Adaptive Lip Feature Point Detection Algorithm for Real-Time Computer Vision-Based Smile Training System

Youngkyoon Jang and Woontack Woo

GIST U-VR Lab.
Gwangju 500-712, S. Korea
{yjang,wwoo}@gist.ac.kr

Abstract. This paper presents an adaptive lip feature point detection algorithm for the proposed real-time smile training system using visual instructions. The proposed algorithm can detect a lip feature point irrespective of lip color with minimal user participation, such as drawing a line on a lip on the screen. Therefore, the proposed algorithm supports adaptive feature detection by real-time analysis for a color histogram. Moreover, we develop a supportive guide model as visual instructions for the target expression. By using the guide model, users can train their smile expression intuitively because they can easily identify the differences between their smile and target expression. We also allow users to experience the smile training system using the proposed methods and we evaluated the effectiveness of these methods through usability tests. As experimental results, the proposed algorithm for feature detection had 3.4 error pixels and we found that the proposed methods could be an effective approach for training smile expressions in real-time processing.[1]

Keywords: Lazy algorithm, lip detection, smile expression recognition, Haar-like classifier.

1 Introduction

Nonverbal information, such as facial expressions, play an important role in human communication [1] because facial expressions visually transmit feelings and intention [2-3]. In particular, smile expressions are an effective way to communicate positively with others [4]. For that reason, various smile training methods have been introduced, such as text books, magazines, and self-image making institutes, to train smile expression techniques. These methods, however, are quite restricted, as they allow for only limited interaction between the user and the training material, especially when relying on text books. Even self-image making institutes, while offering detailed advice from counselors and experts, tend to be time consuming and expensive.

[1] This research is supported by Korea Culture Content Agency(KOCCA) of Ministry of Culture, Sports and Tourism(MCST) in the Culture Technology(CT) Research & Development Program 2009.

M. Chang et al. (Eds.): Edutainment 2009, LNCS 5670, pp. 379–389, 2009.

As a solution to this problem, Kyoko Ito proposed a smile training system, called the 'Facial Expression Training System,' which provided a computer based process for effective facial expression training [5-7]. Ito's system involves four procedures to train a user's facial expression. First, the system registers a general image of the user's face. In the second procedure, the system selects the target facial expression and then determines the current facial expression in the third procedure. For the fourth procedure, the current and target facial expressions are compared. Several points are then marked on the currently registered facial image, representing areas where more expression or training is needed. Through Ito's four procedures, users can more accurately identify weaknesses and train more effectively.

Ito's work, however, does contain several limitations. The first problem is that the system does not support real-time processing and too many steps are required to conduct it. Moreover, the system cannot process more than one image, allowing users to use only one still image at each trial to get feedback. This inefficient procedure does not support seamless and intuitive training because a user has to consider only the marked points on the screen to understand the descriptions from the system. The descriptions, which function similar to a text book and do not easily allow the user to follow the target expression. To provide a more effective smile training system, the system needs to show both the user's facial image and target expression while running the algorithm in real-time. Then, users can intuitively compare their smile expression with the target expression and train efficiently through real-time feedback.

In this paper, we propose an adaptive lip feature detection algorithm for supporting the real-time smile training system using visual instructions. The proposed algorithm can detect lip feature points, irrespective of lip color, by a user-drawn line on a lip on the screen. Consequently, the proposed algorithm supports adaptive feature detection through a real-time analysis for a lip color histogram. In order to support intuitive comparison of a user's smile with a target expression, we develop a supportive guide model to provide visual instructions for the target expression. By using this guide model, users can train their smile expression intuitively because users can easily recognize the differences between their smile and the target expression. We also allow users to experience the smile training system using the proposed methods and we evaluated the effectiveness of the proposed methods through demonstration and usability tests.

The rest of this paper is organized as follows. Section 2 introduces the proposed real-time smile training system using the proposed algorithm and a detailed description of our proposed adaptive lip feature point detection algorithm is presented in Section 3. Section 4 describes the implementation and experimental results of the proposed system and algorithm and conclusions and future works are discussed in Section 5.

2 The Proposed Smile Training System and Algorithm

2.1 The Proposed Real-Time Smile Training System

We present the proposed real-time smile training system using the proposed algorithm, as shown in Figure 1. The proposed system consists of three parts: input devices,

computing device, and display device. A camera on the input device captures a user's facial image and the touch screen receives information through touched pixel points by a line drawn on the lip region of the image. The obtained facial image and touched pixel points are transmitted to the input source manger of the computing device.

The input source manager examines whether the input sources (image and points information) exists or not. When the input sources exist, the source manager forwards them to the feature extractor, which analyzes a color histogram of the touched pixels of the facial image and extracts feature points on the lip of the user's mouth. In the case of the feature points of the guide model, there is no color analysis process because the guide model manager delivers direct feature point coordinates as well as a guide model to the feature matcher. The guide model manager then loads and delivers the model and feature points coordinates from the guide model DB. The extracted feature point coordinates for the user's lip are transmitted only to the feature matcher, which determines whether a user smiles or not by measuring and comparing the distance between a user's lip feature points with the model. Depending on the results, the event manager then selects either visual instructions as tips or visual feedback to help produce a better match.

The visual instruction, a black line connected between the corners of a user's mouth and the guide model, represents the directions for the corners of the mouth where more training is required. The visual feedback also shows the intuitive feedback by changing the color of the supportive guide model when the user successfully trains his/her smile expression. The selected events are displayed through the display

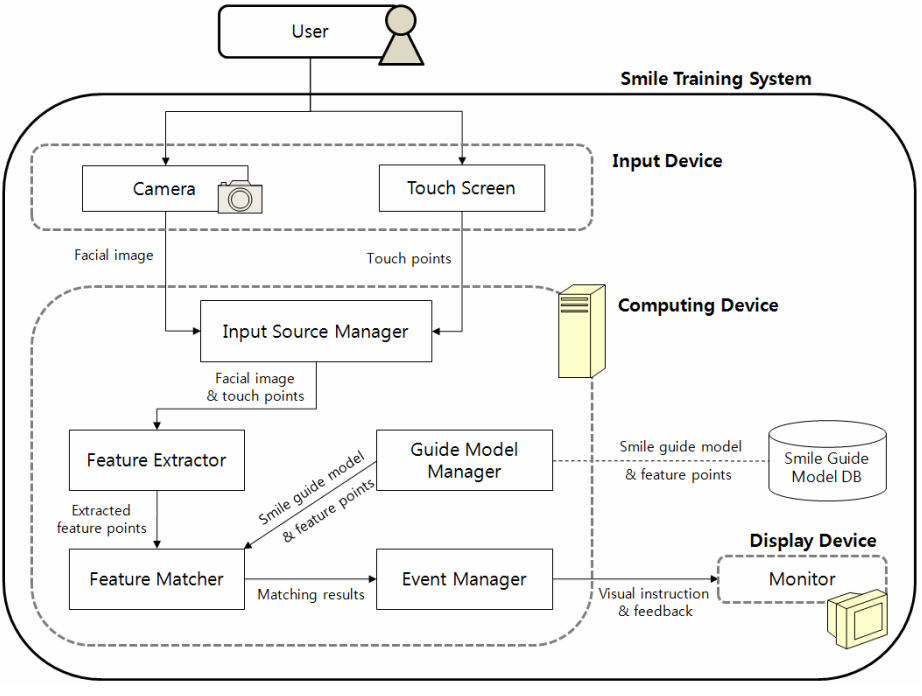

Fig. 1. Proposed the real-time smile training system diagram

device. This proposed method is more perceptive than simply following the comments from a picture as proposed in previous work [7].

2.2 The Proposed Algorithm Flow

An overview of our proposed algorithm can be seen in Figure 2. First, the proposed system captures a facial expression image, obtains touch points from the user's input, and detects mouth ROI (Region of Interests) by using the Haar-like classifiers [8-9]. Next, the corners of mouth are extracted by using the Harris Edge Detector [10]. The proposed system then detects the upper and lower feature points on the mouth by the proposed adaptive lip feature point detection algorithm, as described in section 3. The distances between the extracted feature points of the user's lip and the model are calculated to measure the differences between the user's lip shape and the target expression. When the measured value is lower than the predetermined threshold, indicating a high level of similarity, the system determines that the user's facial expression is a smile, and vice versa. If users have trained their smile expressions with the proposed system, but still want further training, another guide model shape is presented. These procedures can be repeated until users decide to stop.

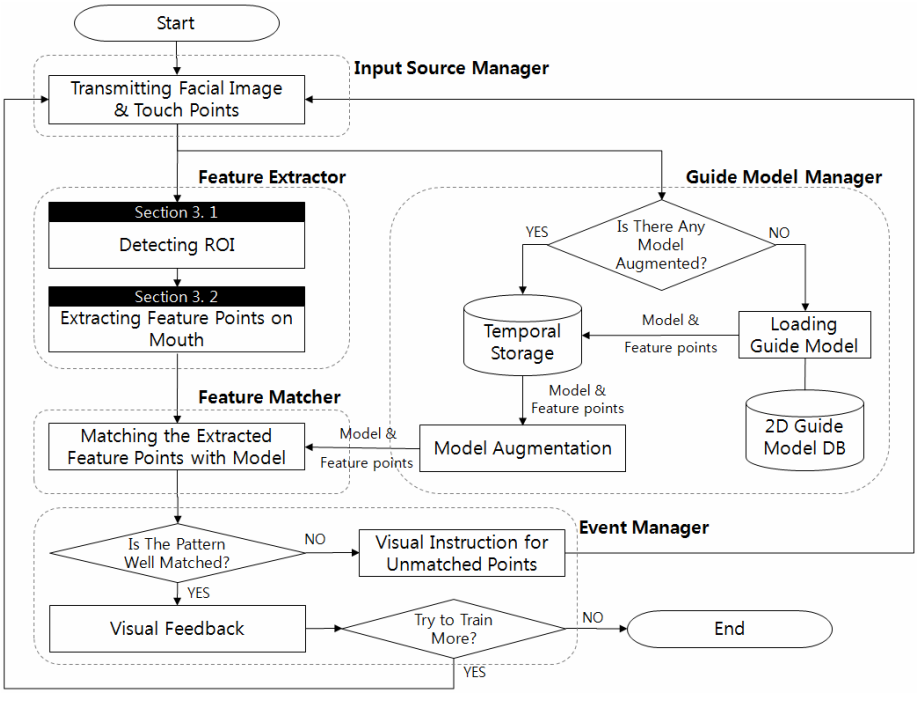

Fig. 2. Proposed algorithm for the real-time smile training system

3 The Adaptive Lip Feature Point Detection Algorithm

Because of the color variances that can occur among users, a smile training system needs to be able to detect lip features irrespective of lip color. To this end, we propose an adaptive lip feature point detection algorithm, used in the 'Feature Extractor' aspect of the system, as shown in section 2.2. Detailed explanations are provided as follows.

3.1 Detecting ROI

The proposed real-time smile training system focuses on the mouth region for training smile expressions. Therefore, before detecting the lip region, we detect the mouth ROI (Region of Interests) using a Haar-like classifier [8-9] integrated with the OpenCV library [11]. One problem with the Haar-like classifier, however, is that when the background of the captured image has a similar color or characteristics with mouth region, the classifier recommends too many candidates for the ROI, many of which are poor. To overcome this problem, we use a checking procedure for the geometric relations between the detected ROIs, as shown in Figure 3. First, we detect the frontal face ROI by using the Haar-like classifier [12-13] and then divide it into two parts, an upper half and a lower half. Using these two parts of the detected frontal face image, we can then detect the ROI for the eyes in the upper image and the mouth in the lower image. Through this procedure, we are able to determine whether or not the detected frontal face ROI is real. If there is no ROI for the eyes and mouth in the detected face ROI, the proposed method will determine that the face ROI is wrong. Therefore, we can detect the real face, eyes, and mouth ROI in more detail (see Figure 6 later in this paper), making the proposed method less affected from noise and reducing the processing time, as shown in Table 1.

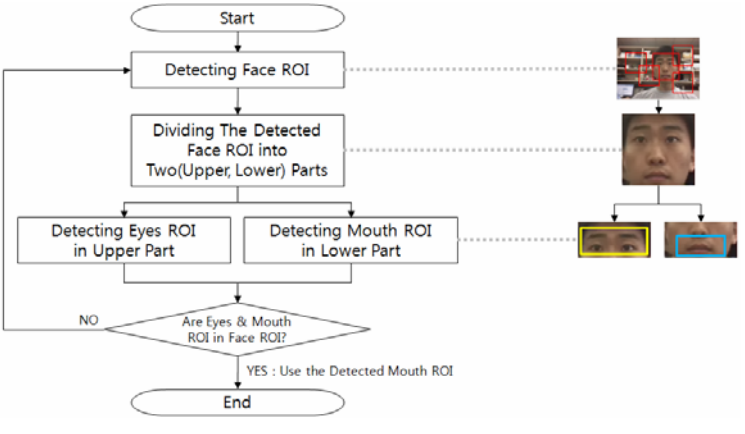

Fig. 3. The proposed procedure to check geometric relation between the detected ROIs

3.2 Extracting Feature Points on Mouth

To recognize the precise position or shape of the mouth, it is necessary to first detect more detailed feature points. Before detecting the precise feature points on an outline of the lip, though, we first detect the corners of the mouth using the Harris Edge detector [10]. We then select both the left and right-most points among the detected edges of the mouth. The predefined threshold for extracting the proper area for the edges of the mouth ROI was determined heuristically to indicate optimal performance.

Our proposed algorithm analyzes the lip color histogram information to accurately extract the lip feature points irrespective of lip color, with some user participation, as shown in Figure 4. The proposed method uses a YC_bC_r color space to analyze the lip color histogram. For each color space, the proposed method calculates the mean and standard deviation values by using all pixel values of the points on the drawn line, as shown in Eq. (1) and (2), respectively. After that, we investigate all pixels of the mouth ROI within the range between the corners of the mouth and all pixels in the range $[C_{1x}, C_{2x}]$ and $[Y_{1y}, Y_{2y}]$ in the x-axis and y-axis, respectively. As shown in Figure 4, C_{1x} and C_{2x} represent the x-axis values of the left and right corners of the mouth, respectively, while Y_{1y} and Y_{2y} represent the y-axis values of the most upper and lower points on the detected mouth ROI, respectively. If a pixel value in the mouth ROI falls in the range [(M-SD), (M+SD)], the proposed method determines that the pixel is in the lip region and vice versa, as shown in Eq. (3), allowing the lip region to be segmented, regardless of lip color.

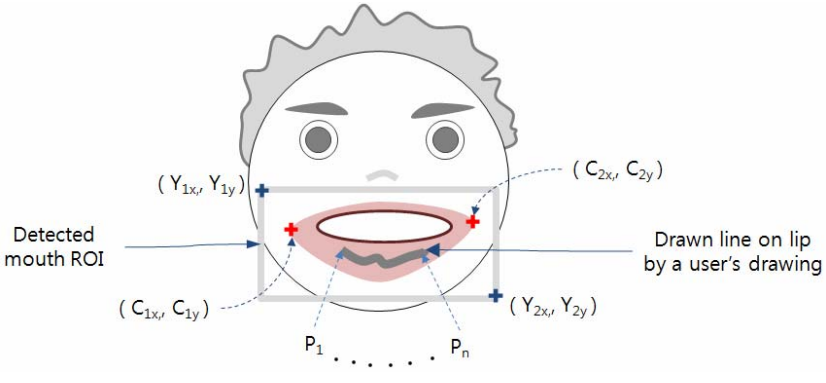

Fig. 4. User-drawn line on the lip for the adaptive feature detection algorithm

$$M = \frac{1}{n}\sum_{i=1}^{n} P_i, \tag{1}$$

$$SD = \sqrt{\frac{1}{n}\sum_{i=1}^{n} (P_i - M)^2}\,, \tag{2}$$

$$F(CPV) = \begin{cases} True, & (M - SD) < CPV < (M + SD) \\ False, & CPV < (M - SD), or (M + SD) < CPV \end{cases}, \tag{3}$$

where n is the number of the points on the line drawn on the lip. P represents the pixel values of a point on the drawn line, as shown in Figure 4. M and SD are mean and standard deviation values of the points on the drawn line. CPV refers to the current pixel value to be evaluated and the function F determines whether or not a pixel is in the lip region.

After segmenting the lip region, the proposed algorithm detects the candidate points in the middle region of the detected mouth ROI by selecting the outside points of the segmented lip region, as shown in Figure 5. Then, according to Eq. (4), the proposed algorithm extracts the feature points for the upper and lower lips, respectively.

$$\begin{aligned} FP_{Uy} &= \frac{1}{n}\sum_{k=1}^{n} P_U(y_k), \\ FP_{Ly} &= \frac{1}{n}\sum_{k=1}^{n} P_L(y_k), \\ FP_{Ux} &= FP_{Lx} = \frac{1}{2}(C_{1x} + C_{2x}) \end{aligned}, \tag{4}$$

where n is the number of the detected candidate points and $P_U(y_k)$ and $P_L(y_k)$ are the vertical positions of the candidate point (y_k) for the upper and lower lips, respectively. FP_{Uy} and FP_{Ly} are the y-axis values of the extracted feature points for the upper and lower lips, respectively. FP_{Ux} and FP_{Lx} are x-axis center point values of the corners of mouth and (FP_{Ux}, FP_{Uy}) and (FP_{Lx}, FP_{Ly}) respectively represent the coordinates of the extracted feature points for the upper and lower lips.

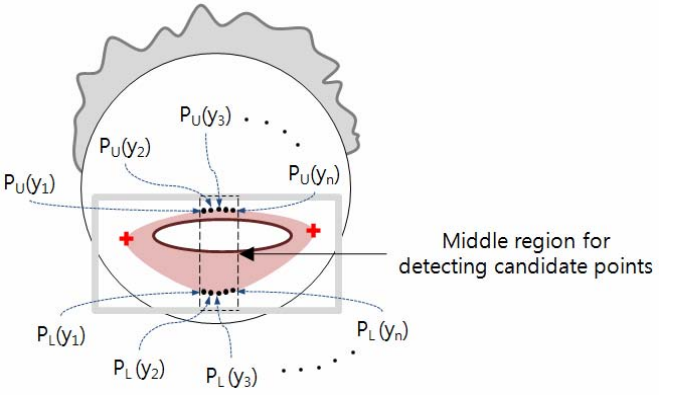

Fig. 5. Detecting candidate points on lip outline before extracting feature points

4 Implementation and Experimental Results

4.1 Implementation

In this work, we implemented a real-time smile training system using the proposed adaptive lip feature point detection algorithm which is irrespective of lip color. We used an OpenCV library [11] for capturing images from a camera and implementing basic image processing algorithms. The proposed system was implemented based on Visual C++ 2005. We used the built-in camera and display on a laptop computer for implementing the proposed algorithm. Instead of using a touch screen, we used a computer mouse for touch input, such as drawing a line on the lip.

We detected the lip feature points by using the proposed algorithm. First, in order to detect the mouth ROI, we used a Haar-like classifier integrated in the OpenCV library. Additionally, by checking the geographical relation between the detected face, eyes and mouth ROI, we were able to precisely detect ROIs, as shown in Figure 6. Moreover, we also detected four feature points on the lip by using the proposed adaptive lip feature detection algorithm irrespective of lip color, as shown in Figure 7. According to Tian's work, the lip shape can be represented by four points on the lip outline [14-15].

Because noise is always present in an image, however, the extracted feature points are shown as quivering in the real-time rendering. Therefore, we used a temporal storage to save the extracted feature points. When the current feature points on the lip were extracted, the proposed method calculated the mean position value among the currently extracted points and the previously stored points. By using the mediated position value of the feature points, the extracted points are able to be tracked more fluidly.

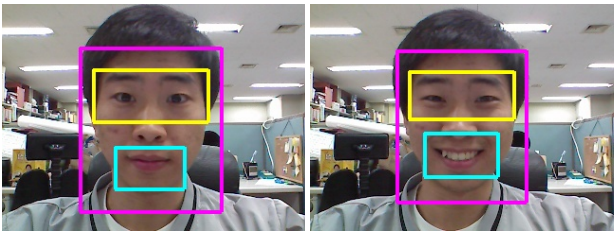

Fig. 6. Detected ROIs for frontal face, eyes and mouth

Fig. 7. Examples of the detection result of the lip feature points on lip

In order to support more intuitive training for a user's smile expression, we used the supportive guide model for the proposed system, as shown in Figure 8. By using the supportive guide model, users can intuitively recognize their target smile expression and can easily identify which direction they have to move to achieve it. By using the proposed visual instruction and feedback, users can more accurately and interactively train their smile expression simply by following the visual directional instructions, as shown in Figure 8(a). This visual instruction is shown by connecting lines from the detected corners of the user's mouth to the corners of the guide model. When a user smiles correctly, the supportive guide model changes the color of the visual feedback, as shown in Figure 8(b).

(a) (b)

Fig. 8. Examples of the visual instruction and feedback: (a) visual instruction showing the connected black line between corners of the mouth and model (b) visual feedback through the model's color change

4.2 Experimental Results

In this paper, we evaluated the performance of our proposed algorithm by using the captured frontal face image DB. The frontal face image DB is composed of 1,000 images acquired from 100 different facial expressions of 10 people using the proposed system. Among the 100 different expressions for each person, the first 50 are non-smile expression images, while the others are smile expression images. Each image in the frontal face image DB has a 24-bit color value with a pixel size of (320*240). We used a laptop computer with an Intel Core 2 Duo 2.4 GHz processor and 2GB SDRAM.

First, we measured the processing time for the proposed algorithm. As a result, our proposed system required 90.55 ms to run in our experimental environment. Specifically, ROI detection required 87.16 ms, the majority of the processing time, while our proposed algorithm required 3.35 ms to detect four feature points on a lip, as shown in Table 1. Even without optimization for the source code, the processing time was fast. With optimization, we could therefore expect even faster results during feature extraction.

In the second set of tests, we measured the detection accuracies of the corners of the mouth and feature points of the lip, shown in Table 2. The accuracies were measured based on the RMS (Root Mean Square) pixel error between the automatically detected points and those chosen manually. As can be seen in Table 2, the average

Table 1. Processing time of the proposed smile training system and algorithm

Algorithm - #1, System - #2		Processing time (ms)		
		Minimum	Maximum	Average
1	ROI detection	76.15	150.83	87.16
	Feature point detection	**1.83**	**5.37**	**3.35**
2	Entire processing time	78.82	154.79	90.55

RMS pixel error indicates a satisfactory working performance because the error pixel is less than the threshold value used for matching. Based on this, we determined a threshold of 4 pixels for detecting the corners of the mouth and 5 pixels for feature points on the lip. If the matching distances were larger than the predetermined threshold, our proposed system determined the facial expression was not a smile.

Finally, we determined the usability of the proposed system and the effectiveness of our algorithm through comments from test subjects. In terms of the proposed system, most of the users provided positive feedback, noting that mirroring the current smile expression with the guide model and direct feedback through visual instruction were intuitive and made it easy to train their smile expression. However, users also commented that the system did not support a wide variety of smile guide models and users wanted to receive a description along with the visual instructions. Users also noted that they needed more indication of success, rather than just showing the color change on the guide model. In terms of the proposed algorithm, most of users indicated that the simple procedure of drawing a line on the lip was effective and easy to use. From these results, we can conclude that, overall, the proposed system using the proposed algorithm is helpful for users to efficiently train for their target expression, despite several limitations indicated by the users.

Table 2. Detection accuracies of the corners of mouth and feature points on the lip

	Average errors for each features (pixel)		Average errors (pixel)
Corners of mouth	Left	3.65	**3.31**
	Right	2.96	
Feature points on lip	Upper	2.53	**3.52**
	Lower	4.51	

5 Conclusions and Future Works

This paper proposes an adaptive lip feature detection algorithm for a real-time smile training system with visual instructions. By using the guide model as a target smile expression, visual instructions and feedback, users can easily and intuitively train their smile expression.

For detecting the feature points of a lip irrespective of lip color, the proposed method does require some user participation, such as drawing a line on the lip. However, the proposed method is able to extract the feature points of a lip accurately. As described in the experimental results, the RMS error pixels for detecting the feature

points using our algorithm was less than 4 pixels. Therefore, we can determine that the proposed real-time smile training system using the proposed algorithm can provide positive results for smile expression training.

As future research, we plan to exchange the 2D guide model with an AR (Augmented reality) guide model. Also, we plan to consider pose tracking for the free movement of a user's head and the appropriate augmentation of the supportive AR guide model. By augmenting the supportive AR guide model to the correct position, the proposed system will provide seamless augmentation for the supportive AR guide model and no restriction of a user's head movement. We also plan to consider other components of smile expression, such as the eyes and furrow. Further, we would like to explore the possibility of running a cell-phone based algorithm to enhance convenience for the user. The user comments from this study, however, will be a primary consideration in any future undertaking in this area.

References

1. Kurokawa, T.: Nonverbal interface. Ohmsha, Ltd., Tokyo (1994) (in Japanese)
2. Yoshikawa, S.: Facial expression as a media in body and computer, pp. 376–388. Kyoritsu Shuppan Co., Ltd. (2001) (in Japanese)
3. Uchida, T.: Function of facial expression. Bungeisha, Co., Ltd. (2006) (in Japanese)
4. Mehrabian, A.: Silent messages, 2nd edn. Implicit Communication of Emotions and Attitudes. Wadsworth Pub. Co. (1981)
5. Ito, K., Kurose, H., Takami, A., Nishida, S.: Development of Facial Expression Training System. In: Smith, M.J., Salvendy, G. (eds.) HCII 2007. LNCS, vol. 4557, pp. 850–857. Springer, Heidelberg (2007)
6. Ito, K., Kurose, H., Takami, A., Nishida, S.: Development and Application of Facial Expression Training System. In: Holzinger, A. (ed.) USAB 2007. LNCS, vol. 4799, pp. 365–372. Springer, Heidelberg (2007)
7. Ito, K., Kurose, H., Takami, A., Nishida, S.: iFace: Facial Expression Training System. Affective computing, 319–328 (2008)
8. Wilson, P.I., Fernandez, J.: Facial Feature Detection Using Haar Classifiers. Journal of Computing Sciences in Colleges 21(4), 127–133 (2006)
9. Castrillón-Santana, M., Déniz-Suárez, O., Antón-Canalís, L., Lorenzo-Navarro, J.: Face and Facial Feature Detection Evaluation. In: International Conference on Computer Vision Theory and Applications (VISAPP 2008) (2008)
10. Harris, C., Stephens, M.: A Combined Corner and Edge Detector. In: Proceedings of the 4th Alvey Vision Conference, pp. 147–151 (1988)
11. Intel Open Source Computer Vision Library, `http://sourceforge.net/projects/opencvlibrary/`
12. Kruppa, H., Castrillón Santana, M., Schiele, B.: Fast and Robust Face Finding via Local Context. In: Joint IEEE International Workshop on Visual Surveillance and Performance Evaluation of Tracking and Surveillance (VS-PETS), pp. 157–164 (2003)
13. Viola, P., Jones, M.J.: Robust Real-time Face Detection. International Journal of Computer Vision 57(2), 137–154 (2004)
14. Tian, Y., Kanade, T., Cohn, J.F.: Recognizing Action Units for Facial Expression Analysis. IEEE Transactions on Pattern Analysis and Machine Intelligence 23(2), 97–115 (2001)
15. Tian, Y., Kanade, T., Cohn, J.F.: Robust Lip Tracking by Combining Shape, Color and Motion. In: ACCV 2000, pp. 1040–1045 (2000)