

---

# Dynamic Recognition and Tracking of Barium Flow Field Based on Deglutition Video

---

Guofeng Qin<sup>1,\*</sup>, Jianhuang Zou<sup>1</sup>, Qiufang Xia<sup>2,\*</sup> and Jiahao Qin<sup>1</sup>

<sup>1</sup>*Department of Computer Science and Technology, Tongji University, Shanghai, China*

<sup>2</sup>*Shanghai first rehabilitation hospital, Shanghai, China*

*E-mail: gfqing@tongji.edu.cn; 13651899419@163.com*

*\*Corresponding Author*

Received 04 December 2020; Accepted 24 December 2020;  
Publication 09 March 2021

## Abstract

Dynamic fluoroscopy was used to study swallowing in 84 adult patients. We proposed a method to extract the barium contrast region by improved interframe difference method, and to indirectly determine the position of epiglotic cartilage and cricopharyngeal muscle according to the location of barium meal. The method is easy to understand, and the extraction effect is good, with 85% probability of successful extraction. On the other hand, in order to evaluate the degree of deglutition difficulty, we used calculation to evaluate variables including displacement, duration, residual quantity, etc., except that there were gender differences in variables and external factors, such as illumination, most of the measurement variables had very good reliability. The experimental results showed that the moving target fluid barium was extracted by quantifying dynamic fluorescence deglutition and using gaussian based background subtraction algorithm. We conclude that this approach significantly reduces the time it takes clinicians to examine moving images. This paper describes how to study swallowing disorders

*Journal of Web Engineering, Vol. 20.2, 533–556.*

doi: 10.13052/jwe1540-9589.20214

© 2021 River Publishers

by X-ray barium fluoroscopy, explains the application of interframe difference algorithm and background subtraction in deglutiography, and extracts the residual amounts in three locations: oral cavity, epiglottic cartilage and piriform fosse.

**Keywords:** Dynamic fluoroscopy, X-ray barium fluoroscopy, swallowing disorders, interframe difference algorithm, background subtraction algorithm.

## 1 Introduction

Swallowing is not an arbitrary activity, but a complex reflex involving simultaneous and sequential contraction of the muscles of the mouth, pharynx, larynx, and esophagus [1–4]. Dysphagia is an eating disorder known to result from the smooth functioning of the jaw, lips, tongue, soft palate, throat, esophageal mouth sphincter, or esophagus. Typical manifestations are: (1). Mouth control ability and food chewing ability are weakened, swallowing reflex is delayed; (2) after swallowing, there is residual food left in the pharynx; (3) before, during or after swallowing, residual food is inhaled into the organ [5].

The study of swallowing X-ray fluoroscopy (barium fluoroscopy) is the “gold standard” of dysphagia [6]. Therefore, in order to detect whether there is the phenomenon of aspiration during swallowing, often in together with the clinical doctors in professional barium perspective tests of barium swallow instrument, observe the upright position, different concentrations of barium swallowing mouth and esophagus, to remove the potential impact of swallowing function of the esophagus, and whether there is any aspiration to hide organ dysfunction of the trachea.

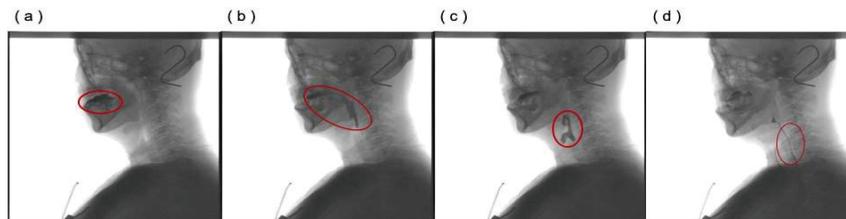
The pharyngeal period is a critical period in the swallowing process, and abnormalities can cause serious medical complications, such as aspiration pneumonia or asphyxia [7]. Dysphagia occurs in 30%~50% of the elderly (60 years old), 30%~65% of stroke patients, and 40%~90% of Parkinson’s patients [8, 9].

There are three stages of barium X-ray fluoroscopy that can be left:

- Barium retention. The condition of the accumulation of contents in the epiglottic valley and pyriform fosses before swallowing;
- Barium residue. The condition of the contents remaining in the epiglottic valley and the pyriform fosse after swallowing;
- Barium spilt. A condition in which the contents of the epiglottic valley and pyriform fosses accumulate beyond their capacity to overflow.

Among them, the vertical movement in the pharyngeal esophageal segment (PE segment) and the rapid change of pressure in this segment made the measurement complicated. Under normal circumstances, clinicians will conduct screening and diagnosis of dysphagia through repeated observation of images and videos. However, due to the previous coordination between patients and doctors, a large number of swallowing videos are redundant. Therefore, how to assist physicians in making judgments is a very important research content.

Normal swallowing is divided into four stages: oral preparation, oral, pharyngeal and esophageal [10]. The X-ray attenuation coefficient of each voxel was obtained by scanning the upper part of the body, which was arranged into a matrix to form a grayscale map. Different grayscale reflected the X-ray absorption degree of organs and tissues. Those with high bone were white, while those with black, and some soft tissues were gray. In order to facilitate observation and detection, clinicians will contrast the two when they are examined. As shown in the sample in this paper, the background is gray, the head, including the cervical vertebra and hyoid bone are gray in different degrees, the pharynx is black due to the shielding of clothing, and the barium agent is gray in different concentrations and volumes.



**Figure 1** (a) Oral preparatory phase. Fluid barium is injected into the mouth through a syringe (b) Oral phase (<math>1\sim 1.5</math>). Trigger swallowing reflex (c) Pharyngeal phase (<math>< 1</math> s). trigger ring pharynx muscle (d) Esophageal phase. The pharyngeal contraction, the fluid extract into the esophagus.

However, due to the lack of chewing process based on the swallowing observation of fluid barium, and the swallowing speed of fluid barium is ten times faster, usually  $1\sim 2$  s fluid through, so it is crucial to observe the flow of barium in stages, tracheal aspiration and observe the change of residual quantity in this experiment.

In addition, obtain VFSS video has redundancy, sample quality and difficult to unity. Lateral views of the head and neck area were recorded by X-ray fluoroscopy throughout the VFSS procedure, and patients developed

different symptoms, such as difficulty swallowing barium at first and needing the help of a syringe and physician. Therefore, for this kind of patients, not only do we need to have enough patience, but also the inappropriate shooting time will lead to too many samples of the patients, resulting in a certain degree of redundancy in the study.

## 2 Background and Motivation

Dysphagia is an increasingly serious problem in the aging of China's population [11]. Frame-by-frame analysis lacks the reliability of biomechanics in deglutition, and clinicians usually rely on their own empirical analysis for deglutition abnormalities without accurate data verification. That is, imaging diagnosis is a subjective interpretation based on visual examination, and there is no accurate standard judgment tool.

Through video fluoroscopy, a number of studies have been conducted on the correlation between the clinical indicators of dysphagia and the actual presence of dysphagia. Studies have found that the position of the stationary superior esophageal sphincter is closely corresponding to the position of the cricopharyngeal muscle, and deglutition usually causes the anterior and upper displacement of the hyoid bone, which leads to the elevation of the posterior part and the opening of the superior esophageal sphincter, which is the key to judge the changes in the swallowing stage [12, 13].

Swallowing is the process of swallowing in a fast sequence of seconds or less. Developing and analyzing software tools for studying VFSS in order to objectively assess the data from the swallowing process was a key breakthrough. The clinician examines the video and manually demarcates the area of interest, the area through which fluid barium has passed. Similar studies have used interactive methods, in which clinicians manually adjust filters to adapt the sharpness of the image, and enhance the contrast through video enhancement, and then use a computer to determine the reference position. However, manually marking ROI areas requires a certain amount of specialization and takes time [14]. Open source software was used to identify the relationship between parameters and the swallowing process based on 13 quantitative physiological variables related to swallowing, including swallowing duration and swallowing interval. However, in clinical practice, the calculation of this kind of data has many interfering factors, and its accuracy is easily affected by external factors.

For the research related to swallow angiography, a method based on deep neural network has been proposed abroad. This method uses a three-layer

cascade framework [15]. The framework includes three stages: First, the optical flow method is used to generate the candidate set. Then use the inflatable 3D convolutional network to train the candidate set, and finally use the sliding window method to obtain the time series of the pharyngeal phase. This study can extract the swallowed pharyngeal phase time series in VFSS, but the direct extraction of the time series does not study the swallowing performance, and the practicability is low.

Domestic for swallowing imaging study towards medical practical, in the third affiliated hospital of zhongshan university rehabilitation medicine group leader zu-lin dou and others with wei soft technology co., LTD. Jointly developed soft star swallowing function digital image acquisition and analysis system, to meet the requirements of the analysis of 30 frames/s collection procedures feeding for 6 pair of target image, including the starting point, the maximal displacement of hyoid swallowing, etc., and by using statistical analysis, using the Spearman correlation coefficient analysis correlation between parameters. Semi-quantitative analysis has certain reliability, but people need technicians to obtain the target image [16].

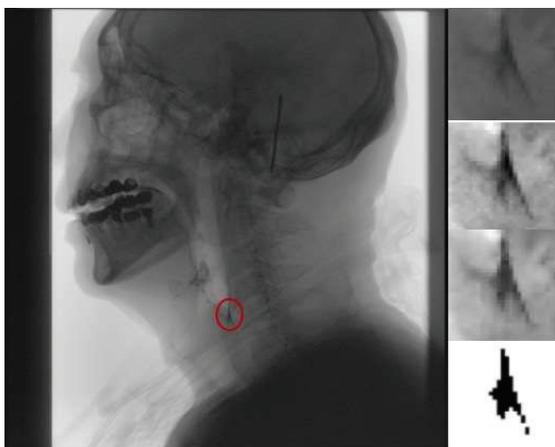
The detection of fluid barium agent can involve the detection of moving objects in sequence images, the detection of changing regions and the extraction of background regions. The difficulties in the detection of moving objects in deglutigraphy are as follows: 1. The moving objects are irregular, which can only be viewed from the perspective, so it is impossible to measure the volume of fluid barium and judge the specific behavior of barium. 2. Dynamic background. The dynamic background is no longer a single wind blowing, but the overall movement of the whole upper body in order to match the swallowing of barium. 3. Noise interference. X-ray noise interference has no fluoroscopic effect on bones, such as false teeth, hyoid bone, cervical vertebra, etc., so it will be displayed in the video. Among them, the false teeth, hyoid bone brings certain interference to the oral motion, and the cervical vertebra has certain usefulness to the judgment of pharyngeal organs.

### **3 Works and Method**

Prospective, continuous experimental data. Thanks to the help of relevant professors in the affiliated hospital of tongji university, we collected the experimental data of related patients in two times, namely 55 cases (2018.07–2019.04) and 39 cases (2019.04–2019.10). These patients were referred to the department of stomatology which admitted that the swallowing ability was suspected to be dysphagia. They included stroke patients (specialist patients

and outpatient patients) and adults with learning disabilities (inpatients and outpatient patients), acute inpatients. Most patients have severe delayed aspiration or swallowing reflex. In cooperation with the patient's treatment, we adopted 3 ml, 5 ml and 7 ml of barium in different solubility, and recorded the lateral perspective at the rate of 30 frames/second. The density and viscosity of barium had a certain influence on the time of injection through the mouth and the opening of the upper esophageal sphincter [17].

Make full use of the modified swallowing research tool, this tool is used to quantify swallowing disorder. The clinician attempts to make a longitudinal observation of the degree of treatment by observing the flow of fluid barium at different concentrations. Therefore, we mainly observed the presence of throat penetration, trachea aspiration and pharyngeal residue. The residual amounts in the oral cavity, epiglottic cartilage and pyriform fossa were extracted, which were classified as lightweight, moderate and severe.



**Figure 2** This case demonstrates a simplified procedure for the extraction of a patient's piriform fossa.

The premise of our work is to judge a single frame from the original barium recognition, and gradually transfer to the determination of tissues and organs through the barium recognition. Why accurate tracking, only the complete track of barium agent, to facilitate the later determination of barium agent in each organ stay position, time and quantitative study. We used morphological methods to extract barium. Due to the quality difference of the original film, the expected errors of the barium trace area deviation in this experiment were within  $-25\% \sim 25\%$ .

### 3.1 Image Pre-processing

The process of image processing mainly applies mathematical morphology algorithm, which is based on set operation and provides a method for the development of nonlinear operators, including expansion and corrosion, two basic morphological operators [18]. Between morphological processing, we will first image pixel linear normalization, image as an array, the image was normalized maximum range for array  $(source(i, j))_{max}$  with minimum  $(source(i, j))_{min}$  short of each value from the array  $source(i, j)$  divided by the minimum value, multiplied by the normalized scheduled range  $beta - alpha$ , and then divided by the original scope of normalization, then add predefined normalized minimum, computation formula is as follows.  $alpha = 0, beta = 1$ .

$$dst(i, j) = \frac{[source(i, j) - source(i, j)_{min}] * (beta - alpha)}{(source(i, j))_{max} - (source(i, j))_{min}} + alpha \quad (1)$$

For the binary grayscale image, first set the threshold  $thresh = 30$ , and the determination of the threshold is determined by sampling and collecting the optimal threshold for the X-ray image. In order to obtain the best visual effect, take the maximum value  $maxval = 255$ . We assume that the element structure  $K$  is an image matrix more than the image thinks, and the moving process of the element structure is consistent with the moving process of the convolution kernel. Dilation utilizes the expanded “ $\oplus$ ” operation of the collection, and scans with the elementary structure  $K$  in the original binary image  $M$ . If the area where the elemental structure intersects is not empty, it is marked as a new collection. The effect of dilation is to fill in some holes in the pixel area, including the subtle noise points remaining after binarization. Corrosion uses the contraction “ $\ominus$ ” operator of the collection. If  $K$  is completely contained by  $A$ , it is marked as a new collection. Corrosion is the opposite of expansion and is used to eliminate small and meaningless noise points. Using the order of corrosion and expansion, we can agree to open and close operations.

Expansion formula:

$$M \oplus K = \{x, y | (K)_{xy} \cap M \neq \emptyset\} \quad (2)$$

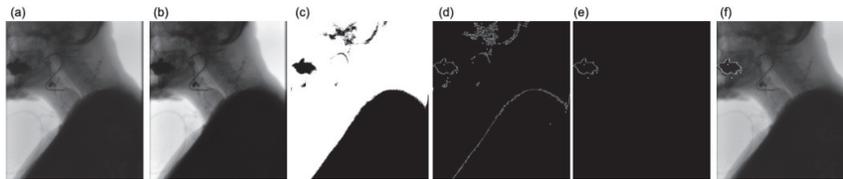
This formula represents expansion with structure  $K$ ,  $M(x, y)$  represents the pixel position.

Corrosion formula:

$$M \ominus K = \{x, y | (K)_{x,y} \subseteq M\} \quad (3)$$

This expression indicates that the structure  $K$  is used for etching, and  $M(x, y)$  indicates the pixel position.

We through the after image preprocessing, reduce some of the noise at the same time also strengthened the edge character, then we use the Canny algorithm for edge extraction, the basic principle of using the gaussian filter  $g_\sigma(m, n)$  first to smooth the image  $f(m, n)$ , and then through the calculation of sobel operator gradient value of  $\sqrt{g_x(m, n)^2 + g_y(m, n)^2}$ , reuse of threshold value will have a strong change point out [19]. We extracted the position of pharynx, and the process of image preprocessing took about 153 s. In this process, due to the influence of noise, different iterative processes were set in the process of processing materials. For example, CSN has 20 iterations in the expansion link and 17 iterations in the corrosion link. GJ sets 26 expansion iterations and 17 corrosion iterations. An example of the experimental results is shown below.



**Figure 3** Image pretreatment process (a) Original image; (b) Normalized image; (c) Binarization processing; (d) Canny edge detection; (e)–(f) Extract the pharyngeal area and mask the results.

### 3.2 Improved Inter-frame Difference

The basic idea of inter-frame difference method is to detect the moving region by the difference of brightness or color pixels and extract moving objects of the same position between different frames [20]. Based on two frames difference will be easy to appear the shadow and the cavity effect, namely overlapping shadows or moving objects, resulting in the shadow of the big or the result of the blank, we use three interframe difference method is used to eliminate certain influence, which USES the current frame  $I(t)$  minus a frame on the  $I(t - 1)$  and the next frame  $I(t + 1)$  images, the results to

the intersection, for the three frame difference of gray image. The operation process is as follows.

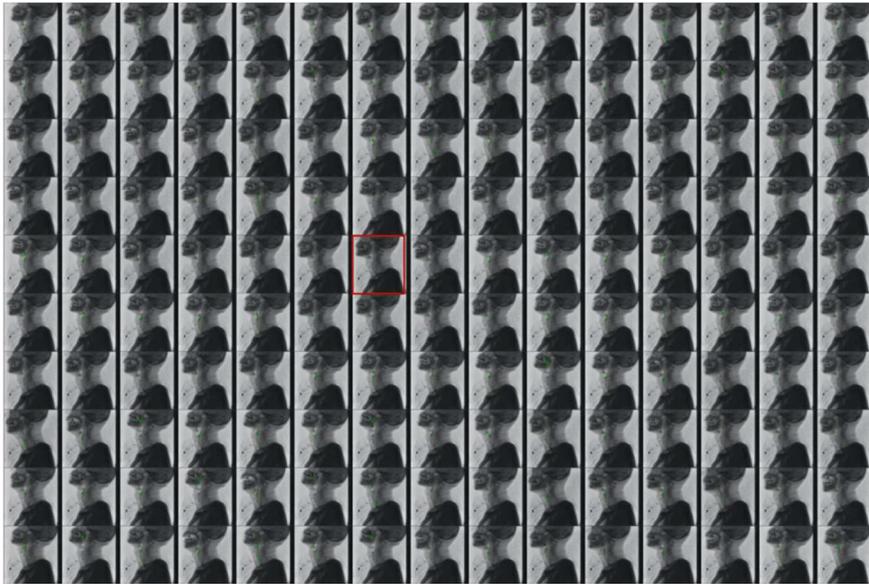
$$D_1 = |p_i(x, y) - p_{i-1}(x, y)| \tag{4}$$

$$D_2 = |p_{i+1}(x, y) - p_i(x, y)| \tag{5}$$

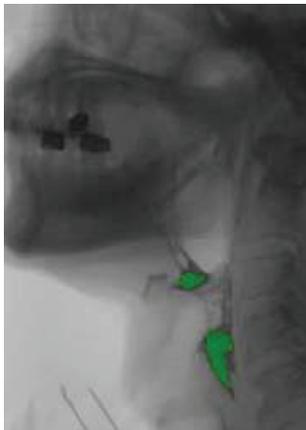
$$D = D_1 \cap D_2 \tag{6}$$

Among them,  $D_1$  represents the foreground image obtained by  $p_i(x, y)$  and  $p_{i-1}(x, y)$ ,  $D_2$  represents obtained by  $p_{i+1}(x, y)$  and  $p_i(x, y)$ ,  $D$  is the intersection of the two to get the final foreground image.

To frame, the use of three consecutive frames for morphological operation, and use the mask method makes the display on the current frame, and obtain the location of the mobile area namely barium [21]. We use the advantage of the three-frame difference algorithm in the extraction of fluid barium agent in the image and video. It can be found that the improved three-frame difference algorithm has a significant effect on the extraction of large amount of variation, and the average accuracy reaches 75% according to the correctly labeled image comparison data. The results are shown below.



**Figure 4** Based on improved three-frame difference experiment results.



**Figure 5** Enlarged display of experimental results.

### 3.3 The Background Subtraction Algorithm Gets the Moving Target

The background subtraction algorithm with global threshold can enhance the robustness. The goal of background extraction is to find the background value according to the VFSS image sequence. Its basic idea is not to obtain the static background, but to build a dynamically updated background frame. Therefore, the background frame is not fixed, but constantly updated with the number of frames [20–25]. Background deduction algorithm is usually used for the extraction of traffic conditions, relative to swallow angiography, traffic has the background (water waves, the wind blowing trees, etc.) the characteristics of relatively fixed, swallowing imaging video background is the body part of the tissues and organs, fluid extract as the only prospect, however, in the process of swallowing booster movement will drive the movement of the body, which greatly strengthened the background subtraction in addition to the application of the algorithm.

Therefore, based on the above problems and the adaptability of inter-frame difference to the dynamic environment, we used it to establish the background model, and then used the background subtraction algorithm to obtain the moving target, and then carried out morphological processing to obtain the motion trajectory of fluid barium agent.

We use the background subtraction method to use it in the VFSS video image, first randomly select a frame of image  $p_i(x, y)$ , and then update the

next frame, what needs to be noted is, this next frame is not necessarily the next frame corresponding to the image, and when the pixel value of the current scene value reaches the critical value, it is converted to background pixels, and it is the next frame obtained by sampling random probability  $\frac{1}{\varphi}$  update. Therefore, the update measurement strategy we adopted for the background subtraction algorithm is: use a conservative update strategy, do not use the foreground to fill the model, and at the same time count the pixels, update each pixel when the threshold is reached, and the update probability is  $\frac{1}{\varphi}$ . When the sample value  $(x, y)$  of the pixels in the updated image is not classified as the background, and it is classified as the foreground, the background is updated, and after the  $\Delta t$  time elapses, the probability that its sample value will remain is,

$$P(t, t + \Delta t) = \left( \frac{M - 1}{M} \right)^{(t+\Delta t)-t} \quad (7)$$

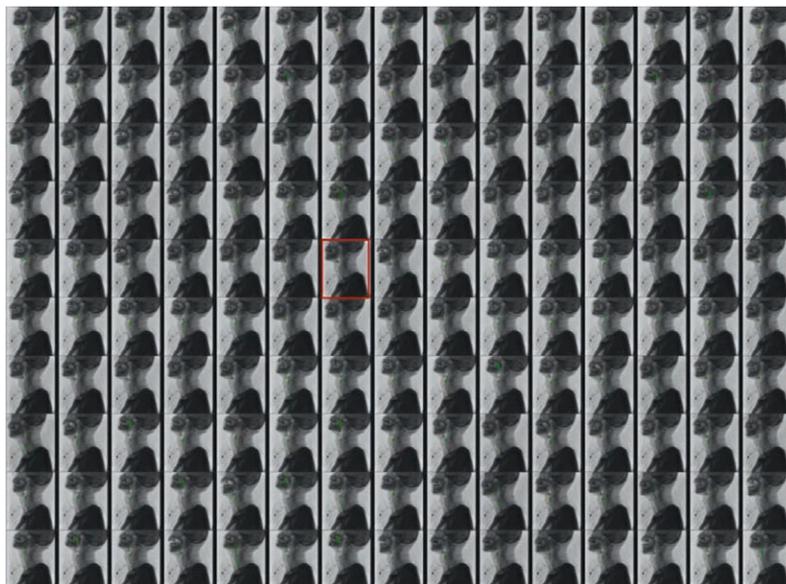
Can also be written as,

$$P(t, t + \Delta t) = e^{-\ln\left(\frac{M}{M-1}\right)\Delta t} \quad (8)$$

Where  $p(x, y)$  is the current pixel at the pixel point  $(x, y)$ ,  $M(x, y) = \{p_1(x, y), p_2(x, y), \dots, p_N(x, y)\}$  is the background sample set of pixel  $(x, y)$  (The sample set size is  $N$ );

However different lighting conditions will lead to VFSS video image research, therefore, we need from a frame to use gaussian background model of gray image algorithm to adjust balance in the appeal of models, we the pixels by weighted gaussian mixture distribution model, results show that fluid of VFSS barium extraction has a certain impact. The picture shows the result. The background subtraction algorithm greatly improved the effect of body part shaking, and the accuracy of barium extraction was up to 85% (relative to the expert mark), which was significantly better than the results obtained by the frame difference algorithm alone, but the overall calculation time was improved.

Fuzzy logic is a Boolean logic extension that deals with some real concepts. According to the conditions, we calculate the residual amount of the mouth, epiglottic cartilage and piriform fosse, and there will still be some noise influence. Therefore, when the pixel value of the system is not estimable, we default to 0 residual [26].



**Figure 6** Experimental results of barium extraction based on background subtraction.



**Figure 7** Enlarged display of experimental results.

### **3.4 Optical Flow Method**

The optical flow field is similar to the motion field. The optical flow of the background in the video sequence acquired by VFSS is unchanged, and the optical glide of the moving target, namely the barium agent, changes,

thereby forming the optical flow field. Therefore, for the complex noise interference and illumination changes in VFSS, we can extract the barium agent trajectory according to the change of optical flow. In the process of watershed segmentation, we use the Lucas-Kanada algorithm to obtain a matrix of coefficients, considering the change in the flow rate of barium under a given number of frames [27], that is, to solve the optimization problem,

$$\begin{aligned} & \min \frac{1}{2} \|f_x u + f_x v + f_t\|^2 \\ & s.t. u_x^2 + v_x^2 = 0, u_y^2 + v_y^2 = 0 \end{aligned} \quad (9)$$

We differentiate the results to calculate the optical flow iteratively. The estimation of local optical flow is based on velocity smoothing and brightness balance. We need to divide the VFSS image into small areas to solve the problem of noise error propagation. The results are shown in the following figure.

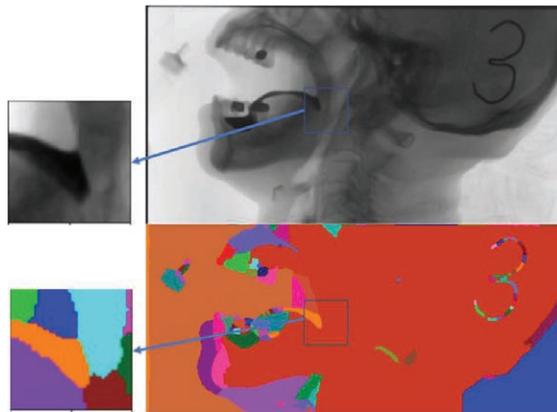


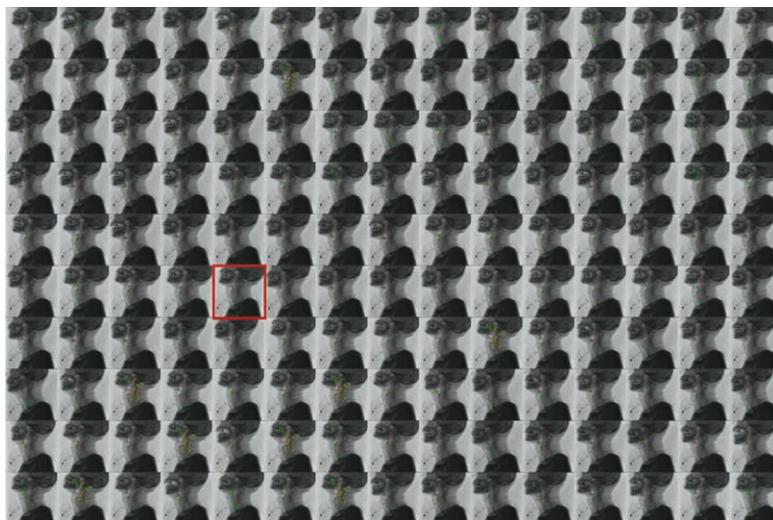
Figure 8 Optical flow method experiment results.

### 3.5 Confirm ROI Based on X-ray Spine Position

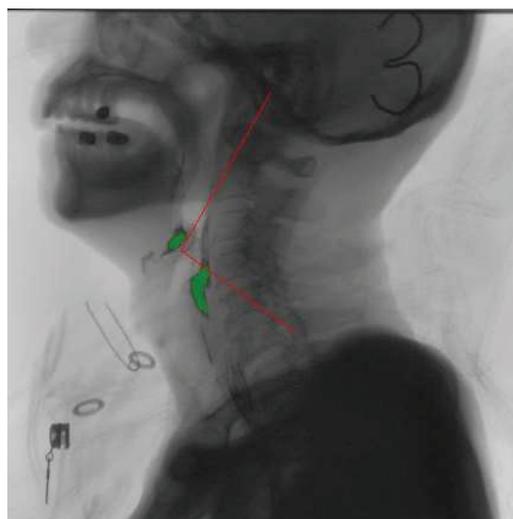
Recognizing the direction of an X-ray photograph is a challenging subject. Normally, the swallowing posture and position of the patient vary, and the position of the body's oral cavity and the length of the pipe will also vary at different ages, as shown in the following figure. As a result, the acquired video frame sequence is largely related to the radiologist's shooting situation.

Through observation, we can find that the position of the cervical spine has a relatively high degree of recognition. We can calibrate the position

of the cervical spine to track the movement trajectory of the fluid barium. As a model verification and correction link, it will help us to improve the recognition accuracy. Experimental results As shown in the figure below.



**Figure 9** Taking the cervical spine as the coordinate system.



**Figure 10** Taking the cervical spine as the coordinate system (enlarged display).

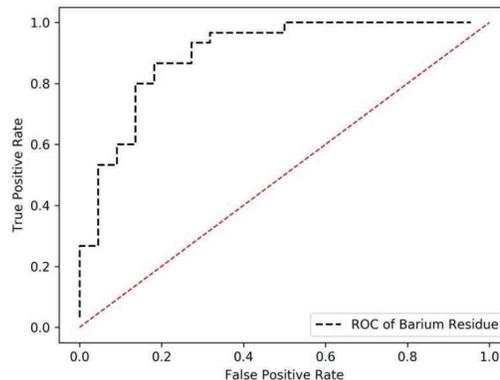
## 4 Experiments and Results

### 4.1 Evaluation Metrics

In order to evaluate the performance of the moving target extraction method during swallowing, we classified the lateral image data set in VFSS into two categories: normal swallowing and dysphagia. Due to the limited number of dicom files provided by hospitals for VFSS of patients and the uneven quality of data sets, in order to make good use of these data sets, when training the network, we randomly divided the data into a training set (80%) and a test set (20%). In the initial stage, we will manually go out the long interference frames in the data set, for example, the process of clinician assisting the patient to inject barium, and the abnormal swallowing of the patient leads to the long-term retention of barium.

During the training process, we load 30 frames (1 s) into the evaluated network each time, and then advance 5 frames one by one, repeat this process, and guide until all frames are evaluated. The training uses a binary cross loss function. The literature [28] gives us a good example, let us determine the possibility of the realization of this training method. We observed the barium residues in 24 examples. When it exceeded a certain threshold, it was determined that it was difficult to swallow, otherwise it was normal to swallow.

We simplified the relevant training process, created a binary classifier to assess whether the patient's VFSS swallowing was normal, and used the test data to draw a ROC (Receiver Operating Characteristic) curve. Note that in order to facilitate the calculation, we extracted three total amounts of epiglottis (oral, pharyngeal, and esophageal), and plotted the curve as shown in the figure below.



**Figure 11** Schematic diagram of ROC curve.

We can observe that the ROC curve is relatively stable, and there is basically not much overfitting, but for the determination of difficulty and normal swallowing, as “There are a hundred Hamlets in a hundred personal cores.” said Yes, even for the same patient, different clinicians will have some errors in the process of determining difficulty in swallowing. Therefore, the performance in terms of time and space detection will be more complicated than the target detection in the image, such as whether it is a cat or a dog. In order to judge in this regard, we will also strengthen and improve in the later period.

## 4.2 Swallowing Imaging Application System

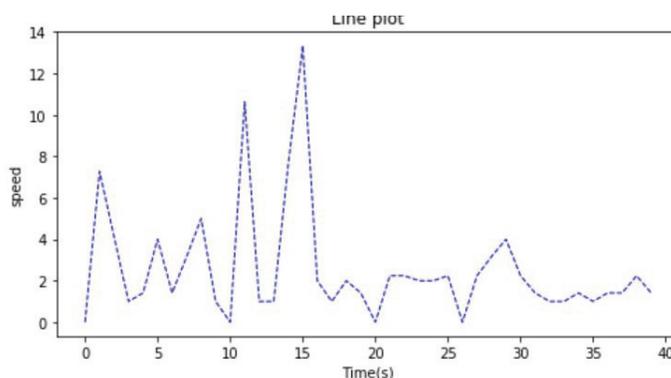
During the experiment, we extracted a separate sample for the experiment. The whole is divided into four stages:

### 4.2.1 Digitized recording of each frame in the video

Dicom (Digital Imaging and Communications in Medicine) has become one of the medical standards. The digitization of Dicom files by X-ray is different from that by tomography, 3D images can be obtained by stacking, so we can only obtain the pixel position brightness from the grayscale image, as well as the change information between each frame over time [29].

### 4.2.2 Determine the reference position

A given image size  $480 * 480$ , according to the pharyngeal variation line (see Figure 7), calibration pharyngeal overall position. As shown in Figure 8, the blue box represents the entire pharynx position.



**Figure 12** Pharyngeal variation line.

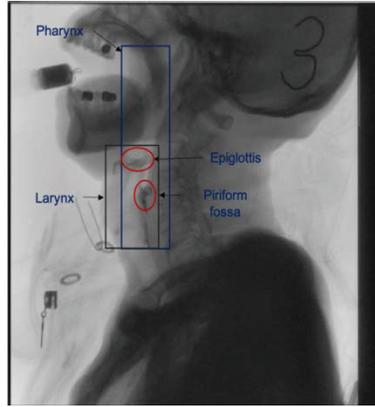


Figure 13 The blue box represents the entire pharynx position.

#### 4.2.3 Calculate the flow velocity

After the extraction of barium, we carried out the next quantitative analysis. The quantitative analysis of fluid is difficult, but the conversion to pixel tracking will simplify the difficulty of the problem. Since the swallowing time is too short (1~2 s), the phased velocity analysis is adopted. The calculation method of speed is as follows:

$$v = \frac{\Delta s}{(t_1 - t_0)} \quad (10)$$

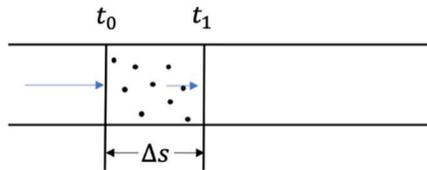


Figure 14

#### 4.2.4 Draw the motion of the target point

By using an improved background subtraction algorithm, we extracted fluid barium targets as much as possible and masked them into the image for clinicians to observe.

Data recording. For different patients, the keyframe method is used to extract the keyframe, and the image at the moment of swallowing is usually extracted. According to the duration of swallowing, the keyframe is extracted

for 1–2 frames after which the residual data of Oral cavity, Epiglottic cartilage and Piriform fossa are analyzed.

**Table 1** Results of different samples

| Sample Id  | Key Frame Number | Oral Residue | Epiglottic Cartilage Residue | Pyriform Fossa Residue |
|------------|------------------|--------------|------------------------------|------------------------|
| KLH-01-414 | 414              | 103          | 90                           | 92                     |
| GJ-02-281  | 281              | 0            | 23                           | 126                    |
| GJ-01-449  | 449              | 0            | 114                          | 91                     |
| MJS-02-303 | 303              | 205          | 1                            | 0                      |
| MJS-03-231 | 231              | 0            | 2                            | 203                    |
| GJ-01-261  | 261              | 59           | 142                          | 224                    |
| CSN-03-56  | 56               | 139          | 289                          | 3                      |
| SH-01-280  | 280              | 1009         | 108                          | 0                      |
| SH-01-186  | 186              | 0            | 78                           | 0                      |
| LAM-01-185 | 185              | 188          | 29                           | 23                     |
| BZY-01-138 | 138              | 29           | 58                           | 46                     |
| YSL-01-128 | 128              | 169          | 38                           | 99                     |
| ZCX-01-233 | 233              | 194          | 36                           | 0                      |
| ZHS-03-197 | 197              | 529          | 33                           | 0                      |
| LML-01-69  | 69               | 851          | 117                          | 179                    |
| GJ-02-361  | 361              | 0            | 78                           | 294                    |
| GJ-03-216  | 216              | 0            | 91                           | 175                    |
| GJ-01-254  | 254              | 0            | 197                          | 400                    |
| YWX-03-146 | 146              | 585          | 12                           | 33                     |
| ...        | ...              | ...          | ...                          | ...                    |

We are developed based on computer-assisted diagnosis and treatment. The visual calibration process is helpful to compare the changes of patients before and after treatment, and assist clinicians in diagnosis and treatment. At the same time, it is beneficial to understand professional medical X-ray images with patients and build a bridge of trust between patients and physicians. Therefore, developing a complete set of system software for practical applications is also the goal of this article.

A complete set of system software includes five basic modules. 1. Video image sequence preview module based on X film. Realize frame-by-frame playback and playback, and obtain basic information of patients and images. 2. Based on X-film video frame preprocessing module. Analyze the preprocessing process of initialization process, threshold adjustment and spine adaptive rotation. 3. Quantitative analysis module based on X-ray image sequence. Analyzed the software calibration, barium meal morphology analysis and extraction during swallowing, so as to obtain the residual amount of barium meal, movement speed and other data. 4. Information collection system. Based on the diagnosis parameters of swallow angiography determined by clinicians, including time and kinematic parameters. 5. Check the report printing module. The purpose is to achieve a horizontal comparison between different patients and a longitudinal comparison of the same patient in different periods.

## **5 Conclusion**

We used semi-quantitative analysis to obtain residual volume results combined with the clinician's diagnosis. On the one hand, the quantitative data was recognized by the clinician to enhance the reliability; on the other hand, the data results verified the clinician's diagnosis, so as to improve the professionalism. The background subtraction algorithm combined with gauss operation has the advantages of fast speed and high accuracy compared with the inter-frame difference method, which balances the influence of patients' movement during the shooting process. However, there is still a low sensitivity to the low peak stage, that is, the place where the fluid change in the swallowing process is low, which affects the accuracy of the data.

There are some defects in quantitative analysis of swallowing contrast. First, using X-ray observations, there is a certain amount of radiation to both the patient and the clinician. Secondly, the data obtained through quantitative examination have certain errors, which still need further observation and detection by clinicians. Therefore, the purpose of the study is to focus on adjuvant therapy. Third, the research lacks a large number of standardized data for analysis, and the evaluation standard can only be used as a reference. Such, however, difficulty swallowing as the symptoms of the elderly today increasingly obvious, blindly by clinicians subjective and professional judgment of the reliability is not high, therefore, quantitative analysis using the computer aided design VFSS image will become the future

development trend. Hope that through this study, for the treatment of patients with dysphagia is of certain clinical significance.

## Acknowledgement

The project is supported by the three-year action plan of Shanghai to further accelerate the development of traditional Chinese Medicine. (Project No. ZY (2018-2020)-FWTX-8015).

## References

- [1] Martin-Harris B, Jones B. The Videofluorographic Swallowing Study [J]. *Physical Medicine and Rehabilitation Clinics of North America*, 2008, 19(4):769–785.
- [2] Kim SM, McCulloch TM, Rim K. Pharyngeal pressure analysis by the finite element method during liquid bolus swallow [J]. *Ann Otol Rhinol Laryngol*, 2000, 109:585–589.
- [3] McConnel FM. Analysis of pressure generation and bolus transit during pharyngeal swallowing [J]. *Laryngoscope*, 1988, 98:71–78.
- [4] Cook IJ. Normal and disordered swallowing: new insights [J]. *Baillieres Clin Gastroenterol*, 1991, 5:245–267.
- [5] Shengli Li. Rehabilitative evaluation and treatments for patient with multiple sclerosis [J]. *Chinese Journal of rehabilitation Technology and Practice*, 1998(4):178–181.
- [6] Palmer, J.B., Kuhlemeier, K.V., Tippett, D.C., Lynch, C. A protocol for the videofluorographic swallowing study. *Dysphagia* 1993, 8, 209–214.
- [7] Ertekin, C., Aydogdu, I. Neurophysiology of swallowing. *Clin. Neurophysiol.* 2003, 114, 2226–2244.
- [8] Yangwei Chen, Luodan Xu. Effect of early nasal feeding intervention on dysphagia complicated with pulmonary infection in stroke patients [J]. *Jilin medicine*, 2014, 35(15):3357.
- [9] Tjaden K. Speech and swallowing in Parkinson's disease [J]. *Topics in geriatric rehabilitation*, 2008, 24(2):115.
- [10] Mosselman MJ, Kruitwagen CL, Schuurmans MJ, et al. Malnutrition and risk of malnutrition in patients with stroke: prevalence during hospital stay [J]. *J Neurosci Nurs*, 2013, 45(4):194–204.
- [11] Ekberg O, Feinberg MJ. Altered swallowing function in elderly patients without dysphagia: radiologic findings in 56 cases [J]. *AJR. American journal of roentgenology*, 1991, 156(6):1181–1184.

- [12] Dodds WJ, Man KM, Cook IJ, et al. Influence of bolus volume on swallow-induced hyoid movement in normal subjects [J]. *American Journal of Roentgenology*, 1988, 150(6):1307–1309.
- [13] Cook IJ, Dodds WJ, Dantas RO, et al. Opening mechanisms of the human upper esophageal sphincter [J]. *American Journal of Physiology-Gastrointestinal and Liver Physiology*, 1989, 257(5): G748–G759.
- [14] Dengel G, Robbins JA, Rosenbek JC. Image processing in swallowing and speech research [J]. *Dysphagia*, 1991, 6(1):30–39.
- [15] Lee JT, Park E, Jung TD. Automatic detection of the pharyngeal phase in raw videos for the videofluoroscopic swallowing study using efficient data collection and 3d convolutional networks [J]. *Sensors*, 2019, 19(18):3873.
- [16] Zulin Dou, Yue Lan, Fang Yu, et al. Application of videofluoroscopy digital analysis in swallowing function assessment for brainstem stroke patients with dysphagia [J]. *Chinese Journal of Rehabilitation Medicine*, 2013, 28(9):799–805.
- [17] Dantas RO, Dodds WJ, Massey BT, et al. The effect of high- vs low-density barium preparations on the quantitative features of swallowing [J]. *American Journal of Roentgenology*, 1989, 153(6):1191–1195.
- [18] Hsiao YT, Chuang CL, Jiang JA, et al. A contour based image segmentation algorithm using morphological edge detection [C]//2005 IEEE International Conference on systems, man and cybernetics. IEEE, 2005, 3:2962–2967.
- [19] Canny J. A computational approach to edge detection [J]. *IEEE Transactions on pattern analysis and machine intelligence*, 1986 (6):679–698.
- [20] Weng M, Huang G, Da X. A new interframe difference algorithm for moving target detection [C]//2010 3rd international congress on image and signal processing. IEEE, 2010, 1:285–289.
- [21] He Zhang. Research on moving object detection algorithm [D]. Wuhan: Wuhan University of Science and Technology, 2011.
- [22] Van Droogenbroeck M, Barnich O. ViBe: A disruptive method for background subtraction [J]. *Background modeling and foreground detection for video surveillance*, 2014:7.1–7.23.
- [23] Barnich O, Van Droogenbroeck M. ViBe: A universal background subtraction algorithm for video sequences [J]. *IEEE Transactions on Image processing*, 2010, 20(6):1709–1724.
- [24] Van Droogenbroeck M, Paquot O. Background subtraction: Experiments and improvements for ViBe [C]//2012 IEEE computer society

- conference on computer vision and pattern recognition workshops. IEEE, 2012:32–37.
- [25] Barnich O, Van Droogenbroeck M. ViBe: a powerful random technique to estimate the background in video sequences [C]//2009 IEEE international conference on acoustics, speech and signal processing. IEEE, 2009:945–948.
- [26] López-Rubio FJ, López-Rubio E. Local color transformation analysis for sudden illumination change detection [J]. *Image and Vision Computing*, 2015, 37:31–47.
- [27] Farnebäck G. Two-frame motion estimation based on polynomial expansion [C]//Scandinavian conference on Image analysis. Springer, Berlin, Heidelberg, 2003:363–370.
- [28] Wilhelm P, Reinhardt JM, Van Daele D. A Deep Learning Approach to Video Fluoroscopic Swallowing Exam Classification [C]//2020 IEEE 17th International Symposium on Biomedical Imaging (ISBI). IEEE, 2020:1647–1650.
- [29] Mildenerger P, Eichelberg M, Martin E. Introduction to the DICOM standard [J]. *European radiology*, 2002, 12(4):920–927.

## Biographies



**Guofeng Qin** is with The Department of Computer Science and Technology as associate Prof, Tongji University, Shanghai, 201804, China. Received BA form Hunan University in 1995, MA form Science and Technology of Huazhong University in 2001, and PHD from Tongji University in 2004. e-mail: [gfqing@tongji.edu.cn](mailto:gfqing@tongji.edu.cn).



**Jianhuang Zou** is a graduate student in School of Electronics and Information Engineering at Tongji University, China. Received BA Fuzhou University e-mail: [jianhuang\\_zou@tongji.edu.cn](mailto:jianhuang_zou@tongji.edu.cn). Her research focuses on image recognition.



**Qiufang Xia** is with the director and deputy chief physician of the Department of Chinese Medicine Rehabilitation at Shanghai First Rehabilitation Hospital, China Acupuncture and Moxibustion, Shanghai, 200092, China. Received BA from Anhui University of Traditional Chinese Medicine in 2001, MA from Shanghai University of Traditional Chinese Medicine in 2011. e-mail: [13651899419@163.com](mailto:13651899419@163.com).



**Jiahao Qin** is a graduate student in School of Electronics and Information Engineering at Tongji University, China. e-mail: [jiahao\\_qin@tongji.edu.cn](mailto:jiahao_qin@tongji.edu.cn). His research focuses on image recognition.

