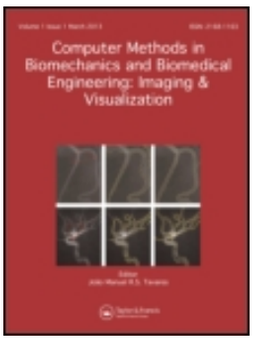

Taylor & Francis
Taylor & Francis Group

Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization

ISSN: 2168-1163 (Print) 2168-1171 (Online) Journal homepage: http://www.tandfonline.com/loi/tciv20

# Semi-automatic assessment of hyoid bone motion in digital videofluoroscopic images

Ishtiaque Hossain, Angela Roberts-South, Mandar Jog & Mahmoud R. El-Sakka

**To cite this article:** Ishtiaque Hossain, Angela Roberts-South, Mandar Jog & Mahmoud R. El-Sakka (2014) Semi-automatic assessment of hyoid bone motion in digital videofluoroscopic images, Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization, 2:1, 25-37, DOI: 10.1080/21681163.2013.833859

**To link to this article:** http://dx.doi.org/10.1080/21681163.2013.833859

*Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization*, 2014
Vol. 2, No. 1, 25–37, http://dx.doi.org/10.1080/21681163.2013.833859

Taylor & Francis
Taylor & Francis Group

# Semi-automatic assessment of hyoid bone motion in digital videofluoroscopic images

Ishtiaque Hossain[a], Angela Roberts-South[b], Mandar Jog[c] and Mahmoud R. El-Sakka[d]*

[a]*Department of Electrical Engineering and Computer Science, North South University, Dhaka, Bangladesh;* [b]*Health and Rehabilitation Sciences, Western University, London, Ontario, Canada;* [c]*Department of Clinical Neurological Sciences, Western University, London, Ontario, Canada;* [d]*Computer Science Department, Western University, London, Ontario, Canada*

(*Received 15 June 2013; accepted 6 August 2013*)

The swallowing process involves triggering the movements of a number of muscles in the throat that transports the food from the mouth to the stomach successfully and at the same time prevents it from getting into the airway and the lung. In order to detect abnormalities in the swallowing process, radiologists use a technique called videofluoroscopic swallowing study. It is a video of X-ray images that are taken while the patient swallows food, which is later visually inspected by the radiologist to evaluate the patient's swallowing ability. It has been reported that measuring the movement of the hyoid bone plays an important role in the evaluation process. However, due to the subjective nature of visual inspection, radiologists have difficulty reaching unanimous decision about the outcome of the evaluation. In this research, a semi-automatic method is proposed which tracks the hyoid bone and quantifies its movement. Using a classification-based approach, the proposed method automatically identifies the region of interest before identifying the hyoid bone. This allows limiting image processing procedures to the relevant area in the image. Results show that the proposed method identifies and tracks the hyoid bone with significant accuracy.

**Keywords:** dysphagia; videofluoroscopic swallowing study; hyoid bone; object detection; Haar classifier; tracking

## 1. Introduction

The process of swallowing starts with chewing the food inside the mouth and ends when the food is transported to the stomach. It is a complex process where a number of nerves and muscles work in a synchronised way to make the transportation process successful.

Abnormalities in the swallowing process are not rare. Studies in past years indicate that patients who suffer from movement disorders (especially patients with a history of Parkinson's disease, head and neck cancer, myopathy or stroke) are likely to have swallowing disorders (Stroudley and Walsh 1991; Leopold and Kagel 1996, 1997; Coates and Bakheit 1997; Martin et al. 1997, 2001; Smithard et al. 1997; Han et al. 2001; Potulska et al. 2003; Higo et al. 2005; Daniels et al. 2006; Warabi et al. 2008; Foley et al. 2009; Suh et al. 2009).

In order to detect swallowing disorders, radiologists use a technique called videofluoroscopic swallowing study. In this technique, the patient is seated in front of an X-ray machine and instructed to swallow food mixed with barium sulphate. While the patient swallows the food, the swallowing process is recorded in the form of a video. Barium makes the food visible in the resulting video, and this allows the radiologist to watch the activities inside the patient's throat while the patient swallows the food. A detailed description of the protocol for this method can be found in the work of Palmer et al. (1993).

There are a number of measures that radiologists usually inspect when evaluating the swallowing process. Rosenbek et al. (1996) proposed an eight-point scale to quantify aspiration in terms of penetration. This scale was later used by Robbins et al. (1999) to differentiate between normal and abnormal airway protection. Kendall et al. (2000) studied the timing of events occurring during a normal swallow. In addition, Stephen et al. (2005) reported the location of the bolus in a normal swallow as the pharyngeal phase is initiated. Kendall and Leonard (2002) investigated epiglottic movement in elderly patients. During a normal swallow, the hyoid bone moves in an upward and forward direction. This kind of upward and forward movement allows the oesophagus to open up so that the food can enter the oesophagus and continue to the stomach. At the same time, it allows the epiglottis to close the airway in order to prevent food from getting into the airway. When the swallowing process ends, the hyoid bone moves in the opposite direction, returning to its normal position. Figure 1 shows the normal movement of the hyoid bone and the epiglottis during swallow, where the vertical axis is the line connecting the anteroinferior corner of the 4th cervical vertebra (the origin) and the anteroinferior corner of the 2nd cervical vertebra and the horizontal axis is the line perpendicular to the vertical axis crossing the origin (Paik et al. 2008). Paik et al. demonstrated that the movement of the hyoid bone is

*Corresponding author. Email: elsakka@csd.uwo.ca

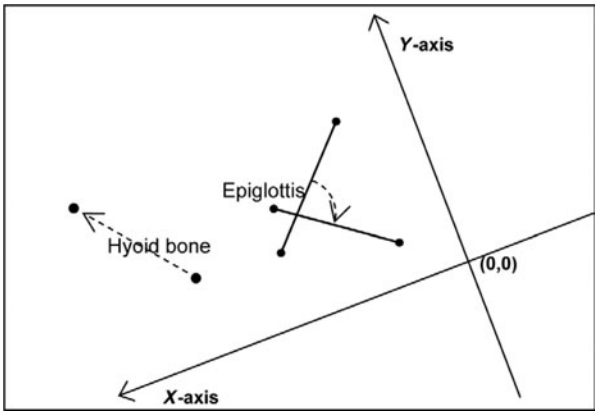


Figure 1. Normal movement of the hyoid bone and epiglottis during swallow (Paik et al. 2008).

significantly different from its normal trajectory for patients with swallowing disorders (Paik et al. 2008). Therefore, inspecting the degree of the movement of the hyoid bone can be helpful for radiologists when evaluating the swallowing process.

As current practice, radiologists perform this type of evaluation by means of visual inspection. This approach relies heavily on the expertise of the radiologist. Due to the subjective nature of this approach, radiologists find it difficult to reach an unanimous decision when multiple radiologists attempt to evaluate the swallow of the same patient (Scott et al. 1998; Stoeckli et al. 2003). Even when the same radiologist performs the evaluation for the same patient multiple times, the result of the evaluation tends to vary (Kuhlemeier et al. 1998; McCullough et al. 2001).

In order to help radiologists with quantitative evaluation of a patient's swallowing ability, computer assistance has been sought. In some studies, computer assistance is applied to the evaluation process in very primitive ways, such as going through the frames in the fluoroscopic video one by one and identifying various landmarks using a mouse pointer (Crary et al. 1994; Stephen et al. 2005; Dyer et al. 2008). This approach can hardly be considered a solution to the problem at hand, because it is labour intensive, time consuming and prone to inaccurate results.

It is evident that, in order to assist radiologists in the evaluation process, more objective methods are required. Such a method should be able to determine the measures accurately and should require the least possible amount of expertise from the user. However, as of the writing of this article, this particular area of research remains an open problem.

Recently, there have been only a few studies that attempt to meet these demands. Chen et al. (2001) proposed computer-aided measurement of oral movement in the fluoroscopic videos. Aung, Goulermas, Stanschus, et al. (2010) proposed a method that automatically identifies a number of anatomical landmarks (e.g. the hyoid bone, the anterior and posterior laryngeal wall, the corners of the cervical vertebrae) in the images using a 16-point active shape model. However, automating the initial placement of the active shape model can be challenging. In another study, Aung, Goulermas, Hamdy, et al. (2010) determined the transit time of the bolus. Kellen et al. (2010) devised an algorithm to track the hyoid bone in a video sequence. In their research, the authors manually identified the region of interest before identifying the hyoid bone.

This research focuses on measuring the movement of the hyoid bone since its trajectory plays an important role in the evaluation process. In this work, a semi-automatic method is introduced which attempts to identify and track the hyoid bone in the fluoroscopic video. While doing so, the cervical vertebrae are also identified and tracked so that the movement of the hyoid bone can be expressed in terms of a referencing system relative to the patient. This allows isolating the movement of the head from the movement of the hyoid bone. Prior to the identification of the cervical vertebrae and the hyoid bone, the regions of interest are automatically identified so that image processing procedures can be limited to the relevant area in the image. It is worth mentioning that the proposed method does not deliver judgement on what the movement of the hyoid bone implies when performing evaluation of the swallowing process. Rather, it merely serves as a tool that helps radiologists evaluate the swallowing process.

The rest of the paper is organised as follows: Section 2 explains the details of the proposed method. The results are accumulated in Section 3. Section 4 draws conclusion to the article by providing comments on the results and discussing some drawbacks of the proposed method. A number of directions to future work are pointed out in Section 5.

## 2. Proposed method

The proposed method attempts to track the hyoid bone in the fluoroscopic videos and measures its movement. In addition to the hyoid bone, the cervical vertebrae are also identified in order to establish a referencing system relative to the patient. A detailed description of the referencing system is provided in Section 2.3. Prior to the identification, the regions of interest are determined so that image processing operations can be limited to a sub-region of the image. This task is accomplished using a classification-based approach which is fully automatic and requires no input from the user. Objects inside the region of interest are identified using template matching. This part requires the user to identify templates from the video to be processed, and hence, the overall process is a semi-automatic one.

### *2.1 Identifying the region of interest*

Typical image segmentation techniques such as edge detection or image binarisation do not work very well for

identifying the cervical vertebrae or the hyoid bone because their performance relies on the threshold values associated with them. Selecting fixed values for these thresholds that apply for every image can be challenging. Corner detection is not a reasonable option either because the number of false corners can be large. An alternative approach was introduced by Huang et al. (2009) where the lumbar vertebrae are detected using a learning-based method. The advantage of the method proposed by Huang et al. is that it is fast, requires no user interaction and can be tuned to achieve high accuracy. In this research, a similar approach is followed, where the region of interest is automatically identified using the Haar classifier.

The Haar classifier, first proposed by Viola and Jones, uses Haar features to search an image for a target object (Viola and Jones 2001). The features comprise the adjacent rectangular regions at specific locations in a detection window. To detect the presence of a feature inside a detection window, the Haar features are applied to the window as convolution masks. The pixel intensities inside the detection window are convoluted with the mask and the convolution result is compared with a threshold to determine whether the feature is present at that location. The Haar features have advantage over pixelwise features in terms of efficiency. In their original work, Viola and Jones (2001) proposed an intermediate form of image (integral image) that allows features to be detected in a constant time. However, pixelwise features cannot be detected in a constant time. For example, if such a feature (e.g. line, circle or ellipse) is represented by a parametric equation, then the time to detect the feature is proportional to the number of parameters required to represent the feature. In their original work, Viola and Jones proposed three kinds of features: two rectangle, three rectangle and four rectangle. Later on, Lienhart and Maydt (2002) proposed an extended set of features which allows features to be tilted. In this research, the extended feature set is used for identifying the region of interest.

The classifier is trained first so that it can identify the region of interest for the cervical vertebrae. For training purpose, two sets of images need to be determined first: one set containing images of the cervical vertebrae and the other set with images that do not contain the cervical vertebrae. Images from the former set are the positive samples and images from the latter are the negative samples. The optimum number of samples in a set is not conclusive. In their original method, Viola and Jones (2001) used 4916 positive samples and 9544 negative samples. In another study, Lienhart et al. (2003) performed an empirical analysis based on the training process with 5000 positive samples and 3000 negative samples, where the positive samples are derived from 1000 images. Currently, a large database containing images of the cervical vertebrae that can be used as a set of positive samples is unavailable. Therefore, in this research, positive samples are generated from a small number of images by applying distortion to the images. A total of 5000 positive samples are used in this work, which are generated from 10 images (500 samples from each image) containing the cervical vertebrae. Figure 2 shows the images used for generating positive samples. The source image is rotated randomly on the sagittal plane. This type of distortion is applied in order to imitate the movement of the head, e.g. when the patient swallows liquid food and the cervical vertebra are tilted backward. In addition, white noise is added to the source image to mimic the addition of noise during the data acquisition process. The cervical vertebrae are surrounded by soft tissue which has a grey shade in the acquired images. Therefore, in the final step of applying distortion, the image obtained is placed onto a specified background colour (127 in this case). For the negative samples, 300 high-resolution random images

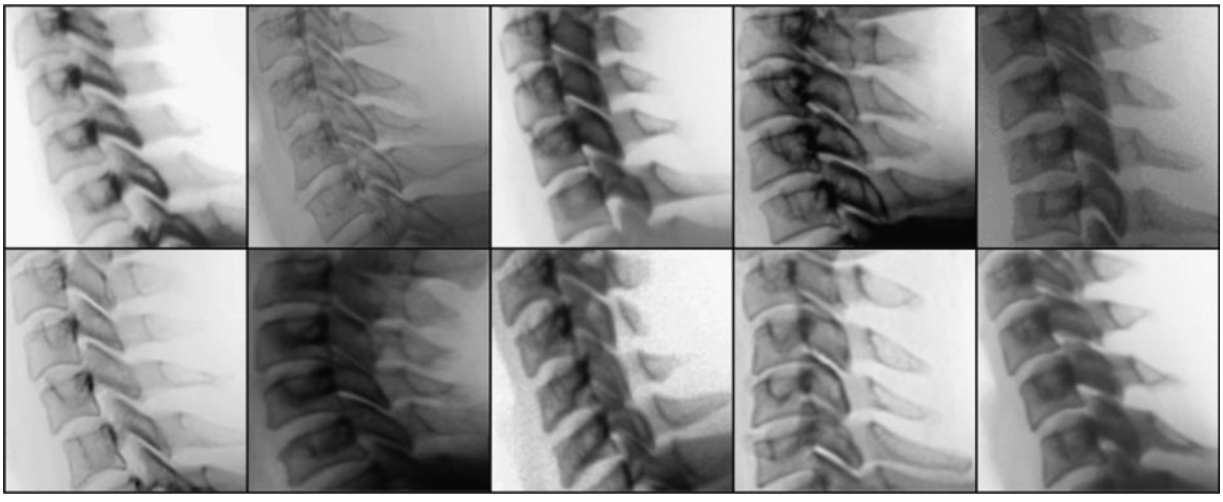

Figure 2. Positive images used while training the Haar classifier.

Figure 3. Some of the negative images used while training the Haar classifier.

are used. Figure 3 shows some of the negative images. From each of the negative images, 10 samples are generated by applying distortion. A total of 3000 negative samples are generated in this process.

The training process classifies the positive samples and the negative samples into two classes: one containing the samples with the cervical vertebrae and the other containing samples without the cervical vertebrae. A Haar feature serves as an input to a weak classifier. To generate a strong classifier out of the weak classifiers, the *adaboost* method is used. The *adaboost* (Adaptive Boosting) method is an iterative process which increases the classification performance (*boosting*) by combining multiple classifiers together (Freund and Schapire 1997).

To speed up the detection process, a cascade of boosted classifiers is utilised. The training process is performed in multiple stages where a boosted classifier is generated in each stage. Upon completion of the training process, the boosted classifiers are organised in a cascaded fashion. For a given region in the image, the cascade is applied to the region starting with the first classifier in the cascade. If at some point, the region is rejected by a classifier, the region is not passed to successive classifiers in the cascade. The target object is considered to be present in the region only if it is passed by all the classifiers in the cascade.

To identify the location of the target object (the cervical vertebrae in this case) in the image, a detector window is moved across the image. Since the scale of the cervical vertebrae can vary for different patients, the scan procedure is done in multiple stages with various scaling factors. The detector window starts with a size the classifiers are trained at, and at each stage, the size of the detector window is multiplied by a scaling factor. For scanning, the detector window is moved round($s\Delta$) pixels where $s$ is the current scaling factor and $\Delta$ is a predetermined constant.

The region of interest for the hyoid bone can be identified without using a separate classifier for it. It can be inferred from the region of interest for the cervical vertebrae. Observations reveal that in the fluoroscopic videos, the hyoid bone is always located on the left side of the region of interest for the cervical vertebrae. Therefore, the region of interest for the cervical vertebrae can be flipped to the left side and it can serve as the region of interest for the hyoid bone. Figure 4 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

Although the regions of interest obtained using the Haar classifier contain the object of interest (a cervical vertebrae or the hyoid bone), it is required to narrow the region of interest down to the smallest rectangle that encloses the object and tracks the object throughout the video. It should be mentioned here that objects can vary among patients in terms of size and structure. For this reason, basic image segmentation techniques such as edge detection or image binarisation do not produce good results for all the patients. In addition, objects can get blurred due to their motion in which case, it is even more difficult to segment them using these methods. In this research, template matching is used for tracking purpose which utilises correlation instead of features to detect objects.

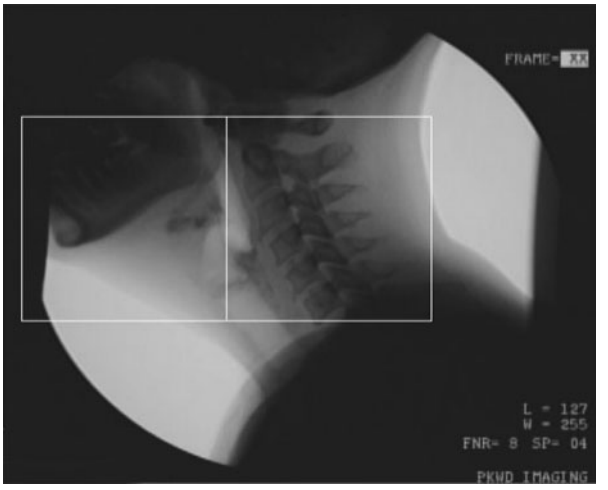


Figure 4. Regions of interest for the hyoid bone and the cervical vertebrae.

Before tracking can be started, the user has to manually identify the regions that enclose the individual objects (C2, C3, C4 and hyoid bone) in one of the frames from the video where the objects are clearly visible and are not affected by motion blur. These regions serve as the templates of the objects to be tracked. The templates are matched inside the corresponding region of interest for all the frames in the video. The reason for using templates from the video to be processed is that the shape and the size of the objects can vary among patients and, therefore, using generic templates of the objects for all patients does not produce good results.

At the beginning of a swallow segment, the templates are matched in the region of interest identified in Section 2.1. Matching templates in the region of interest rather than the entire image saves a significant amount of computation time. It also reduces the probability of encountering invalid match locations for the templates. Match locations are expressed in terms of bounding boxes that enclose the object of interest. For each object in the later frames, the search space for template matching is further reduced to a smaller region which is double the size of the bounding box (the bounding box is small compared with the region of interest and multiplying its size by a factor of two still yields smaller area than that of the region of interest). This approach allows the template to be matched in the proximity of a previously found location, which is an even smaller region than the region identified by the classifier. Figure 5 shows this process where a sample region of interest identified in Section 2.1 is shown. At the beginning of the swallow, the entire region is the search space. The white rectangle is the location of the C3 found in the current frame. In the next frame, the region of interest for the C3 is the black rectangle.

It should be mentioned here that the patient can tilt his/her head back and forth while swallowing. To cope with this situation, the templates are rotated along their centres

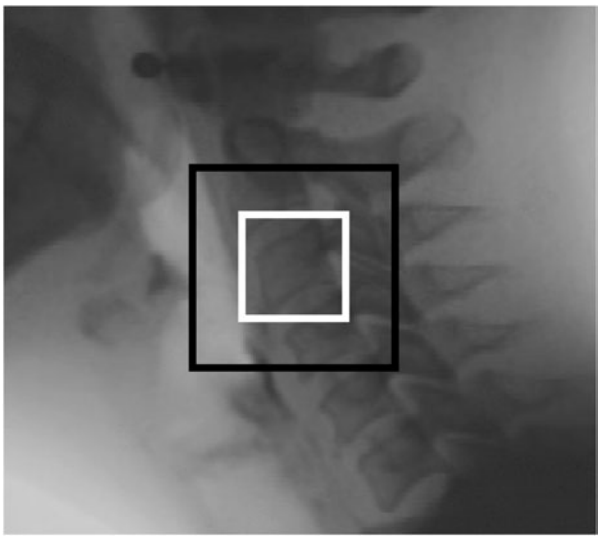

Figure 5. Reduction of region of interest for C3.

from $-10^{\circ}$ to $+10^{\circ}$ with a step size of $1^{\circ}$, and template matching is performed for each of the rotated templates. Twenty-one correlation maps are generated in this process. Out of these 21 correlation maps, the highest correlation value is selected as the best match for a vertebrae.

To ensure that the locations of the cervical vertebrae are identified correctly, a number of heuristics are applied. The template matching process yields three rectangles for the cervical vertebrae, each enclosing one vertebrae. It can be observed that the enclosing rectangles of two consecutive vertebrae overlap with each other slightly. To be more precise, the rectangle for C2 overlaps with the rectangle for C3 and the rectangle for C3 overlaps with the rectangle for C4. Also, the rectangle for C3 cannot be located above the rectangle for C2, and the rectangle for C4 cannot be located above the rectangle for C3 or C2. The upper left corner and the lower left corner of the enclosing rectangle correspond to the anterosuperior corner and the anteroinferior corner of the enclosed vertebra, respectively. Figure 6(a) shows the ideal arrangement of the rectangles for each of the cervical vertebrae. Figure 6(b) 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 should be mentioned here that the template matching process is sensitive to motion blur, which can lead to inaccurate results. Motion blur usually occurs when big movement of the head is involved or the hyoid bone moves too fast. By using heuristics and matching templates in the proximity of previously found locations, this effect is reduced.

### 2.3 *Coordinate transformation*

While swallowing, the patient can move his/her head back and forth and this motion contributes to the motion of the

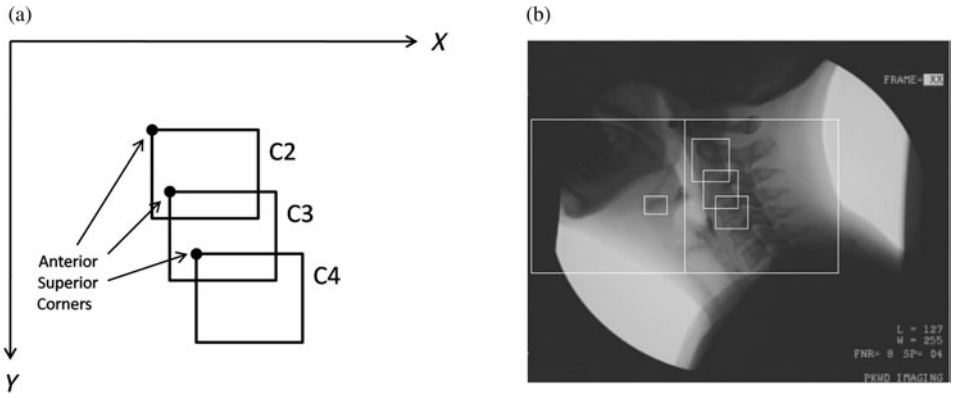


Figure 6. Locations of objects inside the regions of interest identified using template matching. (a) Ideal arrangement of the locations of the cervical vertebrae and (b) identified locations of the hyoid bone and the individual cervical vertebrae.

hyoid bone at the image level. To compensate for the movement of the head, the location of the hyoid bone is required to be expressed in terms of a referencing system relative to the patient. Originally, locations of the objects in the image 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. The relative referencing system is derived from the alignment of the cervical vertebrae. This is accomplished by selecting the line through the anterosuperior corners of C2 and C4 as the vertical direction (*V*-axis) and the line passing through the anterosuperior corner of C4 and perpendicular to the *V*-axis as the horizontal direction (*U*-axis). Figure 7(a) illustrates the relative referencing system.

Figure 7(b) shows the transformation of coordinates. Let the coordinates of the anterosuperior corners of C2, C4 and the hyoid bone be C2$(x_2, y_2)$, C4$(x_4, y_4)$ and H$(x_h, y_h)$ respectively, where the origin of the coordinate system is the upper left corner of the image. Let $\vec{A}$ is the vector from C2 to C4 and $\vec{B}$ is the vector from H to C4. The angle ($\theta$) between $\vec{A}$ and $\vec{B}$ can be obtained by taking the dot product of these two vectors and dividing the result by the product of their magnitudes:

$$\theta = \arccos \frac{\vec{A} \cdot \vec{B}}{\|A\|_2 \|B\|_2} \tag{1}$$

or

$$\theta = \arccos \times \frac{(x_4 - x_2)(x_4 - x_h) + (y_4 - y_2)(y_4 - y_h)}{\sqrt{(x_4 - x_2)^2 + (y_4 - y_2)^2}\sqrt{(x_4 - x_h)^2 + (y_4 - y_h)^2}}. \tag{2}$$

If $r$ is the distance from H to C4 and $(u, v)$ is the transformed coordinate (expressed with respect to the coordinate system derived from the cervical vertebrae) of H, then $u$ and $v$ can be derived as follows:

$$u = r \sin \theta, \tag{3}$$

$$v = r \cos \theta, \tag{4}$$

where

$$r = \sqrt{(x_4 - x_h)^2 + (y_4 - y_h)^2}. \tag{5}$$

Note that the movement of the hyoid bone is assumed to be two dimensional. The patient can move his/her head back and forth and also the hyoid bone can move in its own

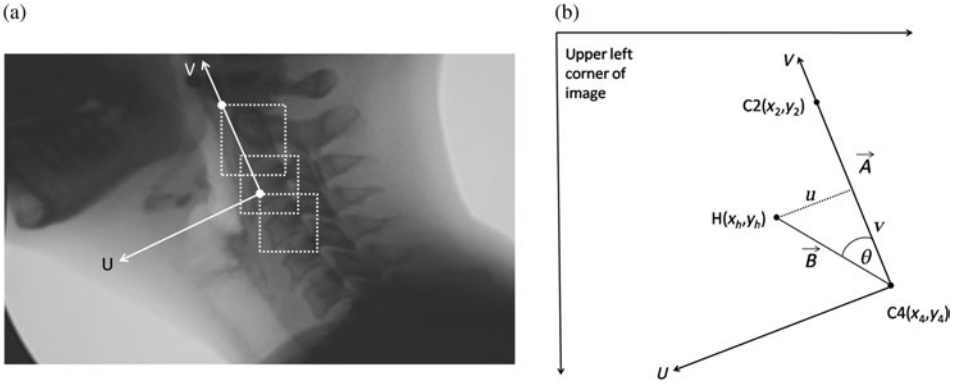


Figure 7. The relative referencing system. (a) Modified axes and (b) coordinate transformation.

trajectory. All of these movements take place in the *XY*-plane (sagittal plane). Since the patient does not move his/her head sideways, motion along the *Z*-axis (coronal plane) can be ignored.

## 3. Results

In this research, the picture archiving and communication system (PACS) is utilised to acquire and store the data. A digital X-ray machine resides at the modality end of the PACS. Images acquired at the X-ray machine are sent to the archive where the data are stored in the form of video sequences (encoding scheme: MPEG-2, frame format: YUV420P, bitrate: 9253 kb/s, frame rate: 30 fps, frame size: 720 × 480) in regular DVD media. Using the libraries provided by the FFMPEG project (available at http://ffmpeg.org), the images are decoded from the videos and processed offline.

Before the proposed method is applied to the images, certain frames are removed from the videos. A swallow segment starts with a sequence of images which displays information about the patient. This information concerns the radiologists and does not play any role in the data processing. Moreover, this information is sensitive, which is why it is not ethical for the radiologist to disclose this information. As a preprocessing step, these images are removed from the videos so that the patient's information cannot be accessed by people other than the radiologists. The images are removed by inspecting their variance. The variance of images with patient information on them is low ($309.8 \pm 85.4$) compared with the rest of the images in the videos ($2123.8 \pm 275.8$).

The validity of the identification process of the region of interest is verified by comparing the result of the automatic identification with that of a manual identification process. Three hundred and fifty sample frames are randomly selected from multiple videos and the region of interest for the cervical vertebrae is identified using the classifier described in Section 2.1. For the same frames, the region of interest is also manually identified using visual inspection. The results obtained from both methods are compared with each other. The region of interest is a rectangle defined by the coordinates of two corners (the upper left corner and the lower right corner of the rectangle). Figure 8 shows the comparison between the coordinates obtained using the classifier and the manual identification process. It should be mentioned here that for the purpose of comparison, coordinates do not need to be expressed in term of a relative referencing system. Therefore, the coordinates for this experiment are expressed in terms of the referencing system of the image where the upper left corner of the image is the origin.

Table 1 shows the result of the comparison between automatic and manual identification of region of interest. The last two columns show the mean square error and the percentage error between the coordinates obtained using automatic and manual identification, respectively.

To demonstrate the final results, seven swallow cycles are identified from multiple videos. In one cycle, the hyoid bone travels along its trajectory (moves forward and upward and comes back to rest) once. The coordinates of the hyoid bone, obtained using the proposed method, are plotted to show the trajectories of the hyoid bone, where the origin is that of the modified referencing system (see Section 2.3) and the coordinates are expressed with respect to the modified referencing system. The trajectories are presented in Figure 9. The horizontal and the vertical axes correspond to the horizontal and vertical displacements of the hyoid bone, respectively. The unit for both the horizontal axis and the vertical axis is in millimetres. To calibrate measured distances, a coin (Canadian penny) of known diameter is secured to the back of a patient's earlobe in one of the studies and its diameter in terms of number of pixels is measured manually.

For validation purpose, the results obtained by using the proposed method (see Figure 9) are compared with that obtained from a manual identification method, where the location of the hyoid bone is identified manually for each frame in the swallow cycles. To ensure that the manual identifications are performed correctly, the result from the manual identification process is validated by a speech-language pathologist. This result is then matched with the result obtained by using the proposed method. Figures 10 and 11 compare the two methods in terms of the horizontal and vertical displacements, respectively, of the hyoid bone for all the swallow cycles. In each plot, the horizontal axis represents the frame number and the vertical axis represents the displacement of the hyoid bone obtained from both methods (solid line represents the manual method and the broken line represents the proposed method). It should be mentioned here that, for the purpose of comparison, locations of the hyoid bone do not require to be expressed in terms of the referencing system described in Section 2.3. Therefore, when comparing the two methods, the coordinate system of the image is utilised to measure the displacement of the hyoid bone where the origin is upper left corner of the image. Hence, the unit for displacement is number of pixels. Note that there is a peak in the plot of Figure 10. This occurs because, in this case, the hyoid bone is partially occluded by the jaw bone for a little while during the cycle, and the template matching process finds the best match for the template in a location close to the desired location, but not the same. The template matching process yields the desired result once the hyoid bone is clearly visible again.

Table 2 shows the mean square error and the percentage error between the two methods for both horizontal and vertical directions of the hyoid bone. From this table, it can be seen that the averages of both the mean

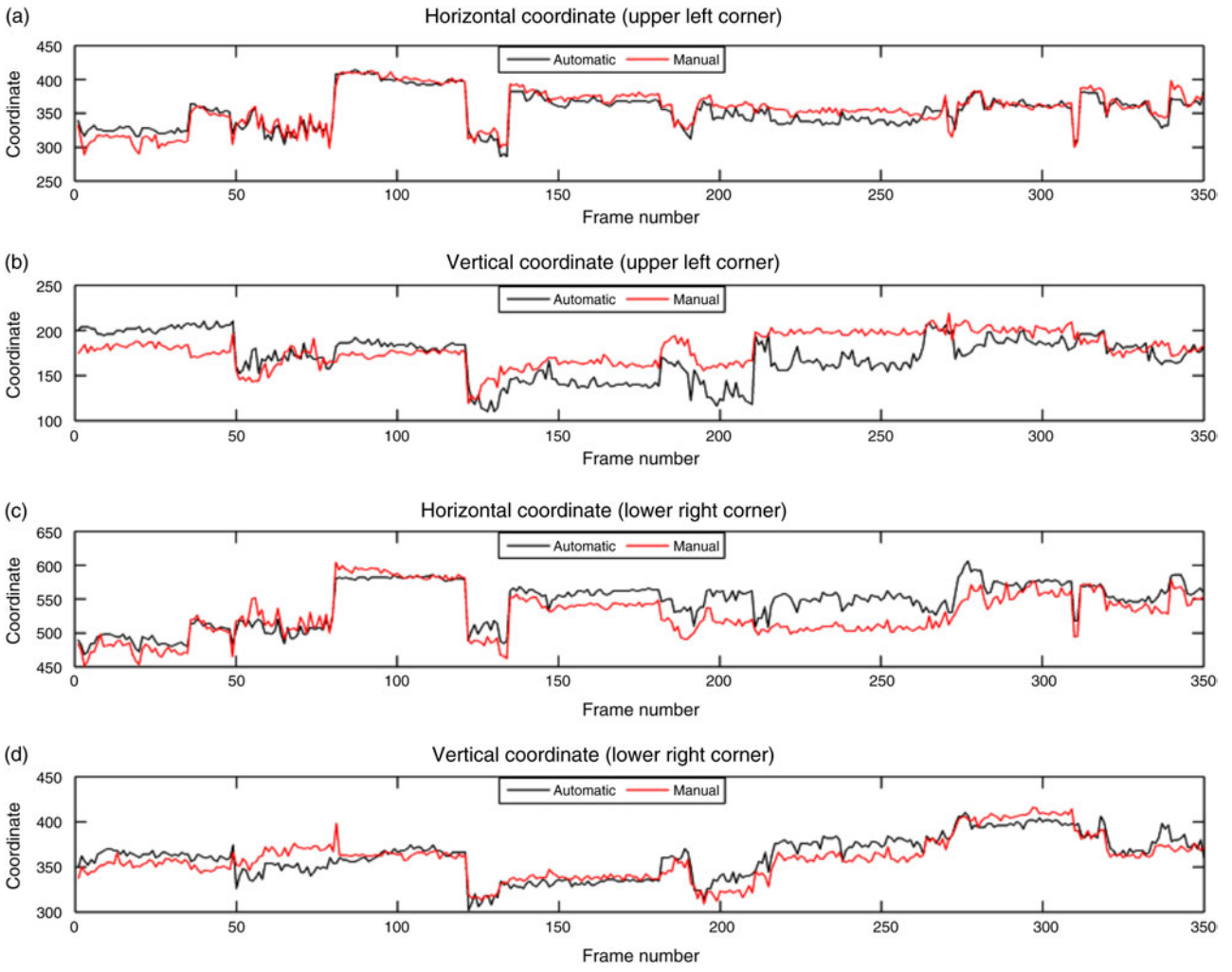


Figure 8. Comparison of automatic and manual identification of region of interest. The horizontal axis represents the frame number and the vertical axis represents the coordinates defining the region of interest.

Table 1. Mean square error and the percentage error between the proposed method and manual method for region of interest detection.

| Corners of region of interest | | Mean square error | Percentage error |
|---|---|---|---|
| Upper left | Horizontal coordinate | 151.63 | 2.89 |
| | Vertical coordinate | 481.45 | 10.41 |
| Lower right | Horizontal coordinate | 584.31 | 3.78 |
| | Vertical coordinate | 348.13 | 3.93 |

square error and the percentage error are low (9.75% and 0.64%, respectively, at most) for both the axes. It indicates that the proposed method can successfully mimic the process of manual identification.

## 4. Discussion and conclusion

The measurement of the movement of the hyoid bone can be useful in a variety of medical applications, including therapeutic use of the proposed method. In order to cope with swallowing disorder, patients are sometimes prescribed to go through various therapies (e.g. chewing gum before having a meal). The proposed method can provide feedback on the therapy and this will allow the radiologist to investigate the results of the therapy. The proposed method can also be utilised to relate swallowing disorder to other diseases. Section 1 briefly discusses some of the studies conducted in this area. So far, the current practice is based on manual inspection. By introducing an objective method that quantifies the measures involved with the swallowing process, the radiologists will be able to better analyse the data.

Although the proposed method successfully identifies and tracks the hyoid bone, it has some limitations as well. Although minimal, it requires some input from the user. A fully automatic method is more preferable. However, this can be challenging, given that the objects we are tracking

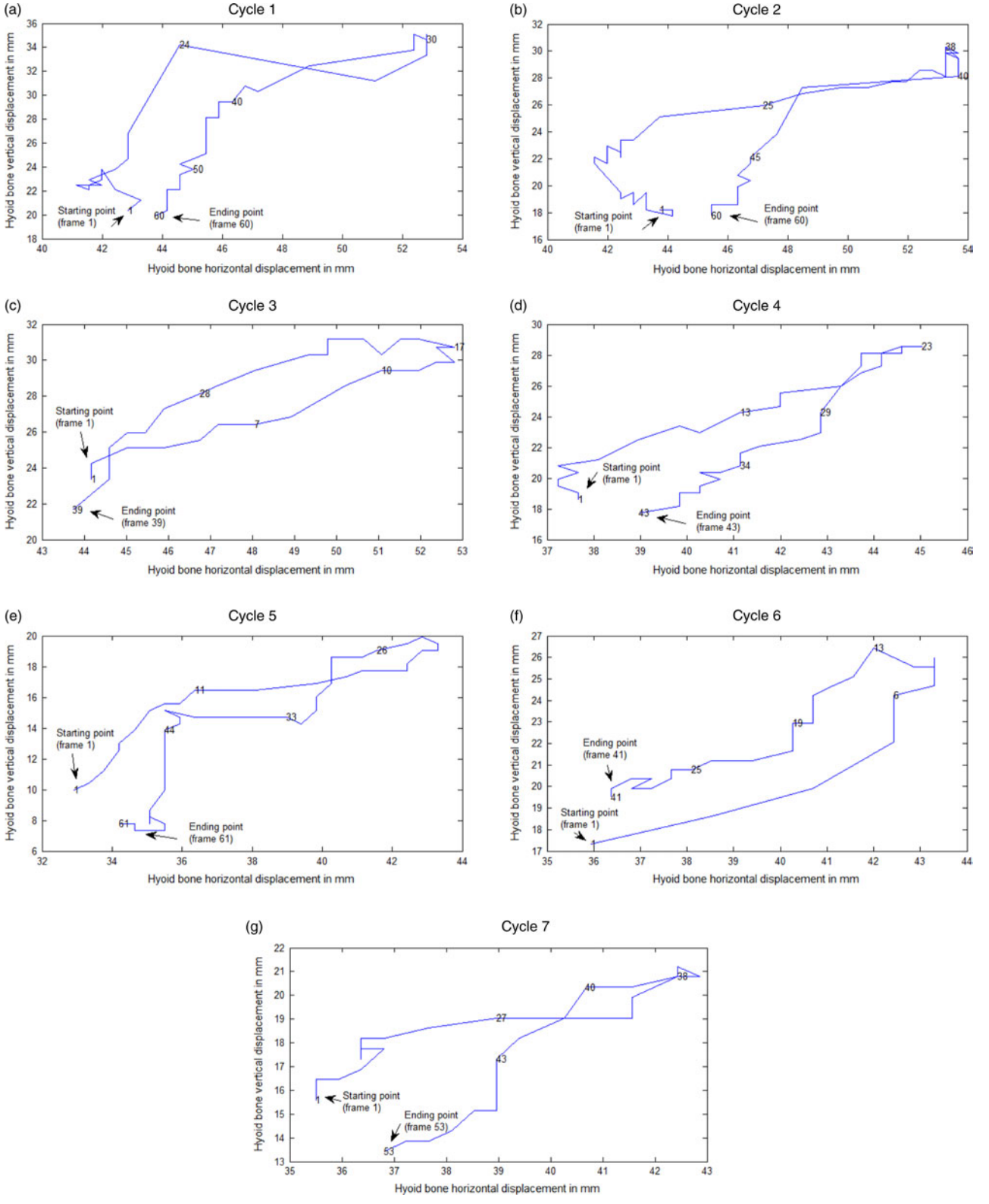


Figure 9. Trajectories of the hyoid bone for various swallow cycles. The horizontal axis and the vertical axis represent the horizontal and vertical displacements of the hyoid bone, respectively.

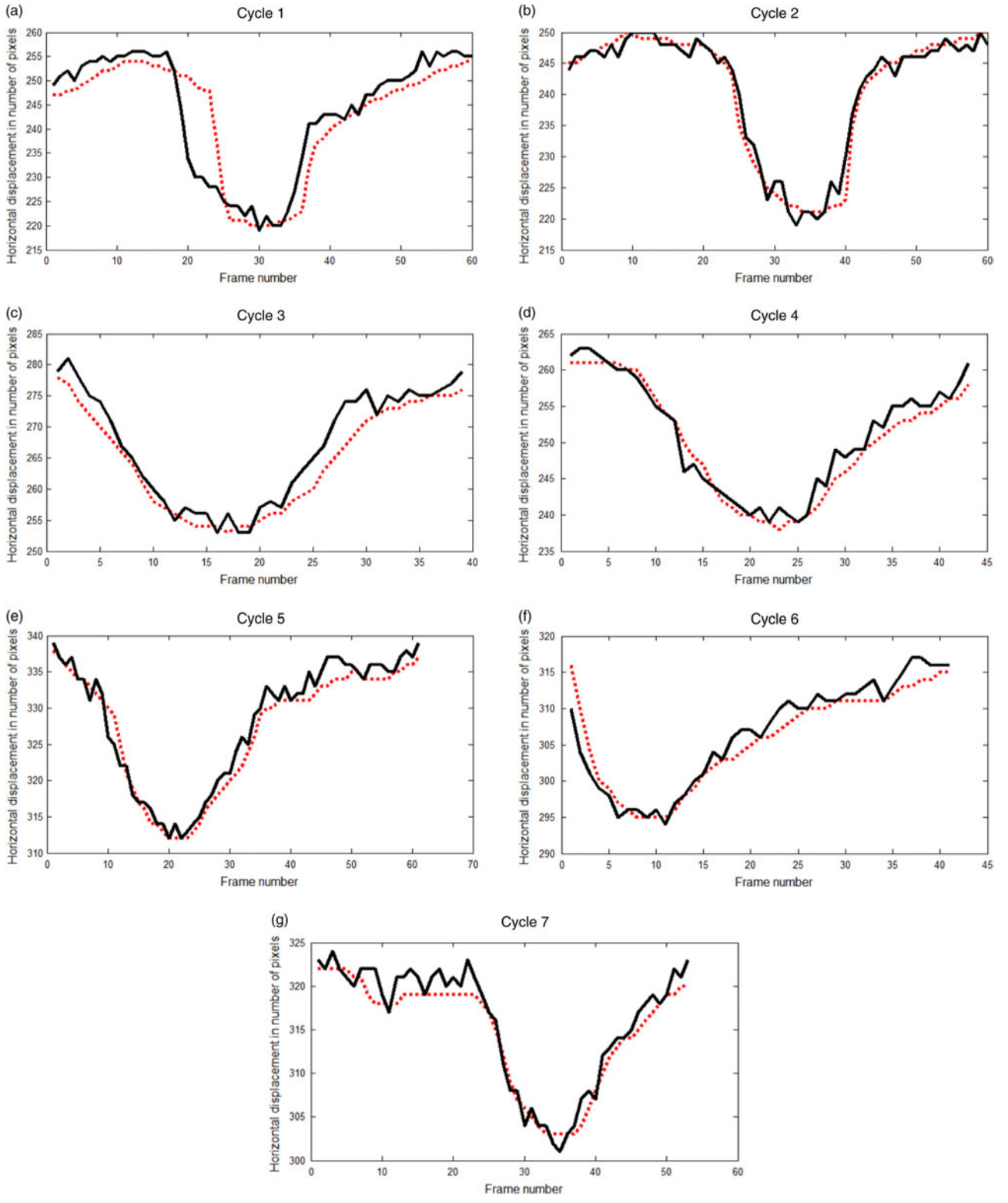


Figure 10. Comparison of proposed and manual methods for horizontal displacement of the hyoid bone (cycles 1–7 in Figure 9). The horizontal axis represents the frame number and the vertical axis represents the horizontal displacement of the hyoid bone obtained from both methods. Solid line represents the manual method and the broken line represents the proposed method.

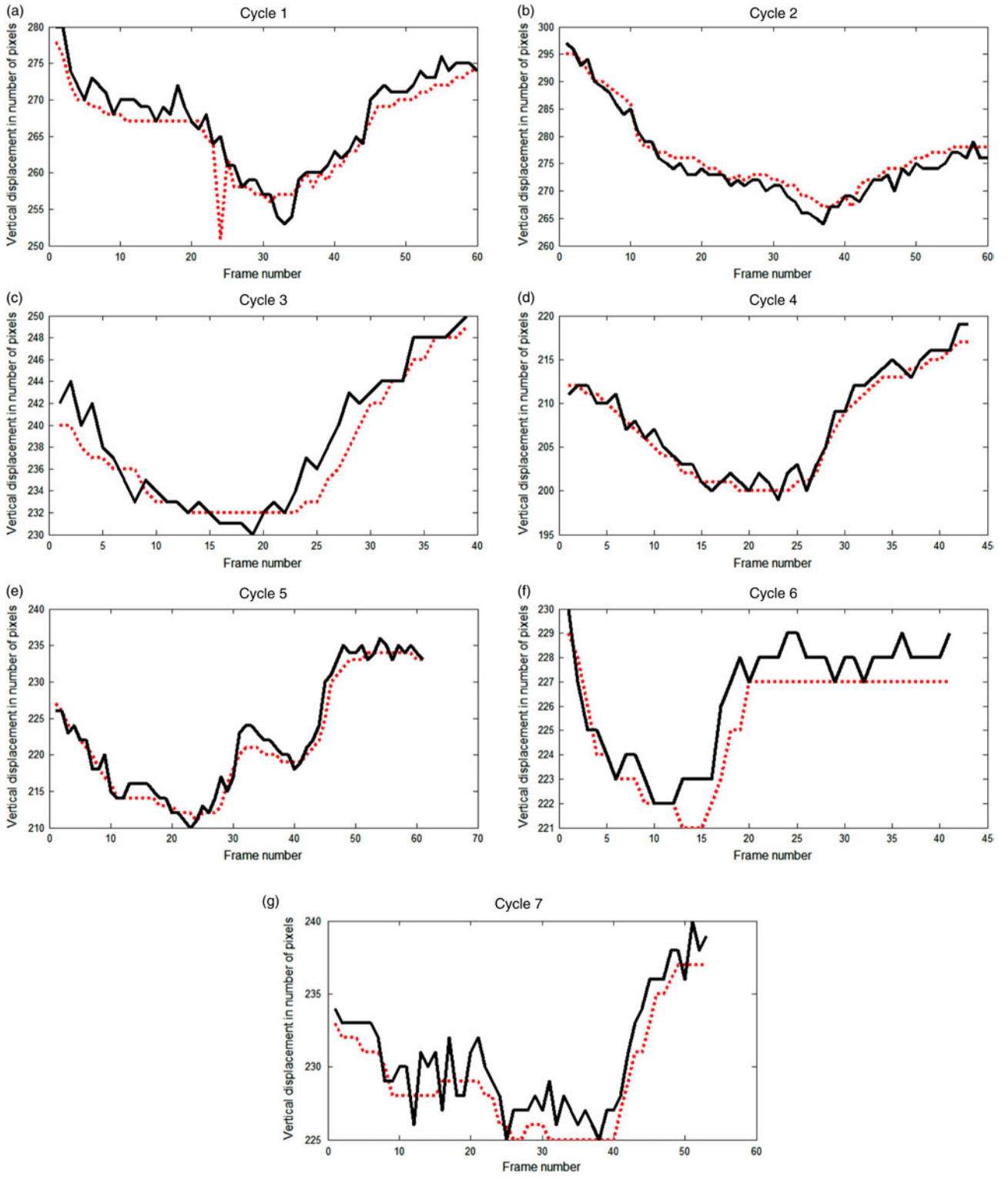


Figure 11. Comparison of proposed and manual methods for vertical displacement of the hyoid bone (cycles 1–7 in Figure 9). The horizontal axis represents the frame number and the vertical axis represents the vertical displacement of the hyoid bone obtained from both methods. Solid line represents the manual method and the broken line represents the proposed method.

Table 2. Mean square error and percentage error between proposed method and manual method for various cycles.

| | Mean square error | | Percentage error | |
|---|---|---|---|---|
| Cycle no. | Horizontal direction | Vertical direction | Horizontal direction | Vertical direction |
| 1 | 45.02 | 3.67 | 1.54 | 0.53 |
| 2 | 5.33 | 1.98 | 0.69 | 0.38 |
| 3 | 7.89 | 2.15 | 0.67 | 0.45 |
| 4 | 2.35 | 2.12 | 0.49 | 0.53 |
| 5 | 2.82 | 1.15 | 0.36 | 0.35 |
| 6 | 2.73 | 2.41 | 0.32 | 0.53 |
| 7 | 2.08 | 1.49 | 0.39 | 0.41 |
| Average | 9.75 | 2.14 | 0.64 | 0.45 |

vary in terms of shape and size among various patients. Also, the proposed method is not applicable where big movements are involved since they introduce motion blur to the images and template matching is sensitive to such blur.

## 5. Future work

In this work, the validation process is performed by one expert. In future work, the validation process should involve more experts. This will allow us to have better insight into the variation among the assessments obtained from various sources and the degree of assistance the proposed method provides to the radiologists.

In the proposed method, the movements are assumed to be limited to the sagittal plane. For the data used in this research, this assumption holds. However, the possibility of movements in the coronal plane cannot be eliminated. In future, the proposed method can be improved so that motion along the coronal plane can be detected as well.

The proposed method can be further improved by utilising techniques used to remove motion blur from images. The proposed method employs a heuristic-based approach to identify the hyoid bone where motion blur is involved. By removing motion blur from the images, this aspect can be reduced. The movement of the hyoid bone can be measured in the existing data by using the proposed method and these measurements can be useful while using the motion blur removal techniques.

## Acknowledgements

The authors would like to thank Professor Mandar Jog, the Director of the Movement Disorders Program, LHSC, and Ms Angela Roberts-South, speech-language pathologist, for providing them with the necessary medical details and for being their primary medical and data resource for this research. Special credit should go to Dr Donald Taves and the staff at the radiology department of Parkwood Hospital for their assistance in collecting the data that have been used in this research. They would also like to acknowledge the Parkinson Disease Society of Canada for funding the entire data acquisition through a grant to Dr Jog and Ms Roberts-South and for allowing them to use the data in their research.

## References

Aung M, Goulermas J, Hamdy S, Power M. 2010. Spatiotemporal visualizations for the measurement of oropharyngeal transit time from videofluoroscopy. IEEE Trans Biomed Eng. 57(2):432–441.

Aung M, Goulermas J, Stanschus S, Hamdy S, Power M. 2010. Automated anatomical demarcation using an active shape model for videofluoroscopic analysis in swallowing. Med Eng Phys. 32(10):1170–1179.

Chen Y, Barron JL, Taves DH, Martin RE. 2001. Computer measurement of oral movement in swallowing. Dysphagia. 16(2):97–109.

Coates C, Bakheit AM. 1997. Dysphagia in Parkinson's disease. Eur Neurol. 38(1):49–52.

Crary MA, Butler MK, Baldwin BO. 1994. Objective distance measurements from videofluorographic swallow studies using computer interactive analysis: technical note. Dysphagia. 9(2):116–119.

Daniels S, Schroeder M, McClain M, Corey D, Rosenbek J, Foundas A. 2006. Dysphagia in stroke: development of a standard method to examine swallowing recovery. J Rehabil Res Dev. 43(3):347–355.

Dyer J, Leslie P, Drinnan M. 2008. Objective computer-based assessment of valleculae residue – is it useful? Dysphagia. 23(1):7–15.

Foley NC, Martin RE, Salter KL, Teasell RW. 2009. A review of the relationship between dysphagia and malnutrition following stroke. J Rehabil Med. 41(9):707–713.

Freund Y, Schapire RE. 1997. A decision-theoretic generalization of on-line learning and an application to boosting. J Comput System Sci. 55(1):119–139.

Han TR, Paik NJ, Park JW. 2001. Quantifying swallowing function after stroke: a functional dysphagia scale based on videofluoroscopic studies. Arch Phys Med Rehabil. 82(5): 677–682.

Higo R, Nito T, Tayama N. 2005. Videofluoroscopic assessment of swallowing function in patients with myasthenia gravis. J Neurol Sci. 231(1–2):45–48.

Huang SH, Chu YH, Lai SH, Novak C. 2009. Learning-based vertebra detection and iterative normalized-cut segmentation for spinal MRI. IEEE Trans Med Imaging. 28(10): 1595–1605.

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

Kendall KA, Leonard RJ. 2002. Videofluoroscopic upper esophageal sphincter function in elderly dysphagic patients. Laryngoscope. 112(2):332–337.

Kendall KA, McKenzie S, Leonard RJ, Gonçalves MI, Walker A. 2000. Timing of events in normal swallowing: a videofluoroscopic study. Dysphagia. 15(2):74–83.

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

Leopold NA, Kagel MC. 1996. Prepharyngeal dysphagia in Parkinson's disease. Dysphagia. 11(1):14–22.

Leopold NA, Kagel MC. 1997. Pharyngo-esophageal dysphagia in Parkinson's disease. Dysphagia. 12(1):11–18.

Lienhart R, Kuranov A, Pisarevsky V. 2003. Empirical analysis of detection cascades of boosted classifiers for rapid object detection. In: Michaelis B, Krell G, editors. Pattern recognition. Vol. 2781. Berlin: Springer. p. 297–304.

Lienhart R, Maydt J. 2002. An extended set of Haar-like features for rapid object detection. In: Proceedings of the 2002 International Conference on Image Processing. Vol. 1. Rochester, NY: IEEE. p. 900–903.

Martin RE, Letsos P, Taves DH, Inculet RI, Johnston H, Preiksaitis HG. 2001. Oropharyngeal dysphagia in esophageal cancer before and after transhiatal esophagectomy. Dysphagia. 16(1):23–31.

Martin RE, Neary MA, Diamant NE. 1997. Dysphagia following anterior cervical spine surgery. Dysphagia. 12(1):2–8.

McCullough GH, Wertz RT, Rosenbek JC, Mills RH, Webb WG, Ross KB. 2001. Inter- and intrajudge reliability for videofluoroscopic swallowing evaluation measures. Dysphagia. 16(2):110–118.

Paik NJ, Kim SJ, Lee HJ, Jeon JY, Lim JY, Han TR. 2008. Movement of the hyoid bone and the epiglottis during swallowing in patients with dysphagia from different etiologies. J Electromyogr Kinesiol. 18(2):329–335.

Palmer JB, Kuhlemeier KV, Tippett DC, Lynch C. 1993. A protocol for the videofluorographic swallowing study. Dysphagia. 8(3):209–214.

Potulska A, Friedman A, Królicki L, Spychala A. 2003. Swallowing disorders in Parkinson's disease. Parkinsonism Relat Disord. 9(6):349–353.

Robbins J, Coyle J, Rosenbek J, Roecker E, Wood J. 1999. Differentiation of normal and abnormal airway protection during swallowing using the penetration–aspiration scale. Dysphagia. 14(4):228–232.

Rosenbek JC, Robbins JA, Roecker EB, Coyle JL, Wood JL. 1996. A penetration–aspiration scale. Dysphagia. 11(2): 93–98.

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

Smithard DG, O'Neill PA, England RE, Park CL, Wyatt R, Martin DF, Morris J. 1997. The natural history of dysphagia following a stroke. Dysphagia. 12(4):188–193.

Stephen J, Taves D, Smith R, Martin R. 2005. Bolus location at the initiation of the pharyngeal stage of swallowing in healthy older adults. Dysphagia. 20(4):266–272.

Stoeckli SJ, Huisman TAGM, Seifert BAGM, Martin-Harris BJW. 2003. Interrater reliability of videofluoroscopic swallow evaluation. Dysphagia. 18(1):53–57.

Stroudley J, Walsh M. 1991. Radiological assessment of dysphagia in Parkinson's disease. Br J Radiol. 64 (766):890–893.

Suh MK, Kim H, Na DL. 2009. Dysphagia in patients with dementia: Alzheimer versus vascular. Alzheimer Dis Assoc Disord. 23(2):178–184.

Viola P, 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. Vol. 1. Kauai, HI: IEEE. p. 511–518.

Warabi T, Ito T, Kato M, Takei H, Kobayashi N, Chiba S. 2008. Effects of stroke-induced damage to swallow-related areas in the brain on swallowing mechanics of elderly patients. Geriatr Gerontol Int. 8(4):234–242.