Computer Assisted Quantification of Hyoid Bone Motion in Fluoroscopic Videos

Ishtiaque Hossain[1], Angela Roberts-South[2], Mandar Jog[3] and Mahmoud R. El-Sakka[1]

[1]*Computer Science Department, Western University, London, Ontario, Canada*

[2]*Health and Rehabilitation Sciences, Western University, London, Ontario, Canada*

[3]*Department of Clinical Neurological Sciences, Western University, London, Ontario, Canada*

Keywords: Dysphagia, Swallowing Disorder, Videofluoroscopic Swallowing Study, Hyoid Bone, Object Detection, Haar Classifier, Tracking, Template Matching, Kinesiologic Analysis.

Abstract: The Videofluoroscopic Swallowing Study is a technique commonly used by radiologists to detect abnormalities in the swallowing process. While the subject swallows the food, X-ray images are taken and then compiled in a video form. The video is later analyzed by the radiologist using visual means. Since the nature of the inspection is highly subjective, the result of the inspection can barely be reliable. One of the assessed measures is the elevation of the hyoid bone during the swallow. This research introduces a semi-automatic method which identifies the hyoid bone in fluoroscopic videos and quantifies its motion. Before identifying the hyoid bone, the region-of-interest is automatically identified using a classification-based approach and subsequent image processing procedures are applied to the identified region-of-interest. Results show that the proposed method can accurately quantify the motion of the hyoid bone.

1 INTRODUCTION

The swallowing process begins as the food is chewed inside the mouth and ends when the food reaches the stomach. In order to detect abnormalities in the swallowing process, radiologists use a technique called Videofluoroscopic Swallowing Study, where the patient is instructed to swallow food mixed with barium sulphate and the swallowing process is recorded in the form of a video made of X-ray images. Barium causes the food to become visible in the captured video and this allows the radiologist to watch the activities inside the patient's throat during the swallowing process. The protocol for this method is described in more detail in the work of Palmer et al. (Palmer et al., 1993).

Usually, radiologists inspect a number of measures when evaluating the swallowing process, including the elevation of the hyoid bone. During a single swallow cycle, the hyoid bone is elevated (moves in an upward and forward direction) as the cycle begins. The hyoid bone then moves in the opposite direction, returning to its normal position as the cycle ends. Figure 1 shows the trajectory of the hyoid bone during a normal swallow. Paik et al. reported that the trajectory of the hyoid bone is significantly different from its normal trajectory for patients who have abnormalities in the swallowing process (Paik et al., 2008). It is indicative of the fact that, inspecting the trajectory of the hyoid bone can play an important role when evaluating the swallowing process.

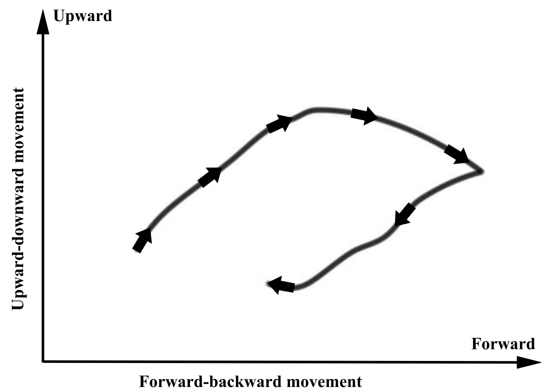


Figure 1: Trajectory of the hyoid bone during a normal swallow.

Currently, the evaluation procedure is performed by means of visual inspection. Due to the highly subjective nature of the evaluation process, achieving reliable result from the assessment can be a very challenging task. The severity of this issue is reported in

Hossain I., Roberts-South A., Jog M. and R. El-Sakka M..
Computer Assisted Quantification of Hyoid Bone Motion in Fluoroscopic Videos.
DOI: 10.5220/0004276707570761
In *Proceedings of the International Conference on Computer Vision Theory and Applications* (VISAPP-2013), pages 757-761
ISBN: 978-989-8565-47-1

a number of studies conducted on intrarater and interrater reliability (Kuhlemeier et al., 1998; McCullough et al., 2001; Stoeckli et al., 2003; Scott et al., 1998).

Evidently, the evaluation process demands more objective methods for quantifying various measures involved with the evaluation process. However, as of the writing of this article, only a few attempts have been made to meet this demand. Chen et al. proposed a computer aided method that measures and quantifies oral movement (Chen et al., 2001). Aung et al. proposed automatic identification of a number of anatomical landmarks using a 16-point active shape model (Aung et al., 2010b). In a different study, Aung et al. introduced a semi-automatic approach to determine the transit time of the bolus (Aung et al., 2010a). Kellen et al. proposed a semi-automatic method to track the hyoid bone (Kellen et al., 2010). It is worth mentioning here that in the work of Kellen et al., the region-of-interest is identified manually by means of user interaction.

This research concentrates on the problem of quantifying the movement of the hyoid bone. In this work, a semi-automatic method is introduced which attempts to identify and track the hyoid bone in fluoroscopic videos. At the same time, the cervical vertebrae are also identified which establish a relative referencing system. In order to limit image processing procedures to the relevant area of the image, the regions-of-interest are automatically identified before identifying the hyoid bone and the cervical vertebrae.

The rest of the paper is organized as follows. Section 2 presents the proposed method. The results are presented in Section 3. Section 4 concludes the article by commenting on the results. A number of directions to future work are pointed out in Section 5.

2 PROPOSED METHOD

The proposed method attempts to quantify the movement of the hyoid bone in fluoroscopic videos. Additionally, a referencing system relative to the patient is established by identifying the cervical vertebrae (see Section 2.3). Using a classification-based approach, the regions-of-interest are automatically identified in order to limit image-processing operations on a subregion of the image. By matching user defined templates, objects inside the regions-of-interest are identified.

2.1 Identifying the Region-of-Interest

The proposed method identifies the regions-of-interest using a method similar to the one proposed by Huang et al. where the lumbar vertebrae are detected using a learning-based method (Huang et al., 2009). Such a method is fast, requires no user interaction and can be tuned to achieve high accuracy. In this research, the regions-of-interest are automatically identified using the Haar classifier. The Haar classifier uses Haar features to classify sub-regions in the image and search the image for target objects (Viola and Jones, 2001). Instead of using the original features, an extended feature-set is used in this research which includes tilted features (Lienhart and Maydt, 2002).

The classifier is trained to identify the region-of-interest containing the cervical vertebrae. For training purpose, two sets of example images are prepared. The cervical vertebrae are present in one set (set of positive samples), and absent from the other (set of negative samples). As of the writing of this article, there is no conclusive study that dictates the optimum number of samples. However, Lienhart et al. conducted an empirical study on the training process with 5000 positive samples and 3000 negative samples and the positive samples are derived from 1000 images (Lienhart et al., 2003). In this research, the same number of samples is used. For the negative samples, high resolution random images are utilized.

The training process utilizes the *adaboost* method to iteratively classify the samples into their corresponding classes, minimizing the classification error at each step (Freund and Schapire, 1995). A single Haar feature performs as an input to a weak classifier. At each step, the *adaboost* method combines multiple weak classifiers in order to generate a boosted classifier. To speed up the detection process, a cascade of boosted classifiers is used.

It is not required to train a separate classifier for the purpose of identifying the region-of-interest for the hyoid bone. In the fluoroscopic videos, the hyoid bone is always located on the left side of the region-of-interest for the cervical vertebrae. This observation suggests that the region-of-interest for the hyoid bone can be inferred from the region-of-interest for the cervical vertebrae by mirroring the latter to the left. Figure 2 shows the identified regions-of-interest for the hyoid bone and the cervical vertebrae in one of the frames from the videos.

2.2 Tracking

After the regions-of-interest are identified, it is required to identify the objects of interest (each cervical vertebra and the hyoid bone) and track the objects throughout the video. Template matching is used to accomplish this task. Before tracking can be started,

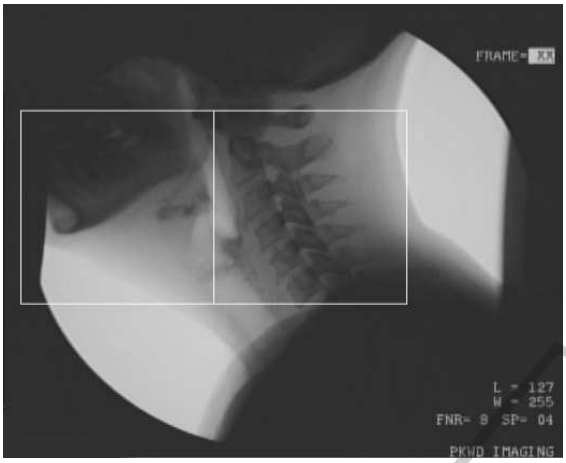


Figure 2: Regions-of-interest for the hyoid bone and the cervical vertebrae.

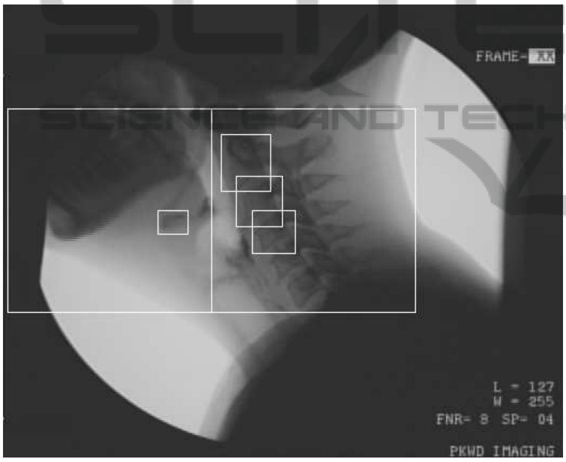


Figure 3: Identified locations of the hyoid bone and the individual cervical vertebrae.

the user has to manually identify the individual objects (the C2 vertebra, the C3 vertebra, the C4 vertebra and hyoid bone) in one of the frames in the video by identifying the smallest rectangle enclosing each object. The purpose of these rectangular regions is to serve as the templates of the objects. Figure 3 shows the identified locations of the hyoid bone and the individual cervical vertebrae inside the regions-of-interest in one of the frames from the videos. It can be argued that using generalized templates is more preferable than using templates from the video to be processed. However, the shape and the size of the objects can vary among patients and therefore, using generalized templates for all patients does not produce good results.

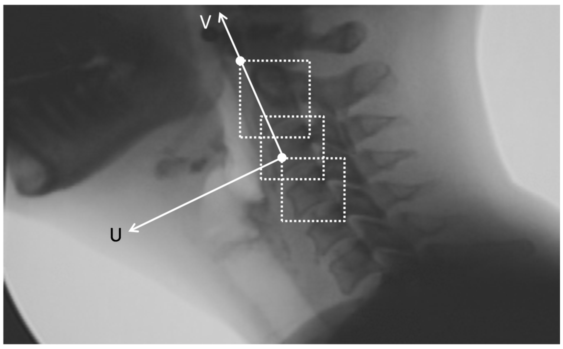


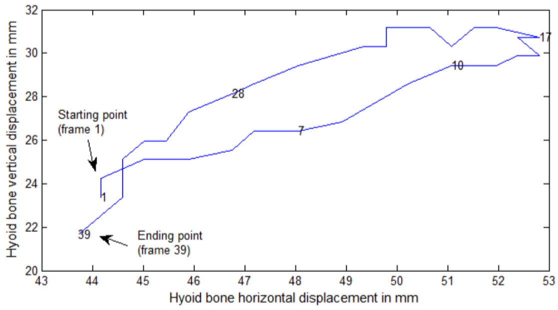
Figure 4: Relative referencing system.

2.3 Coordinate Transformation

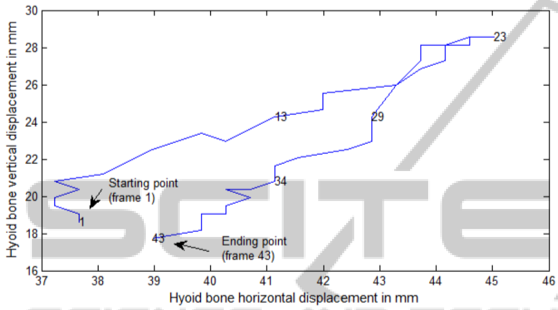
Originally, the results obtained using template matching (locations of objects) are expressed in terms of the coordinate system of the image, where the upper left corner of the image is the origin, the X axis lies in horizontal direction and the Y axis lies in the vertical direction. However, this approach does not allow us to distinguish between the movement of the hyoid bone and the movement of the patient's head. The movement of the hyoid bone needs to be isolated from the movement of the patient's head by expressing the movement of the hyoid bone in terms of a referencing system relative to the patient. In the relative referencing system, the line through the C2 vertebra and the C4 vertebra is the vertical direction (V axis). The line perpendicular to the V axis and passing through the C4 vertebra is the horizontal direction (U axis). Figure 4 shows the relative referencing system.

3 RESULTS

The trajectories of the hyoid bone in two sample swallow cycles are presented in Figure 5. The horizontal axis and the vertical axis correspond to displacement of the hyoid bone in horizontal and vertical direction, respectively (the U axis and the V axis described in Section 2.3). The unit for displacements, both horizontal and vertical, is in millimeters. Measured distances are calibrated by securing a coin of known diameter to the back of the patient's earlobe. Results obtained using the proposed method are compared with results obtained by manually identifying the hyoid bone in the same images. Table 1 shows the average and the standard deviation of distance between the locations of the hyoid bone obtained from both methods. It can be seen from Table 1 that results obtained from both methods are close to each other.



(a) Cycle 1



(b) Cycle 2

Figure 5: Trajectories of the hyoid bone for two sample swallow cycles. The horizontal axis and the vertical axis represent the horizontal and vertical displacement of the hyoid bone, respectively.

Table 1: Average and standard deviation of distances between locations of the hyoid bone obtained from proposed method and manual identification.

Cycle #	Distance (pixels)	
	Mean	Std
1	3.01	1.96
2	1.94	1.11

4 DISCUSSION AND CONCLUSIONS

This research introduces a semi-automatic approach to identify the hyoid bone and quantify its movement in fluoroscopic videos. Results indicate that the propose method measures the movement of the hyoid bone with a significant amount of accuracy. Identifying the region-of-interest allows us to perform image processing procedures to the most promising area in the image and to reduce computing time significantly. Therefore, automatic identification of the regions-of-interest can be useful in quantifying measures other than the elevation of the hyoid bone as well. Therapeutic use of the proposed method is one of the various medical applications where the measurement of the movement of the hyoid bone can be useful. The proposed method can also be utilized in studies that attempt to relate swallowing disorder to other diseases. The proposed method requires minimal input from the user. However, a fully automatic method is more preferable.

5 FUTURE WORK

In this research, the movement of the hyoid bone is assumed to be limited to the sagittal plane. Although this assumption holds for the data used in this research, the possibility of movements in the coronal plane cannot be eliminated. As a future work, the proposed method can be improved by detecting motion along the coronal plane.

ACKNOWLEDGEMENTS

We would like to thank Professor Mandar Jog, Director of the Movement Disorders Program, LHSC, and Ms Angela Roberts-South, Speech-Language Pathologist, for providing us with the necessary medical details and for being our primary medical and data resource for this research. Special credit goes to Dr. Donald Taves and the staff at the radiology department of Parkwood Hospital for their assistance with collecting the data that have been used in this research. We also acknowledge the Parkinson Disease Society of Canada for funding the entire data acquisition process through a grant to Dr. Jog and Ms. Roberts-South and for allowing us to use this data in our research.

REFERENCES

Aung, M., Goulermas, J., Hamdy, S., and Power, M. (2010a). Spatiotemporal visualizations for the measurement of oropharyngeal transit time from videofluoroscopy. *IEEE Transactions on Biomedical Engineering*, 57(2):432–441.

Aung, M., Goulermas, J., Stanschus, S., Hamdy, S., and Power, M. (2010b). Automated anatomical demarcation using an active shape model for videofluoroscopic analysis in swallowing. *Medical Engineering and Physics*, 32(10):1170–1179.

Chen, Y., Barron, J. L., Taves, D. H., and Martin, R. E. (2001). Computer measurement of oral movement in swallowing. *Dysphagia*, 16(2):97–109.

Freund, Y. and Schapire, R. (1995). A desicion-theoretic generalization of on-line learning and an application to boosting. In Vitnyi, P., editor, *Computational Learning Theory*, volume 904 of *Lecture Notes in*

Computer Science, pages 23–37. Springer Berlin / Heidelberg.

Huang, S.-H., Chu, Y.-H., Lai, S.-H., and Novak, C. (2009). Learning-based vertebra detection and iterative normalized-cut segmentation for spinal mri. *IEEE Transactions on Medical Imaging*, 28(10):1595–1605.

Kellen, P., Becker, D., Reinhardt, J., and Van Daele, D. (2010). Computer-assisted assessment of hyoid bone motion from videofluoroscopic swallow studies. *Dysphagia*, 25(4):298–306.

Kuhlemeier, K., Yates, P., and Palmer, J. (1998). Intra- and interrater variation in the evaluation of videofluorographic swallowing studies. *Dysphagia*, 13(3):142–147.

Lienhart, R., Kuranov, A., and Pisarevsky, V. (2003). Empirical analysis of detection cascades of boosted classifiers for rapid object detection. In Michaelis, B. and Krell, G., editors, *Pattern Recognition*, volume 2781 of *Lecture Notes in Computer Science*, pages 297–304. Springer Berlin/Heidelberg.

Lienhart, R. and Maydt, J. (2002). An extended set of haar-like features for rapid object detection. In *Proceedings of the 2002 International Conference on Image Processing*, volume 1, pages 900–903.

McCullough, G. H., Wertz, R. T., Rosenbek, J. C., Mills, R. H., Webb, W. G., and Ross, K. B. (2001). Inter- and intrajudge reliability for videofluoroscopic swallowing evaluation measures. *Dysphagia*, 16(2):110–118.

Paik, N.-J., Kim, S. J., Lee, H. J., Jeon, J. Y., Lim, J.-Y., and Han, T. R. (2008). Movement of the hyoid bone and the epiglottis during swallowing in patients with dysphagia from different etiologies. *Journal of Electromyography and Kinesiology*, 18(2):329–335.

Palmer, J. B., Kuhlemeier, K. V., Tippett, D. C., and Lynch, C. (1993). A protocol for the videofluorographic swallowing study. *Dysphagia*, 8(3):209–214.

Scott, A., Perry, A., and Bench, J. (1998). A study of interrater reliability when using videofluoroscopy as an assessment of swallowing. *Dysphagia*, 13(4):223–227.

Stoeckli, S. J., Huisman, T. A. G. M., Seifert, B. A. G. M., and Martin-Harris, B. J. W. (2003). Interrater reliability of videofluoroscopic swallow evaluation. *Dysphagia*, 18(1):53–57.

Viola, P. and Jones, M. (2001). Rapid object detection using a boosted cascade of simple features. In *Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, volume 1, pages 511–518.