The Use of Tongue Protrusion Gestures for Video-based Communication

Luis Ricardo Sapaico   Hamid Laga   Masayuki Nakajima
Tokyo Institute of Technology
Ookayama 2-12-1, Meguro, Tokyo 152-8552, Japan
{ruisu, hamid, nakajima}@img.cs.titech.ac.jp

Abstract

We propose a system that tracks the mouth region in a video sequence and detects the occurrence of a tongue protrusion event. Assuming that the natural location of the tongue is inside the mouth, the tongue protrusion gesture is interpreted as an intentional communication sign that the user wishes to perform a task. The system operates in three steps: (1) mouth template segmentation, in which we initialize one template for the entire mouth, and one template for each of the left and right halves of the mouth; (2) mouth region tracking using the Normalized Correlation Coefficient (NCC), and (3) tongue protrusion event detection and interpretation. We regard the tongue protrusion transition as the event that begins when a minimum part of the tongue starts protruding from the mouth, and which ends when the protrusion is clearly visible. The left and right templates are compared to their corresponding halves for each new mouth image that has been tracked, and a left-NCC and a right-NCC are obtained for each part. By analyzing the NCCs during the tongue protrusion transition time, the left or right position of the protrusion, relative to the center of the mouth, is determined. We analyze our proposed communication method and demonstrate that it adapts easily to different users. The detection of this gesture can be used for instance as a dual-switch hand-free human-computer interface for granting control of a computer.

1. Introduction

Perceptual user interfaces, as indicated by Turk and Robertson [16], aim at establishing a natural communication between humans and machines, such as it resembles the way we interact with other people in our everyday lives. Face features have been used previously for human-computer communication. Examples include using eye blinks or eyebrow raises as proposed by Grauman *et al*. [5]; or using head-nodding and shaking as proposed by Davis and Vaks [3]. The difficulty with such interfaces is the discomfort or weariness they may create when generating events that should be understood by the computer. On the other hand, arguably, one of the most natural and primitive means of facial communication is tongue protrusion, a gesture so simple that even neonates are able to reproduce just by observation, as Anisfeld substantiated in [1].

Figure 1. Tongue Protrusion events to detect: Right and Left.

In this paper we propose a Perceptual User Interface based on tongue protrusion. The communication is based on two tongue events: (1) tongue protrusion to the left side, and (2) to the right side of the mouth. In both cases, the mouth should remain closed, as illustrated in Figure 1. We use an off-the-shelf webcamera to detect these events so they can be translated into computer commands. The system operates in three stages; (1) initial mouth region detection where the closed mouth region template is automatically obtained from the first frame, (2) mouth region tracking in subsequent frames, and (3) the tongue event protrusion detection and interpretation.

For obtaining localized information regarding the changes on the mouth region, the mouth template is divided into its Left and Right halves, thus we work using three templates. Each sub-template is compared with its corresponding half in the newly tracked mouth template by means of Normalized Correlation. The coefficients of the three correlations (mouth, left and right templates) are used for estimating the protrusion position during the tongue protrusion transition time. We regard the tongue protrusion transition event as the interval of time such as although the mouth region is changing in appearance due to the tongue protrusion, this region is still tracked accurately at each incoming frame. It begins when the tongue starts protruding from the mouth, and finishes when the protrusion is clearly visible

2009 IEEE 12th International Conference on Computer Vision Workshops, ICCV Workshops
978-1-4244-4441-0/09/$25.00 ©2009 IEEE

and the tracked mouth is not accurate anymore. The only requirement for the transition analysis to hold is that the frames are captured in real-time, in order not to lose transition information, necessary for obtaining the correlation coefficients. This allows the tracking procedure to be robust to partial occlusions caused by the appearance of the tongue.

Our ultimate objective is to create a system capable of voluntarily generating customized tasks in a computer by means of the tongue protrusion. Examples of applications range from simple ones such as emulation of YES/NO acknowledgements, or utilization of the tongue for documents and photo albums browsing; to complex ones where for example a physically disabled person would be granted the control of the computer solely with the tongue movement.

The contributions of this paper are:

- We propose a system that tracks the mouth in real-time, and detects whether the tongue has been protruded or not by using mouth templates.
- We define the concept of tongue protrusion transition, which is used for the detection of the protrusion position.
- We perform a thorough analysis of the parameters required for setting-up the system and we prove its adaptability to different users.

The remaining parts of the paper are organized as follows. A short survey on tongue-based human-computer interfaces is presented in Section 2. In Section 3 we describe the algorithm developed for the initial mouth template segmentation. In Section 4 we detail the mouth tracking procedure and the framework used for the detection of the tongue protrusion position. Results and a discussion on the system configuration are presented in Section 5. We conclude and give some ideas for future work in Section 6.

2. Previous Work on Tongue-based Interfaces

In recent years, research on tongue-based interfaces has started to receive significant attention due to the capabilities of the tongue muscle, and its high degree of control. Related work can be classified into two main categories depending whether they are intrusive or non-intrusive.

2.1. Intrusive Interfaces

This type of interfaces requires the insertion of a hardware component inside the oral cavity. Struijk [14] proposed a system that changes the inductance of an air-cored induction coil by moving a ferro-magnetic material attached to the tongue into the core of the coils. Kencana and Heng [7] proposed a passive, haptic interface for the tongue based on induction as well. In this system the user needs to touch a panel in order to generate a linear current that is mapped for controlling the computer. The Tongue-TouchKeypad (TTK) [9] is a commercial product based on a wireless pressure-sensitive keypad that is inserted in the mouth, which the user controls by touching it with the tip of the tongue. Huo *et al*. [6] developed a magnetic tongue-computer interface which requires a small permanent magnet to be implanted on the tongue of the user, and a headset for measuring the changes in magnetic field due to tongue movement.

The main downside of this type of interfaces is the necessity of specialized hardware that is frequently invasive or obtrusive, which not only increases the cost, but also the discomfort of its user.

2.2. Non-intrusive Interfaces

Computer Vision technology for HCI is very promising since camera-based interfaces are widely accepted by users and often considered as non-intrusive. To the best of our knowledge, the only vision-based work has been proposed by Sapaico and Nakajima [12]. In this work, the tongue protrusion is detected by training a three-layer Support Vector Machines classifier for detecting the middle, left and right tongue protrusions. We consider that the inclusion of the middle tongue protrusion classifier hindered the effectiveness of the system. Moreover, the classifier works poorly for cases that were not included in the database, and are distant in terms of color or brightness.

Considering the issues discussed above, we have decided to work on a non-intrusive interface that uses a single end-user web camera; which not only is more affordable to install and maintain, but it is also convenient to utilize.

3. Mouth Template Segmentation

The first step of our system requires to segment accurately a mouth region for it to be tracked afterwards. There has been extensive research coping with mouth segmentation from a face image. Given that our objective is to propose an communicative system adaptable to different users and situations, some considerations apply so as to choose the method that suits our purpose the best:

1. There is only one user to access the system at any given time.
2. The user would be located in front of the computer, in an upright position.
3. For helping to the initial segmentation process, the user is asked to mantain the mouth closed.

Therefore, we choose to work with the method proposed by Sapaico *et al*. [11]. First, Viola and Jones [17] face detector is used. Then, they use a set of Gabor filters oriented

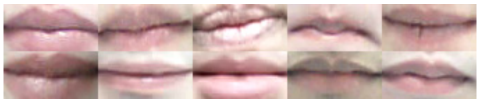

Figure 2. Some of the mouth regions segmented using the method described in Section 3.

horizontally such as they enhance the shadow line between the upper and lower lip, typical of a closed mouth. Additionally, the filters are scaled so as they deal with low-resolution images captured using a webcamera. A faster variant of the Hough Transform proposed by Matas [8], called the Progressive Probabilistic Hough Transform, is applied to the fused version of the filtered images. Hence, the line that connects both mouth corners is expected to be detected. Finally, these corners are taken as initialization points for a 15-point Active Appearance Model fitting process, that was previously trained with more than 350 manually-marked mouth images using the code provided by Stegmann *et al*. [13] in conjunction with the OpenCV Library [2]. By using this technique, we were going to segment mouth regions, some of which are illustrated in Figure 2.

4. Tracking and Protrusion Position Detection

We have segmented a mouth region given an upright frontal face. For speeding up our following procedure, a grayscale version of the mouth is used for the tracking. In order to remove noise while preserving important edges, the image is further processed using the Bilateral filter proposed by Tomasi and Manduchi [15], followed by a unsharpening mask. Next, the mouth template (**MT**) is divide into its left half template (**LHT**) and its right half template (**RHT**), as shown in Figure 3. The three templates will be utilized afterwards for the detection of the tongue protrusion position.

4.1. Mouth Tracking

The next step is to find the mouth at each incoming frame. To achieve this we use the Normalized Correlation

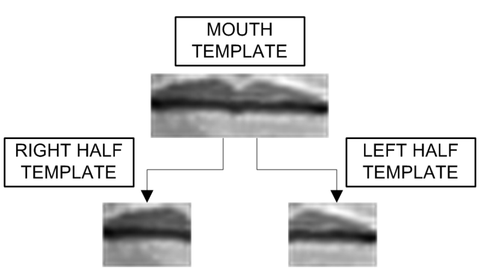

Figure 3. A Mouth Template and the templates corresponding to its Left and Right halves.

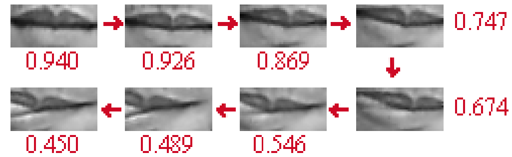

Figure 4. NCC values for a sequence of mouth tracking when protrusion to the right occurs.

Coefficient (NCC) similarity measure to find the area that best matches our mouth template. The use of the NCC for tracking purposes in the Human-Computer Interaction context was analyzed by Fagiani in [4], finding its suitability for this type of task. Additionally, the NCC is robust against illumination changes, and given that the size of our templates is small, it finds the best match at real-time speed. The search area for the new mouth region is defined using the current mouth position, extending the search over an area 50% larger at each side. Hence, the tracking allows for smooth rotation movements of the face.

Although the tracking procedure may seem trivial, in fact, the appearance of the tongue in the mouth image makes the tracking challenging. While the tongue protrudes from the mouth, the mouth region starts to change significantly. Therefore, there is one point at which the NCC tracking is not able to find the correct position where the mouth is located, because the amount of tongue that has protruded is large, producing a drift in the tracked template. This is exemplified in Figure 4, where we can observe in the first frame that the NCC is high, but as the tongue starts protruding, the NCC decreases and by the sixth frame of the sequence ($NCC = 0.546$), the mouth region is not tracked accurately, drifting towards the right side due to the tongue protrusion.

4.2. Detection of Tongue Protrusion Position

We have used the NCC as a similarity measure between the MT and the new mouth region. We will utilize the same measure for comparing the LHT and the RHT with their corresponding new tracked images. However, as we can notice in Figure 4, for low correlation values we cannot assure that an accurate mouth region has been segmented. This implies that we cannot assume we will permanently obtain good new left and right images.

From our observations, the issue shown in Figure 4 is common to all the protrusion events in our video database, i.e., there is a NCC value such as if we fall behind it, the mouth tracking will drift towards one of the sides. We call this value the Minimum Threshold (θ_{MIN}). On the other hand, there exists one threshold value such as it tells us when the mouth is starting to change, possibly due to a tongue protrusion, i.e., there is a NCC value such as if we

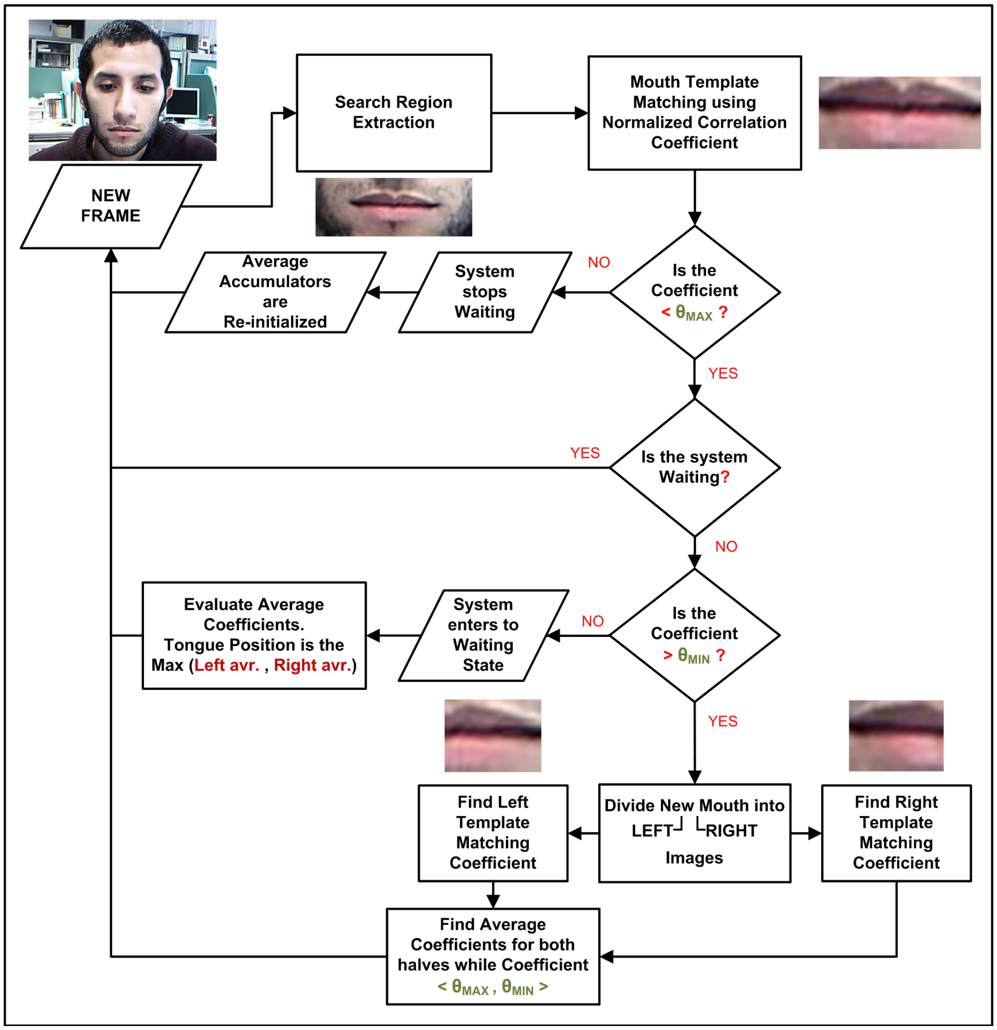

Figure 5. Flowchart of the mouth tracking method that allows the tongue protrusion detection.

fall behind it, the system enters into a waiting mode that will acknowledge if a protrusion event occurs. We call this value the Maximum Threshold (θ_{MAX}). For instance, referring exclusively to the data shown in Figure 4, we would choose $\theta_{MIN} = 0.65$ and $\theta_{MAX} = 0.80$.

The tracking procedure shown in Figure 5 makes use of both thresholds. They cooperate as follows:

1. While the Mouth NCC remains higher than θ_{MAX}, we do nothing, as the changes in the mouth are not meaningful.

2. If the Mouth NCC falls behind θ_{MAX}, it means that a significant change has started to occur in the mouth image, *e.g.*, due to tongue protrusion.

3. While the Mouth NCC stays above θ_{MIN}, we assume that the new mouth region we have obtained by tracking is still reliable and that the protrusion is not fully developed yet. Thus, we divide it into its left and right halves. We compare them with the LHT and RHT, respectively, obtaining two more correlation-based similarity measures. These coefficients show us how much each portion of the mouth has changed.

4. While the Mouth NCC stays between θ_{MAX} & θ_{MIN} (tongue protrusion transition), we continue the process previously described. For each template comparison, we find the LHT and RHT coefficients and update the average value, during this transition, for the NCC corresponding to the left and right halves.

5. Immediately after the Mouth NCC falls behind θ_{MIN}, the tongue protrusion position is detected by comparing the average NCC left/right values. We consider that the smaller the value is, the more its corresponding template has changed in average, i.e., if the average

Video	Length (sec.)	# Left Prot.	# Right Prot.
V001	30s	3	4
V002	36s	2	1
V003	31s	2	4
V004	25s	2	2
V005	33s	2	2
V006	17s	2	1
V007	19s	3	2
V008	19s	4	4
V009	36s	3	2
V010	36s	3	3
TOTAL	282s	26	25

Table 1. Ground truth of the tongue protrusion events.

right value is the smallest, it means that the right side template has changed the most, which is interpreted as a right side protrusion.

6. While the Mouth NCC stays below the θ_{MAX} level, the system waits for the protrusion to end, in order to allow the generation of a new protrusion event.

7. If at any moment, the NCC surpasses the θ_{MAX} value, the system automatically returns to the initial state. Thus, this algorithm allows for changes that are not as large as the tongue protrusion. For instance, in case of small head rotation or minor lip movement, only the θ_{MAX} threshold is surpassed, but not the θ_{MIN}; therefore, no detection is executed.

5. Results and Discussion

We have collected ten videos from seven different people inside an office environment. The users have been asked to follow some easy to learn guidelines to ensure the usability of the system:

- To keep the mouth closed at all times, including the case of a tongue protrusion event.
- In order to generate another protrusion event, the previous protrusion must be finished by returning the tongue back into the oral cavity.

Table 1 shows details about the database we are using.

A crucial part of setting-up the tongue protrusion detection system is the selection of the threshold parameters θ_{MAX} and θ_{MIN}. We have chosen θ_{MAX} candidates as 0.80 and 0.85; and, our θ_{MIN} candidates are 0.60, 0.65 and 0.70, which yields six possible combinations of thresold values. We have analyzed the videos for each one of these combinations, and we have manually labeled and counted the following tongue protrusion-related events:

1. Missed Events (ME): Protrusions that were not detected.
2. Detected Events: Protrusions that were detected. Events are further separated into:
   (a) Uncertain Event (UE): a decision could not be taken due to lack of transition information.
   (b) Misdetected Event (MDE): a decision was taken; however, Right was detected as Left, and viceversa.
   (c) Correctly Detected Event (CDE): a correct decision was taken.

The results from the evaluation using different parameters are shown in Table 2. In this table, Success Rate represents the ratio of correctly detected protrusion events to the total number of detected protrusion events as given by Equation 1.

$$Success\,Rate = \frac{CDE}{CDE + MDE + UE} \tag{1}$$

From Table 2, choosing the values $\theta_{MAX} = 0.85$ and $\theta_{MIN} = 0.70$ gives the best balance between Detection Accuracy and Detection Success Rate for the entire video database. However, for a specific video condition, there are cases in which using another threshold pair offers a better response, *e.g.*, for video *V006* choosing $\theta_{MAX} = 0.80$ and $\theta_{MIN} = 0.60$ yields a better result. While the chosen parameters are a good default option, it is possible to calibrate the system by configuring the threshold parameters that best fit the protrusion characteristics of the user. Figure 6 exemplifies clearly this trade-off between thresholds. For this specific case, setting $\theta_{MIN} = 0.85$ (red line) and $\theta_{MIN} = 0.70$ (blue line) yields the detection of eight protrusion events, being the sixth a false positive since it is actually another mouth-based gesture. The other events are detected accurately. However, if we initially set $\theta_{MIN} = 0.65$ (purple line), all of the events are detected satisfactorily, without any false positive. The same situation was found for other videos containing different users, such as if we set correctly, for each person, the pair or thresholds $(\theta_{MAX}, \theta_{MIN})$, the protrusion events are detected with high accuracy within that video sample.

Regardless of the parameters, the tracking process takes in average 30ms. for a 320×240-pixel image which demonstrates the real-time capability of the proposed algorithm.

5.1. Lower Threshold Parameter Discussion

There are some issues when decreasing the θ_{MIN} value. First, given the low similarity between templates when using this coefficient, the tracked mouth region may not be accurate causing the new detected mouth to be very poorly related to the original one. This subsequently affects the right

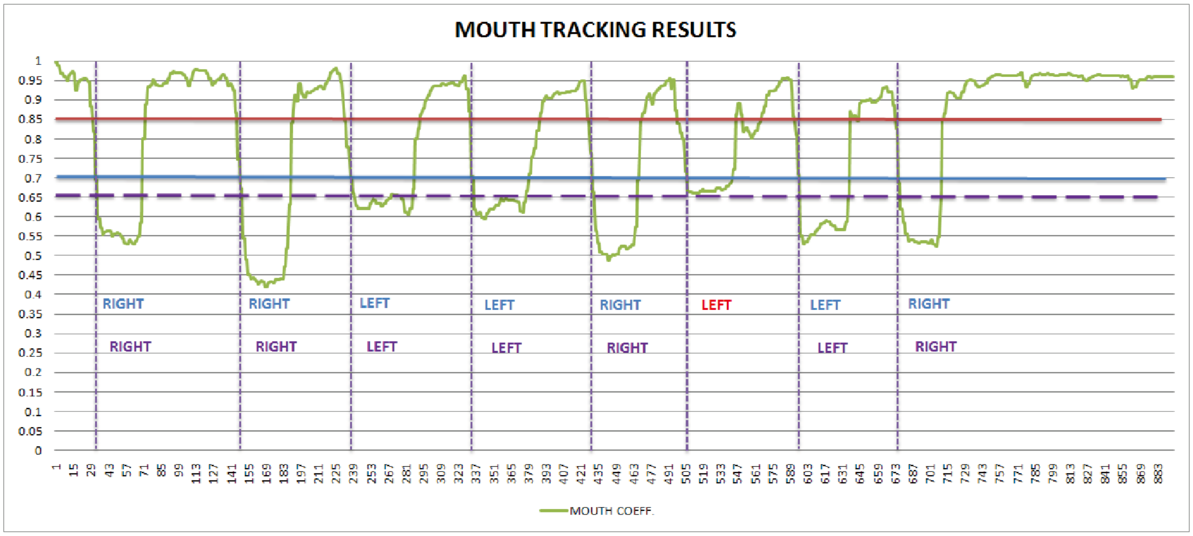

Figure 6. Results of Mouth tracking in V001 sequence. Parameters $\theta_{MAX} = 0.85$ and $\theta_{MIN} = [0.70, 0.65]$

and left templates used for the protrusion detection. Hence, the detection is not reliable any longer. Furthermore, having a lower θ_{MIN} value means that we require a larger change in the image for the protrusion to be detected, i.e., a larger portion of the tongue needs to be taken outside the mouth. However, a larger portion of protruded tongue means less mouth similarity compared to the original template, which causes the tracking to drift as discussed in Section 4.

θ_{MAX} / θ_{MIN}		0.80 0.60	 0.65	 0.70	0.85 0.60	 0.65	 0.70
V001	ME	1	0	0	1	0	0
	UE	0	0	2	0	0	0
	CDE	4/6	7/7	5/5	4/6	7/7	7/7
V002	ME	0	0	0	0	0	0
	UE	0	0	0	0	0	0
	CDE	2/3	1/3	3/3	2/3	1/3	3/3
V003	ME	6	0	0	6	0	0
	UE	0	0	1	0	0	0
	CDE	0/0	5/6	5/5	0/0	5/6	6/6
V004	ME	2	2	0	2	2	0
	UE	0	0	0	0	0	0
	CDE	2/2	2/2	3/4	2/2	2/2	3/4
V005	ME	1	1	1	1	1	1
	UE	0	0	0	0	0	0
	CDE	3/3	2/3	2/3	3/3	2/3	2/3
V006	ME	0	0	0	0	0	0
	UE	0	0	0	0	0	0
	CDE	3/3	2/3	1/3	3/3	2/3	1/3
V007	ME	4	1	0	4	3	3
	UE	0	0	0	0	0	0
	CDE	1/1	3/4	3/5	1/1	1/2	2/2
V008	ME	3	0	0	3	1	1
	UE	0	0	0	0	0	0
	CDE	5/5	5/8	7/8	5/5	5/7	6/7
V009	ME	1	0	0	1	0	0
	UE	0	0	0	0	0	0
	CDE	3/4	3/5	3/5	3/4	3/5	3/5
V010	ME	3	2	0	3	2	0
	UE	0	0	0	0	0	0
	CDE	2/3	4/4	6/6	2/3	4/4	6/6
OVERALL	ME	21	6	1	21	9	5
	UE	0	0	3	0	0	0
	CDE	25/30	34/45	38/47	25/30	32/42	39/46
Success Rate		83.33%	75.56%	76.00%	83.33%	76.19%	84.78%

Table 2. Evaluation of Threshold parameters for Table 1 data. **ME:** Missed Events, **UE:** Uncertain Events, **CDE:** Correctly Detected Events.

On the other hand, larger values of θ_{MIN} have the opposite effect: a smaller portion of the tongue now suffices for the protrusion to be detected. However, other actions such as head rotation or lip movement could affect the detection and trigger false positives. Additionally, while more events are detected, less time is given for the protrusion transition analysis. Therefore, we are forced to get an average of left and right coefficient templates based on very few transition frames. In an extreme case, it might happen that the tongue is protruded so quick that no transition data is obtained, which will prevent the system from taking a decision.

5.2. Higher Threshold Parameter Discussion

Setting θ_{MAX} to higher values results in a larger time window for tracking the protrusion transition. Moreover, considering that the correlation score is relatively high and the mouth region is accurate, the correlation coefficients obtained for the left and right template after the tracking are highly reliable. However, since θ_{MAX} is used as well for detecting when the protrusion event has finished, which requires a high similarity between the templates, it could lead to missing some other event or it could hinder the recovery of the system from the protrusion event.

On the other hand, decreasing θ_{MAX} will shorten the protrusion transition time. However, the system will be stronger against false negatives.

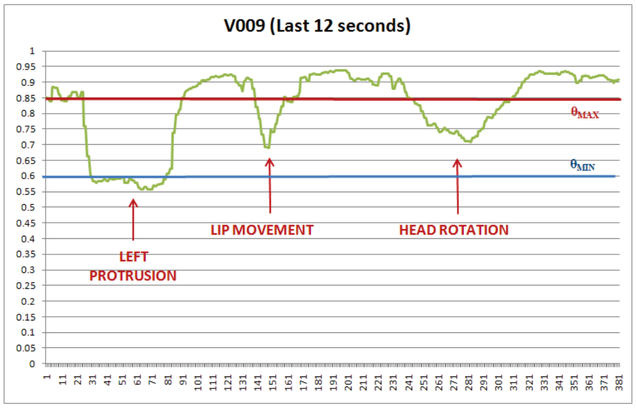

Figure 7. Results of Mouth tracking in V009 sequence. Parameters $\theta_{MAX} = 0.85$ and $\theta_{MIN} = 0.60$

5.3. Handling of Unwanted Events

After setting up a suitable pair of threshold values, our detection system is able to disregard events where the mouth region has not changed considerably. For instance, in Figure 7 we show the last twelve seconds of video *V009*, where a left protrusion of the tongue is followed by lip movement caused by swallowing saliva. Then, the head is rotated approximately 20° to both sides. Using the proposed method, only the left protrusion event is processed and detected because the two latter events do not cause the mouth NCC value to go below the θ_{MIN} threshold. Additionally, in *V002* the mouth is slightly open for breathing through it; nonetheless, this change is irrelevant to the system because of the same reason as the previous example. These unwanted events are repeated as well in other videos of the database and they are managed following the same principle.

The system proves to be robust against changes in the shape of the mouth that are not significant, giving freedom to the user for performing small head rotations and actions such as breathing or swallowing.

6. Conclusions and Future Work

In this paper we have proposed a system that detects in real-time the tongue protrusion and finds its position. The system relies on two threshold parameters θ_{MIN} and θ_{MAX} that have been deeply discussed. For the parameter selection, we consider the trade-off that exists between Missed Protrusion Events and the Success Rate of the Detection. In Figure 2 we can observe that for different sequences of video, there are different threshold pairs that minimize the missed events, while maximizing the correctly detected event. Thus, for designing an interface based on these gestures, these parameters can be adjusted according to the physical characteristics of the user, *e.g.*, for a person with a short tongue we can adjust both θ_{MAX} and θ_{MIN} to a higher value compared to a user who can protrude a larger portion of the tongue. In other words, by adjusting the threshold values we are adjusting the sensitivity of the response of the system. This calibration can be executed at the beginning by asking the user to protrude the tongue. Then, by evaluating the coefficients obtained in this action, we can obtain a good pair of threshold values.

We introduced the *tongue protrusion transition* concept, which we employ in order to detect the position where the tongue has been protruded. In our experiments, the tongue protrusion is carried out at normal speed and the video is captured in real-time. However, we need to set up suitable threshold parameters so as not to lose important transition frames that are necessary for the analysis.

Our future work includes the development and evaluation of an interface that will allow to use the proposed communication method in everyday life. For instance, the left and right protrusion can be used for emulating the left and right click of the mouse or for emulating a *go previous-go next* command for hands-free browsing of photo albums. They can be used as well as a dual-switch input device that would grant physically disabled people the control of the screen cursor of the computer. Since our trial videos are relatively short, it is necessary to evaluate the usability of the system during a long period of time.

Finally, although the system is robust against non-significant changes in the mouth region, it restricts the user to keeping a tight control of the mouth. Therefore, it is necessary to develop means of controlling other actions such as speaking or yawning so the user can feel more relaxed in the long term. Since the mouth region changes significantly for those actions, a method that tracks the mouth more efficiently is required. For instance, the OKAO Vision library by Omron Corporation [10] can be used for the tracking, before comparing the threshold values for each half part of the mouth.

Acknowledgement

Hamid Laga is supported by the Japanese Ministry of Education, Culture, Sports, Science and Technology program Promotion of Environmental Improvement for Independence of Young Researchers under the Special Coordination Funds for Promoting Science and Technology.

References

[1] M. Anisfeld. Only tongue protrusion modeling is matched by neonates. *Developmental Review*, 16:149–161(13), June 1996.

[2] G. Bradski and A. Kaehler. *Learning OpenCV: Computer Vision with the OpenCV Library*. O'Reilly, 2008.

[3] J. W. Davis and S. Vaks. A perceptual user interface for recognizing head gesture acknowledgements. In *PUI '01: Pro-*

ceedings of the 2001 workshop on Perceptive user interfaces, pages 1–7, New York, USA, 2001. ACM.

[4] C. Fagiani, M. Betke, and J. Gips. Evaluation of tracking methods for human-computer interaction. In *Applications of Computer Vision, 2002. (WACV 2002). Proceedings. Sixth IEEE Workshop on*, pages 121–126, 2002.

[5] K. Grauman, M. Betke, J. Lombardi, J. Gips, and G. R. Bradski. Communication via eye blinks and eyebrow raises: Video-based human-computer interfaces. *Universal Access in the Information Society*, 2(4):359–373, 2003.

[6] X. Huo, J. Wang, and M. Ghovanloo. A magneto-inductive sensor based wireless tongue-computer interface. *Neural Systems and Rehabilitation Engineering, IEEE Transactions on*, 16(5):497–504, Oct. 2008.

[7] A. P. Kencana and J. Heng. Experiment on a novel user input for computer interface utilizing tongue input for severely disabled. In *i-CREATe '07: Proceedings of the 1st international convention on Rehabilitation engineering & assistive technology*, pages 114–117, New York, USA, 2007. ACM.

[8] J. Matas. Robust detection of lines using the progressive probabilistic hough transform. *Computer Vision and Image Understanding*, 78(1):119–137, 2000.

[9] newAbilities Systems. The tonguetouchkeypad (ttk), August 2009. http://www.newabilities.com.

[10] Omron. Okao vision, August 2009. http://www.omron.com.

[11] L. R. Sapaico, H. Laga, and M. Nakajima. Mouth region localization based on gabor features and active appearance models. In *NICOGRAPH International*, Kanazawa, June 2009. The Society of Art and Science.

[12] L. R. Sapaico and M. Nakajima. Toward a tongue-based task triggering interface for computer interaction. *Applications of Digital Image Processing XXX*, 6696(1):66960J, 2007.

[13] M. B. Stegmann, B. K. Ersbøll, and R. Larsen. FAME - a flexible appearance modelling environment. *IEEE Transactions on Medical Imaging*, 22(10):1319–1331, 2003.

[14] L. Struijk. An inductive tongue computer interface for control of computers and assistive devices. *Biomedical Engineering, IEEE Transactions on*, 53(12):2594–2597, Dec. 2006.

[15] C. Tomasi and R. Manduchi. Bilateral filtering for gray and color images. In *Computer Vision, 1998. Sixth International Conference on*, pages 839–846, Jan 1998.

[16] M. Turk and G. Robertson. Perceptual user interfaces (introduction). *Commun. ACM*, 43(3):32–34, 2000.

[17] P. Viola and M. J. Jones. Robust real-time face detection. *Int. J. Comput. Vision*, 57(2):137–154, 2004.