Mouth Shape Detection Based on Template Matching and Optical Flow for Machine Lip Reading

Tsuyoshi Miyazaki, Department of Information and Computer Science, Kanagawa Institute of Technology, Atsugi, Kanagawa, Japan

Toyoshiro Nakashima, Department Culture-Information Studies, Sugiyama Jogakuen University, Chikusa, Nagoya, Aichi, Japan

Naohiro Ishii, Department of Information Science, Aichi Institute of Technology, Yakusa, Toyota, Aichi, Japan

ABSTRACT

The authors describe an improved method for detecting distinctive mouth shapes in Japanese utterance image sequences. Their previous method uses template matching. Two types of mouth shapes are formed when a Japanese phone is pronounced: one at the beginning of the utterance (the beginning mouth shape, BeMS) and the other at the end (the ending mouth shape, EMS). The authors' previous method could detect mouth shapes, but it misdetects some shapes because the time period in which the BeMS was formed was short. Therefore, they predicted that a high-speed camera would be able to capture the BeMS with higher accuracy. Experiments showed that the BeMS could be captured; however, the authors faced another problem. Deformed mouth shapes that appeared in the transition from one shape to another were detected as the BeMS. This study describes the use of optical flow to prevent the detection of such mouth shapes. The time period in which the mouth shape is deformed is detected using optical flow, and the mouth shape during this time is ignored. The authors propose an improved method of detecting the BeMS and EMS in Japanese utterance image sequences by using template matching and optical flow.

Keywords: Japanese Vowel, Machine Lip Reading, Mouth Shape Detection, Optical Flow, Template Matching

INTRODUCTION

We have studied Japanese machine lip reading by modeling skilled lip readers, who pay attention to the sequence of mouth shapes when they read lips. Therefore, we studied how to detect mouth shapes in images of Japanese speakers. We considered that words and phrases can be recognized from the detected mouth shape sequences and proposed a method of detecting the basic mouth shape (BaMS) by using template matching (Miyazaki, Nakashima, & Ishii, 2011). The BaMS is the set of mouth shapes associated with Japanese vowels and the closed mouth.
Japanese language has only five vowels (/a/, /i/, /u/, /e/, and /o/). The BaMS is defined as:

\[ \text{BaMS} = \{A, I, U, E, O, X\} \]  

(1)

where the first five symbols represent the vowels, and \( X \) represents the closed mouth. Conventional studies of machine lip reading have adopted a method based on words (Kiyota & Uchimura, 1993; Nakata & Ando, 2002; Okumura, Hamaguchi, Okano, & Miyazaki, 1998; Saitoh & Konishi, 2007; Uda, Tagawa, Minagawa, & Moriya, 2001). This method requires real utterance images because the features of each word or phrase are calculated from the images. If the number of words and phrases to be recognized increases, this method is cumbersome. In contrast, our proposed method uses the mouth shape. It is easy to identify a sequence of mouth shapes associated with a word or phrase because the features of each word or phrase are calculated from the images. However, in some cases the method did not detect the beginning mouth shape (BeMS), as defined in (2):

\[ \text{BeMS} = \{I, U, X\} \]  

(2)

We believe that the BeMS frames may be dropped because the mouth shape is formed for a very short time period. The digital video camera that we used was intended for home use, and the frame rate was only 30 fps. Therefore, we used a video camera with a higher frame rate (60 fps) to capture the BeMS frames. The new camera could capture the necessary frames; however, we faced another problem. When one mouth shape changes to another, the deformed mouth shape at that transitional moment was misdetected as the BeMS. To correct this, the motion of the lips is measured to prevent the detection of these deformed mouth shapes. This study describes the previous method and how optical flow (Farnebäck, 2003) is adopted to measure the distance that the lips move.

**PREVIOUS BAMS DETECTION METHOD**

We have proposed a method of detecting the BaMS using template matching (Miyazaki, Nakashima, & Ishii, 2011). Six mouth shape images were used as template images, and the similarity to each BaMS was calculated for a sequence of images of Japanese utterances. In addition to the BeMS, we used the ending mouth shape (EMS), as defined by (3).

\[ \text{EMS} = \{A, I, U, E, O, X\} \]  

(3)

When the EMS is formed, the similarity waveform of the corresponding mouth shape becomes flat. In contrast, when the BeMS is formed, the corresponding waveform becomes convex. Using these characteristics, our previous method first detected the BeMS and EMS periods and then detected the BaMS for each period.

However, as mentioned above, the correct BeMS was not always detected because a digital video camera for home use was employed. The similarity waveforms in Figure 1 depict the case of a closed mouth changing to the mouth shape /a/. The x axis shows the frame number, and the y axis shows the similarity to the BaMS. Each waveform describes the similarity for one of the six BaMSs. In frames 0-19, a closed mouth shape is formed, and the mouth shape /a/ is formed in frames 30-40. Therefore, deformation appears in frames 20-29, and a deformed mouth shape similar to /i/ is formed during this time period (indicated by an arrow in Figure 1). The convex waveform here caused the misdetection.

To detect the BeMS correctly, it is necessary to ignore such deformed mouth shapes.
Design Churn as Predictor of Vulnerabilities?
*International Journal of Secure Software Engineering* (pp. 16-31).
[www.igi-global.com/article/design-churn-as-predictor-of-vulnerabilities/118146?camid=4v1a](www.igi-global.com/article/design-churn-as-predictor-of-vulnerabilities/118146?camid=4v1a)

Identifying the Opinion Orientation of Online Product Reviews at Feature Level: A Pruning Approach
[www.igi-global.com/article/identifying-the-opinion-orientation-of-online-product-reviews-at-feature-level/199005?camid=4v1a](www.igi-global.com/article/identifying-the-opinion-orientation-of-online-product-reviews-at-feature-level/199005?camid=4v1a)