Bolus Inflow Detection Method by Ultrasound Video Processing for Evaluation of Swallowing

Takato MATSUZAKI,* Yutaka SUZUKI,*, # Morimasa TANIMOTO,** Keisuke MASUYAMA,*** Masashi OSANO,† Osamu SAKATA,†† Masayuki MORISAWA*

Abstract To prevent aspiration pneumonia, a system for non-invasive and quantitative evaluation of the swallowing function is required. Therefore, we have previously proposed a method of using ultrasound videos to establish evaluation indicators of the swallowing function. The proposed method aims to automatically estimate the velocities of the esophageal wall region and the bolus region in the ultrasound video. In this method, estimation of the bolus region comprises two steps: estimating the esophageal region through which the bolus flows and extracting only the frame in which the bolus passes through the esophageal region. However, the step of extracting the frame in which the bolus passes is still performed manually. Therefore, to automate this step, the purpose of this study was to automatically determine the frame in which the bolus flowed into the screen. This method was tested five times on five healthy adult male subjects by recording a cervical ultrasound video while swallowing a bolus of water. We identified the different elements of the esophageal region in the video by first identifying the esophageal wall region with the maximally stable extremal regions (MSER). Then, we used the luminance histogram of each frame to establish the graph of the histogram similarity. This, in turn, was used to detect a change in the observed region, thus indicating the inflow of the bolus. Moreover, we could distinguish the change caused by the inflow from the change caused by the elevation of the esophageal wall using the velocity results obtained by optical flow estimation in the anterior esophageal wall region. Our results showed that in most cases, the proposed method was successful in recognizing the inflow of the bolus and distinguishing it from the elevation of the esophageal wall. Furthermore, an accuracy sufficient for estimation of the velocity of the bolus was achieved.

Keywords: ultrasound video, swallowing, optical flow, esophagus, bolus, maximally stable extremal regions.

Adv Biomed Eng. 10: pp. 18–25, 2021.

1. Introduction

In the super-aged Japanese society, pneumonia is the third leading cause of death, and most cases of pneumonia amongst the elderly are cases of aspiration pneumonia. The number of patients suffering from aspiration pneumonia is increasing with the growing elderly proportion, which saturates the hospital capacity. Therefore, it is necessary to prevent aspiration pneumonia and reduce the number of patients. There is a high demand for a method to prevent aspiration pneumonia by quantitative and non-invasive measurement of the swallowing ability and by providing suitable food for the patient’s swallowing ability. Currently, video fluorography (VF) [1–5], video endoscopy (VE) [6–8], and food testing are the main methods of measuring swallowing ability. However, VF and VE are invasive methods and food testing [9] is not a quantitative evaluation.

Therefore, we have previously attempted to establish a method for non-invasive and quantitative measurement of the swallowing function using B-mode ultra-
sound videos [10–12]. VF can be used mainly to observe the flow of the bolus from the oral stage to the pharyngeal stage and the esophageal stage, and VE can be used mainly to observe the movement of the epiglottis in the pharyngeal stage and the bolus remaining in the pharynx after swallowing. In contrast, we focused on the esophageal stage and the movement of the bolus using ultrasound videos, and attempted to extract quantitative indicators for evaluation of the swallowing function. An existing method analyzes tongue movement as an evaluation of swallowing using ultrasound [13]. On the other hand, we focused on the movement of the esophagus and the flow characteristics of the bolus. In addition, a study [14] has investigated the flow velocity of the bolus in the pulse Doppler mode by fixed position observation. However, we expected improved accuracy in swallowing evaluation by analyzing additional information using B-mode moving images. In our previous study [10], we calculated the estimated amount of movement of the bolus and esophageal wall, which are considered to be effective as evaluation indicators of the swallowing function. This method involves estimating the esophageal wall region and the bolus region in the ultrasound video, and then calculating the estimated amount of movement by optical flow estimation for each pixel. Finally, the estimated amount of movement at the pixels of the esophageal wall region and the bolus region is extracted. However, in this method, the esophageal region is first estimated as the esophageal region through which the bolus flows, and among all the frames of the video containing the esophageal regions, only those in which the bolus passes is extracted as the bolus region. That is, in order to estimate the bolus region, it is indispensable to estimate the esophageal region and the frames in which the bolus passes. However, in the previously proposed method [10], there is no system that automatically determines the frame in which the bolus passes. Therefore, the frame in which the bolus flows into the moving image and the frame in which the entire bolus passes are determined by visual observation. Thus, the purpose of this study was to automatically determine the frame in which the bolus flows into the screen in order to automate a part of the previously proposed method [10].

2. Methods

2.1 Recording

A FUJIFILM FC1 (B-mode) was used for the recording. We used a linear probe HF-50xp (frequency band 6 to 15 MHz). We recorded the videos at a frame rate of 30 fps. The relationship between the actual length of the measured object and the number of pixels in the video was 133 pixels/cm. Figure 1 shows an ultrasound image of the esophageal area during swallowing. A cross-section of the tubular esophagus is shown, and the anterior esophageal wall is displayed brightly and horizontally at the center of the screen. The element appearing underneath is the bolus, and at the bottom, the posterior esophageal wall is visible. The measurements were performed by excluding the esophageal sphincter and focusing on the movement of the esophagus and bolus.

2.2 Data acquisition

The probe was placed so as to avoid the subject’s thyroid cartilage and at a position where the esophagus could be observed with the FC1 (on the right in Fig. 2). The subject was seated at a 30° angle to the vertical (on the left in Fig. 2). Five healthy adult male subjects participated in the study. Each subject swallowed a bolus consisting of 5 ml of drinking water, five times. We instructed the subjects to swallow the drinking water in sips. The measurement time of moving images was 5 s per data sample. Table 1 lists the height and weight of each subject. The experiment was conducted with consent from all subjects in accordance with the Declaration of Helsinki.

![Fig. 1 Pharyngeal ultrasound image during swallowing.](image1)

![Fig. 2 Subject’s sitting position and probe position.](image2)
2.3 Data processing

Figure 3 illustrates the overall processing flow, which can be divided into three stages.

The first stage was the extraction of the bolus and the esophageal wall region and estimation of the velocity of the anterior esophageal wall. Here, we first cut out the region of interest (ROI) from the measurement video and extracted the area around the esophageal region alone. The processing after this ROI extraction step was performed automatically. The esophageal wall and the esophageal region were identified. In parallel, the amount of movement of each pixel was calculated by optical flow estimation. Only the x-direction velocity in the anterior esophageal wall region was determined. The x-direction was the direction from the mouth to the stomach (from left to right in Fig. 1). A detailed description of this stage is presented in Section 2.3.1.

In the next stage, we calculated the histogram similarity for each frame. To do so, only the leftmost 30-pixel width area of the esophageal region was used. Then, the luminance histogram of each frame was generated in this area. Next, the histogram similarity was calculated in comparison with the preceding and following frames. We plotted a graph representing the histogram similarity of each frame to easily identify a change in similarity between the frames. A detailed explanation of the method is presented in Section 2.3.2.

Finally, we determined the bolus inflow frame and distinguished it from the elevation of the esophageal wall. The moment of the elevation of the esophageal wall was deduced from the change in velocity of the anterior esophageal wall in the x-direction. We identified the frame representing the end of elevation of the esophageal wall by identifying the frame after the peak velocity (representing the elevation of the esophageal wall) in which the x-direction velocity was back to normal. The bolus inflow frame was the first frame after the end of wall elevation, which had a local minimum value in the histogram similarity graph. Detailed explanations are provided in Section 2.3.3.

2.3.1 Identification of the esophagus and esophageal wall

The method for estimating the esophageal wall region and the esophageal region described in this section was the same as the method proposed previously [10]. First, we cut out the ROI from the ultrasound video. Next, we estimated the amount of movement (optical flow) in all pixels. For this, we used the Farnebäck method [15, 16], which is more accurate than other methods such as Lucas-Kanade method [17] and Horn-Schunck method [18]. In addition, the regions in the image with high
brightness, corresponding to the anterior and posterior esophageal walls, were identified by MSER [19] independent of the optical flow estimation. Then, using the method proposed previously [10], we identified the esophageal region by performing threshold processing based on the positional relationship of each tissue on the screen. Figure 3 shows a schematic diagram of each of the identified tissues. In Fig. 4, \((p_1, q_1)\) and \((p_2, q_2)\) indicate the central coordinates of the esophageal wall region detected in an elliptical shape by MSER, and \(\theta_1\) and \(\theta_2\) represent the inclination of the esophageal wall region. The boundaries \(y_1\) and \(y_2\) of the esophageal region are determined based on the detected esophageal wall region. \(\alpha\) and \(\beta\) are constants for the boundary line to be located inside the esophageal wall region. The esophageal region \(y\) is defined by \(y_1 < y < y_2\). However, since the esophagus is expected to appear like a band on the screen, the range of inclination of \(y_1\) and \(y_2\) is limited between \(-\pi/60\) and \(\pi/60\).

We managed to identify the different parts of the esophageal region successfully using our method, as shown in Fig. 5. The areas marked by the green ellipses are the areas identified as the esophageal walls, and the area between the yellow and red lines is the esophageal region. When the recordings were clear enough to be exploited, the leftmost area of the image was used to calculate the histogram similarity.

### 2.3.2 Calculation of histogram similarity graph

In the esophageal region identified in Section 2.3.1, we extracted a region 30 pixels in width on the left side, which was the location from which the bolus entered the esophagus, and created a luminance histogram with 256 dimensions in each frame. Then, we determined the similarity of the histogram by comparing with the histograms in the preceding and following frames. Equation (1) is used to calculate the histogram similarity \(S\).

\[
S(H_1, H_2) = \frac{\sum (H_1(I) - \bar{H}_1)(H_2(I) - \bar{H}_2)}{\sqrt{\sum (H_1(I) - \bar{H}_1)^2} \cdot \sqrt{\sum (H_2(I) - \bar{H}_2)^2}}
\]

Here, \(H_1, H_2\) are the histograms of the preceding and following frames, respectively. The histogram similarity takes a value from 0 to 1, and the closer the value is to 1, the more similar are the histograms. A significant decrease in the histogram similarity value indicates that the area on the left-end of the bolus region has changed significantly between the preceding and following frames. Therefore, when the graph presenting the histogram similarity for each frame is plotted, the inflow of the bolus can be detected by identifying the local minimum value.

### 2.3.3 Distinguishing the timing of esophageal wall elevation and bolus inflow

Using the process described in Section 2.3.2, we observed that the histogram similarity value decreased significantly during two events: elevation of the esophageal wall and inflow of the bolus. Elevation of the esophageal wall here refers to the movement in preparation for the bolus to flow into the esophagus, and the esophageal wall moves in the direction of the head in conjunction with the pharynx before the bolus passes. In healthy individuals, the order of the two events is elevation of the esophageal wall followed by inflow of bolus. To detect the bolus inflow timing, it is thus necessary to distinguish these two moments. We visually observed the timing of elevation of the esophageal wall and the timing of inflow of bolus in further detail, and found that each event occurred at the timing shown in Fig. 4. The elevating movement of the esophageal wall can be divided into three periods: the beginning of movement, the most active period, and the period when the magnitude of movement decays and the movement ends. We found that one of the two events that exhibited a large response in histogram similarity in Section 2.3.2 was the period of the most active movement of the esophageal wall elevation. Further, we found that the timing of the bolus inflow was during the period when the movement of the esophageal wall elevation decayed and the movement ended.

We searched the array \(v_x\), which stored the estimated average velocity of the anterior esophageal wall region in

---

**Fig. 4** Overview of the esophageal wall and the bolus identification.

**Fig. 5** Image of a frame after the identification of the different elements of the esophageal region.
each frame, for the smallest frame number \( k \) in which the esophageal wall moved the most in the direction of the head. The movements in this direction are represented by negative velocities and \( k \) is defined by the following equation:

\[
k = \text{argmin}(v_x)
\]

For every frame after frame \( k \), we defined an array \( L_{vx}(n) \) that stored all local minimum velocity values and an array \( I_{vx}(n) \) that stored the frame numbers of each value of \( L_{vx}(n) \) on the original video. Here, \( n \) is a variable that indicates each array index of \( L_{vx}(n) \), which is an array that stores the local minimum value, and \( n \) is different from the frame number in the original video. Equation (2) was used to search for the most active period of esophageal wall elevation movement in Fig. 6. We then found the smallest element number \( m \) that satisfies the following equation:

\[
m = \text{argmin}_n\{L_{vx}(n−1) < L_{vx}(n) < L_{vx}(n+1)\}
\]

(3)

Here, \( m \) is the array index storing the timing at which the velocity is gradually attenuated in the local minimum value array \( L_{vx}(n) \). This indicates the period when the movement of the esophageal wall elevation decays and the movement ends, as shown in Fig. 6. As the value of \( m \) determined by equation (3) is an array index, the frame number in the corresponding original image is \( I_{vx}(m) \). Therefore, we determined the bolus inflow frame to be the smallest frame number with a local minimum in the histogram similarity calculated in Section 2.3.2 for frames after frame number \( I_{vx}(m) \).

3. Results

Figure 7 shows an example of the estimated x-direction velocity in the anterior esophageal wall region (UEW_x-velocity in Fig. 7) detected from the moving image recorded in subject A. The horizontal axis shows the frame number of the moving image and the numerical value obtained by converting the frame number into real time. The movements in the direction of the head are represented by negative velocities. Elevation of the anterior esophageal wall in the head direction was approximately in the 40th frame. Figure 7 also plots the local minimum value of the estimated velocity in the x-direction (min_vx), which is \( L_{vx} \) in equation (3). The frame number fitting equation (3) was frame 44. The 44th frame marked the beginning of the period when the movement of the esophageal wall elevation decayed and the movement ended, as shown in Fig. 6. Figure 8 shows the graph of histogram similarity score for each frame in the extracted bolus inflow region using the same set of data as the one used to plot Fig. 7. Histogram similarity could not be recorded before frame 25, and was recorded only after that, when the esophageal region could be stably identified. The similarity score decreased significantly around the 40th frame (Fig. 8). However, frame 40 was
still the most active period of esophageal wall elevation movement as shown in Fig. 6. In fact, the velocity of the anterior esophageal wall reached the minimum value at the 40th frame. Thus, the esophageal wall elevation around the 40th frame was not considered. Therefore, in Fig. 8, the frame finally detected as the bolus inflow was a local minimum indicated by the red line at the 44th frame. We visually observed and recorded the timing of the inflow of the bolus for comparison, and observed the inflow at frame 44. No difference was observed between the results of the processing method and those of the visual observation method.

Then, we compared the difference in results among the subjects. Table 2 shows the error between the frame detected by the proposed method and the visually observed frame for the five subjects. The seven crossed-out cells in Table 2 correspond to the measurements for which the histogram similarity was not calculated because the leftmost 30 pixel-wide area could not be extracted when the esophageal region was identified. Eighteen of the 25 recordings produced usable data, and it was also possible to obtain exploitable data for three or more measurements for every subject. This accuracy can be considered sufficient to apply the subsequent process to identify the esophageal wall and the bolus. Only the third measurement of subject D showed a significant error between the frame detected by the proposed method and the actual bolus inflow. This was caused by $I_{oc}(m)$ detecting a distinctly different frame. Excluding this error, there was an average difference of ±2 frames between the automatically detected frame and the manually confirmed frame.

| Subject | 1st  | 2nd | 3rd  | 4th  | 5th |
|---------|------|-----|------|------|-----|
| A       | +2f  | 0f  | −2f  |      |     |
| B       | +1f  | −1f | +1f  | −2f  |     |
| C       | −4f  | −2f | 0f   |      |     |
| D       |      | +3f | +40f | −1f  |     |
| E       |      | −2f | +3f  | +1f  | −3f |

Table 2 Frame detection error for the different subjects.

In addition, one of the reasons why some recordings were unusable is the fact that the posterior esophageal wall was not reflected clearly, and therefore the region could not be identified. It is necessary to reduce the percentage of the unusable data. Hence, there is a need for a system that assists in the measurement and identification when the appearance of the tissue is clear, as presented in a report [20]. Furthermore, there is a need to improve the detection accuracy of the entire system, from the measurement to the analysis. The development of a system that detects and eliminates data in which the appearance of the tissue is not clear could be useful to solve this problem.

In the future, we plan to automatically extract the bolus velocity using this method. In addition, we plan to calculate the opening of the esophageal wall and investigate the correlation between the swallowing function and the estimated indicators: the velocities of the esophageal wall and the bolus and the opening of the esophageal wall. Furthermore, as a non-invasive measurement method, there also exists myoelectric potential measurement using surface electrodes. Comprehensive evaluation combined with methods such as evaluation of the swallowing function based on EMG signals during swallowing [21] is considered to be effective.

5. Conclusion

We aimed to automate the method of estimating the position and velocity of the esophageal wall and bolus in an ultrasound video proposed in our previous study. Therefore, we proposed a method that can automatically determine the frame in which the bolus flows into the screen. We tested this method five times on five different subjects to obtain a good sample for analysis. We were able to automatically detect the inflow of the bolus in most of the cases with an acceptable margin of error. However, this method can be further improved as there are still errors in the identification of the different tissues in the video. Therefore, there is a need to develop methods that can help reduce these errors throughout the process.
from the recording to the analysis.

Conflicts of Interest

The authors declare no conflicts of interest.

References

1. Tokifuji A, Matsushima Y, Hachisuka K, Yoshioka K: Texture, sensory and swallowing characteristics of high-pressure-heat-treated pork meat gel as a dysphagia diet. Meat Sci. 93(4), 843–848, 2013.
2. Schatz K, Langmore SE, Olson N: Endoscopic and videofluoroscopic evaluations of swallowing and aspiration. Ann Otol Rhinol Laryngol. 100(8), 678–681, 1991.
3. Sellars C, Campbell AM, Stott DJ, Stewart M, Wilson JA: Swallowing abnormalities after acute stroke: a case control study. Dysphagia. 14(4), 212–218, 1999.
4. Takahashi N, Kikutani T, Tamura F, Groher M, Kuboki T: Videoendoscopic assessment of swallowing function to predict the future incidence of pneumonia of the elderly. J Oral Rehabil. 39(6), 429–437, 2012.
5. Bastian RW: Videoendoscopic evaluation of patients with dysphagia: an adjunct to the modified barium swallow. Otolaryngology-Head and Neck Surgery. 104(3), 339–350, 1991.
6. Smith HA, Lee SH, O’Neill PA, Connolly MJ: The combination of bedside swallowing assessment and oxygen saturation monitoring of swallowing in acute stroke: a safe and humane screening tool. Age Ageing. 29(6), 495–499, 2000.
7. Ozaki K, Kagaya H, Yokoyama M, Saitoh E, Okada S, González-Fernández M, Palmer JB, Uematsu H: The risk of penetration or aspiration during videofluoroscopic examination of swallowing varies depending on food types. Tohoku J Exp Med. 220(1), 41–46, 2010.
8. Yamamoto T, Ikeda K, Usui H, Miyamoto M: Validation of the Japanese translation of the Swallowing Disturbance Questionnaire in Parkinson’s disease patients. Qual Life Res. 21(7), 1299–1303, 2012.
9. Osawa A, Maeshima S, Tanahashi N: Food and liquid swallowing difficulty in stroke patients: a study based on the findings of food tests, a modified water swallowing test and videofluoroscopic examination of swallowing. Jpn J Rehabil Med. 49(11), 838–845, 2012.
10. Suzuki Y, Sakata O, Tanimoto M, Hatsushika K, Masuyama K, Morisawa M: Evaluation of swallowing capacity based on esophageal and bolus movements by ultrasound video processing. Proc of Tenth International Conference on Signal Processing Systems. Vol. 11071, Singapore, 2018.
11. Suzuki Y, Tanimoto M, Hatsushika K, Masuyama K, Sakata O, Morisawa M: Classification of esophageal and bolus movements for evaluation of swallowing function by ultrasound video processing using optical flow and stable extremal region method. IIEJ Trans Electron Inf Syst. 139(5), 603–608, 2019.
12. Matsuzaki T, Suzuki Y, Tanimoto M, Masuyama K, Takahashi M, Arakawa M, Sakata O, Morisawa M: Determining the influence of viscosity on quantitative indicators of swallowing function using ultrasound video processing. Proc of IEEE 9th Global Conference on Consumer Electronics. Osaka, pp. 913–917, 2019.
13. Tomi Y, Uehara T, Torii T, Matsuoka H, Toyoda K, Minematsu K: Tongue and oral function test with ultrasonography (TOFUS): A dynamic ultrasound imaging system for swallowing studies in stroke patients. Neurosonol. 23(1), 5–8, 2010.
14. Hasegawa A, Ootoguro A, Kumagai H, Nakazawa F: Velocity of swallowed gel food in the pharynx by ultrasonic method. Nippon Shokuhin Kagaku Kogaku Kaishi. 52(10), 441–447, 2005.
15. Farnebäck G: Two-frame motion estimation based on polynomial expansion. 13th Scandinavian Conference on Image Analysis. Vol. 2749, pp. 363–370, 2003.
16. Farnebäck G: Polynomial expansion for orientation and motion estimation. Diss. Linkoping University Electronic Press Vol. 790, 2002.
17. Lucas B, Kanade T: An iterative image registration technique with applications to stereo vision. Proc of Darpa IU Workshop. pp. 121–130, 1981.
18. Horn BKP, Schunck BG: Determining optical flow. Artif Intell. 17(1–3), 185–203, 1981.
19. Matas J, Chum O, Urban M, Pajdla T: Robust wide-baseline stereo from maximally stable extremal regions. Image Vision Comput. 22(10), 761–767, 2004.
20. Matsuzaki T, Suzuki Y, Tanimoto M, Masuyama K, Takahashi M, Sakata O, Morisawa M: Examination of image processing method and image selection method for region of interest extraction in an ultrasound moving image for swallowing evaluation. J Signal Process. 23(6), 285–291, 2019.
21. Suzuki M, Sasaki M, Kamata K Nakayama A: Swallowing pattern classification method using multichannel surface EMG signals of suprathyroid and infrathyroid muscles. Adv Biomed Eng. 9, 10–20, 2020.

Takato Matsuzaki

Takato Matsuzaki received the B.E., from Univ. of Yamanashi, Japan, in 2019. He is now M.E degree student, in Graduate Faculty of Interdisciplinary Research, University of Yamanashi.

Yutaka Suzuki

Yutaka Suzuki received the B.E., M.E., and Ph.D. degrees from Univ. of Yamanashi, Japan, in 2003, 2005 and 2008, respectively. He had been an Assistant Professor at Center for Life Science Research, Univ. of Yamanashi, Japan in 2008–2014. Since 2020, he is currently an Associate Professor at faculty of Engineering, Univ. of Yamanashi, Japan. He is a member of IEEE, IEICE, JSMBE, JSWE and JSKE.

Morimasa Tanimoto

Morimasa Tanimoto received the B.S. and E.M. degrees from Hokkaido University, Japan, 1978 and 1980 respectively. He had worked at a Japanese Food Company from 1980 to 2011. He received the Ph.D. degree from Tohoku University, Japan, 1998. Currently, he is a Professor in the Department of Food Science at Tokyo Seiei College since 2020.
Keisuke Masuyama
Keisuke Masuyama completed School of Medicine, Kumamoto Univ., Japan in 1979. He was a Lecturer at Kumamoto Univ. in 1989–1994. He was an Associate Professor at the Dep. of Oto-rhino-laryngology, Kumamoto University in 1995–2002. He was a Professor at Univ. of Yamanashi in 2003–2019 and currently works at Suwa Central Hospital, Japan. He received MD and PhD. He is a member of ORLAJ and Japan Otological Society.

Masashi Osano
Masashi Osano did his residency at the University of Yamanashi Hospital in 2009 (Department of Otolaryngology). In 2011, he became a senior resident for otolaryngology, head, and neck surgery at the University of Yamanashi Hospital. In 2016, he became an assistant professor at the University of Yamanashi Hospital.

Osamu Sakata
Osamu Sakata received the B.S., M.S, and Ph.D. degrees from Univ. of Tsukuba in 1996, 1998, and 2001, respectively. He is currently an associate professor in Tokyo Univ. of Science, after having served as an associate professor in Univ. of Yamanashi. He is a member of IEEJ, IEICE and IEEE.

Masayuki Morisawa
Masayuki Morisawa received a BSc degree, 1987 and MSc degree, 1989 in Electrical Engineering at Yamanashi University, Japan and PhD in 2002 in Electrical Engineering at Tohoku Univ., Japan. He is now Professor at faculty of Engineering, Univ. of Yamanashi, Japan.