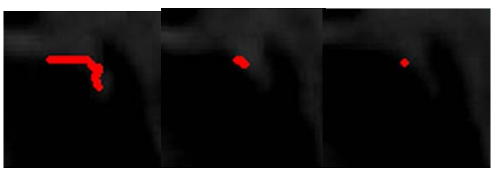

Fig. 10. Snake shrinks due to the velum disappearing

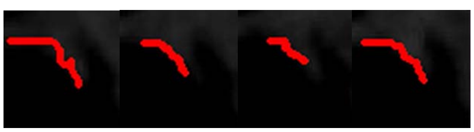

Fig. 11. Snake showing inconsistency while tracking same images

Additionally, EdgeTrak can only be used on a limited number of images due to memory constraints. The manual provided during installation states that less than or equal to 80 images should be used but as image size is increased, the number of images that can be used decreases.

IV. Hidden Markov Model

The HMM has been widely used to track objects. Researchers at Lund University [13] were successfully able to track multiple objects using HMM in image sequences on three different setups – footfall counter, parking lot monitor and car tracking in traffic surveillance videos. The parking lot setup was an initial test where the entrance to a narrow parking lot was monitored over 7 hours long and consisted of 17 events. All but one of the events were correctly detected, giving it an error rate of 3.6%. In the footfall setup, people entering and leaving a building were counted by tracking each person for a short distance to decide if the person was entering or leaving, with an accuracy rate of 96.4%. In the traffic setup, a 7 minute surveillance video was analyzed consisting of 58 cars and several large vehicles, of which 57 cars were detected. Not just objects, HMM can also be used to track signals as shown in a study [14], where the researchers were able to track two slowly varying time tones in additive white Gaussian noise. HMM can also be combined with other tracking methods, such as Augmenting Electro Optical (EO) based tracking systems with Infrared (IR) modality, known as Coupled Hidden Markov Model (CHMM) [15]. In this study, the researchers conducted experiments on real world sequence and reported improvement in tracking accuracy over other integration schemes where the target object is corrupted by noise.

The study that aimed to track the velum using HMM [4] consisted of 300 images tagged by the researcher of which 200 images and their corresponding audio features were used to train the HMM. A 2.5 second audio file was used to test the model where the error rate was considered to be the minimum calculated distance between predicted and actual markers. The model was able to track the velum with an accuracy of 81%.
Although the accuracy is high, there are several problems with the model –

Tracking – An accuracy of 81% means that the model can only successfully track the velum 4 out of 5 times. From a clinical standpoint, this is not sufficient. The purpose of tracking the velum is to gauge its movement with regards to speech. This information can then be used by either speech pathologists to train patients with VPI or clinicians to solve VPI through surgical means. For this purpose, the aforementioned accuracy is not sufficient to make informed decisions.

Human Errors – In the model, the researcher tags the images manually. This induces errors as was self-admitted by the researcher. The images were used to train the model and hence any errors would have continued throughout the image sequence which adds to the model's inaccuracy.

Performance – Machine learning algorithms (ML) such as HMM perform better with increase in data supplied to the model [16]. Given this, the model can never gain 100% accuracy until it obtains all the data required for future predictions, which in this context would mean the entire population of the planet. This is certainly not possible.

Repeatability – The predictions made in the study with the aforementioned accuracy are patient dependent. It is difficult to run the model on every patient.

Cyclical repetitions – The predictions made in the study were of patients uttering speech in cyclical repetitions. In order to be truly effective, the model needs to predict velar movement in regular speech.

These observations are supported by a study performed at François Rabelais University [17], where a new way of using HMM to track objects in video sequences was developed. The goal was to track a football during the entire length of a shot by predicting the approximate object position using a simple motion estimator first, following which the exact object position was computed. The method yielded a success rate of 87%. It only partially succeeded in tracking objects during occlusion (object of interest hidden partially in an image sequence), and faced difficulties when faced with two similar objects, such as the ball and a sock.

V. Conclusion

Computer based automatic tracking using image segmentation is often complicated due to the amount of time required to conduct the process and the inherent noise, motion artifacts, air interfaces and refractions in MRIs. Additionally, poor image quality and lack of a distinct boundary between the velum and the PPW make the process even more difficult [4].

EdgeTrak showed a lot of promise given the success it had in tracking the surface of a human tongue. However, the results with regards to tracking the velum have been disappointing. Several challenges remain for the software to be applicable to muscles other than the tongue. It needs to be able to work with images of poor quality as it is not always possible to obtain high quality MRIs or images of other nature. Its over-reliance on image intensity causes the snake to get attracted to areas other than the region of interest. An option to select the region of interest other than a box would allow the user to select the appropriate region and avoid the snake from getting distracted. An automated method of initializing the snake could help eliminate human error and provide consistent tracking.

ML algorithms such as the review of an HMM study performed in this report are successful in tracking the velum but possess many limitations. Accuracy is always an issue as HMM requires large amounts of data in order to be able to successfully predict velar movement that is beneficial for clinical purposes. In this particular study, there was an element of human error which could possibly be avoided in future models with the help of an automatic marker. Then again, having an automatic marker would mean that it is able to successfully track the velum, in which case other efforts are redundant.

Image segmentation can be used for tracking but it is currently unreliable and requires improvement. The technical limitations mentioned can be tackled with improvement in technology. However, the theoretical challenges remain which require further research for it to be useful for clinical purposes.

VI. References

[1] R. Grycuk, M. Gabryel, M. Korytkowski, R. Scherer and S. Voloshynovskiy, "From Single Image to List of Objects Based on Edge and Blob Detection," in *Artificial Intelligence and Soft Computing*, Zakopane, Springer International Publishing, 2014, pp. 605-615.

[2] P. Sujatha and K. K. Sudha, "Performance Analysis of Different Edge Detection Techniques for Image Segmentation," *Indian Journal of Science and Technology,* vol. 8, no. 14, pp. 1-6, 2015.

[3] Video/Image Modeling and Synthesis Lab, "EdgeTrak System," University of Delaware, 26 April 2012. [Online]. Available: https://www.eecis.udel.edu/wiki/vims/index.php/Main/EdgeTrak. [Accessed 9 September 2015].

[4] P. Rahimian, "Using Synchronized Audio Mapping to Predict Velar and Pharyngeal Wall Locations during Dynamic MRI Sequences," East Carolina University, Greenville, 2013.

[5] Z. Ghahramani, "An Introduction to Hidden Markov Models and Bayesian Networks," *International Journal of Pattern Recognition & Artificial Intelligence,* vol. 15, no. 1, pp. 9 - 42, 2001.

[6] Centers for Disease Control and Prevention, "Facts about Cleft Lip and Cleft Palate," U.S. Department of Health & Human Services, 20 October 2014. [Online]. Available: http://www.cdc.gov/ncbddd/birthdefects/images/cleft_palate_small.jpg. [Accessed 30 August 2015].

[7] E. A. Martin and T. A. McFerran, A Dictionary of Nursing (6 ed.), Oxford University Press, 2014.

[8] J. M. Inouye, J. L. Perry, K. Y. Lin and S. S. Blemker, "A Computational Model Quantifies the Effect of Anatomical Variability on Velopharyngeal Function," *Journal of Speech, Language & Hearing Research,* vol. 58, no. 4, pp. 1119-1133, 2015.

[9] J. L. Perry, "Anatomy and Physiology of the Velopharyngeal Mechanism," *Seminars in Speech and Language,* vol. 32, no. 2, pp. 83-92, 2011.

[10] P. Rong, "Using articulatory adjustment to compensate for hypernasality ? a modeling study based on measures of electromagnetic articulography (EMA)," ProQuest Dissertations Publishing, Ann Arbor, 2012.

[11] M. Li, C. Kambhamettu and M. Stone, "Automatic contour tracking in ultrasound images," *Clinical Linguistics & Phonetics,* vol. 19, no. 6-7, pp. 545-554, 2005.

[12] "About Blender," [Online]. Available: http://www.blender.org/manual/getting_started/about_blender/index.html. [Accessed August 31 2015].

[13] H. Ardo, K. Astrom and R. Berthilsson, "Real Time Viterbi Optimization of Hidden Markov Models for Multi Target Tracking," in *Motion and Video Computing, 2007. WMVC '07. IEEE Workshop on*, Austin, 2007.

[14] X. Xie and R. Evans, "Multiple target tracking using hidden Markov models," in *Radar Conference, 1990., Record of the IEEE 1990 International*, Arlington, 1990.

[15] J. Gai, Y. Li and R. Stevenson, "Coupled Hidden Markov Models for Robust EO/IR Target Tracking," in *Image Processing, 2007. ICIP 2007. IEEE International Conference on (Volume:1)*, San Antonio, 2007.

[16] M. Banko and E. Brill, "Scaling to very very large corpora for natural language disambiguation," in *ACL '01 Proceedings of the 39th Annual Meeting on Association for Computational Linguistics*, Stroudsburg, 2001.

[17] S. Lefevre, E. Bouton, T. Brouard and N. Vincent, "A new way to use hidden Markov models for object tracking in video sequences," in *Image Processing, 2003. ICIP 2003. Proceedings. 2003 International Conference on (Volume:3)*, 2003.