Lip Language Recognition for Specific Words

Yun-Long Lay, Hui-Jen Yang, Chern-Sheng Lin, Bin-Fen Lee

Abstract

In the human language communication process, facial expression and lip shape change of the teller contains extremely rich language information. Previous research about lip-reading recognition focused on proposing the theoretical method to analyze rules of lip shapes. However, little evidence appears that the real lip reading recognition system was implemented. For getting moving lip features, principal component analysis and mouth changing rate were put forward by this paper. We implement a lip language recognition system by using image processing technology, neural network algorithm, and database to help people and computer understand lip language. The dynamic image variation of the lip shape changing could be detected in this system. The results indicated that this method could recognize effectively and correctly 62 Chinese words with 55 users and practically enhanced the lip language recognition.

Keywords: lip reading, dynamic image recognition, principal component analysis

1. Introduction

Owing to the fast growing technology of image processing and computer hardware, computer input interfaces have been becoming more and more important and necessary within the past few decades. Many research institutes have devoted their efforts toward automated lip reading and obtained some useful results in this decade. Some articles act as references on lip reading research. However, most of these articles only discuss the methods of lip shape feature grabbing or tracking, not a complete lip reading recognition system (Chiang et al., 2003; Yao et al., 2010). In this research, a lip reading recognition system is implemented. Supposing a user sits in front of the CCD camera, this system could track the lip position from the user’s face to recognize whether the lip changing of a word pronunciation is correct or not. This system could thus be used as a lip shape control interface.

Until the writing of this paper, very limited lip-language application research was conducted as a useful tool to enhance the communication of computer. The arguments are that only the pronunciations for those characters with the changes of obvious labiates and oral area outward appearance are possibly recognized by lip language. Many pronunciations only used behind the throat or the oral cavity character are not easily recognized by lip language. Therefore the general language probably is that only 30% of Mandarin characters can be recognized by lip language.

Although lip language identification is not entirely accurate, there are still many people who do not have good hearing ability and rely on lip language every day, especially in noisy environments. According to the research results of the University of Manchester (Bauman, 2003), the hearing impaired subjects could only recognize 21% of spoken language. If they used a hearing aid, the recognition rate of speaking could be increased to 65%. If they simultaneously used a hearing aid and lip language, the recognition rate of speaking could reach 90%. Thus, if people want to raise their communication ability, using lip reading should significantly improve their communicative ability. Therefore, implemented a lip language recognition system by using image processing technology, neural network algorithm, and database to create a human computer interactive system. The dynamic image variation of the lip shape changing could be detected in this system for using as a lip shape control interface.

2. Image Processing Steps

The pre-processing of the lip reading image is to get the lip portion from the colored image by localizing the lip position. These steps include brightness adjustment, color space transforms, face color cutting, facial operation region, and lip localization. The lip portion of the image obtained from the pre-processing will be the identification target of all subsequent processes. The recognition portion uses the principal component analysis (PCA) to identify the lip-changing rate as the lip characteristic values. Those lip characteristic values were then learned by a Neural Network to allow the lip image analysis to acquire recognition results. The steps’ flow chart of image processing is shown in Figure 1.
3. Position Detection and Lip Tracking

In this research, skin color is the major feature to find the lip position for implementation of the recognition system. The system input is a color image, which is grabbed by a CCD camera. The facial area detection locates the human face region by the following steps, including lighting compensation, color segmentation, and skin color detection.

3.1 Lighting Compensation

The traditional Gamma Correction is an image adjustment skill relying on a kind of power operation to change the brightness of an image, which can refine gray scale images (Lin et al, 2012; Gonzalez & Woods, 2002; Lin et al, 2011). The Gamma Correction function can be shown as Eq. (1).

\[ V_{out} = c \times (V_{in} + e)^{\gamma} \]  

(1)

Where \( V_{in} \) is the gray scale of input pixel; \( V_{out} \) is the gray scale of output pixel; \( c \) and \( \gamma \) are the constants; \( e \) is the value of parallel shift. The \( e \) value is small and can be neglected.

The traditional Gamma procedure uses a constant Gamma to process the image so that it cannot have a better performance for the adjustment. Hence, used an adaptive Gamma procedure to get better gray scale enhancement as seen below in Eq. (2).

\[ V_{out} = 255 \times \left( \frac{V_{in}}{255} \right)^{1/(\gamma(V_{in})} \]  

\[ \gamma(V_{in}) = 1 + a \times \cos(p(V_{in})) \]  

(2)

\[ p(V_{in}) = \begin{cases} \pi V_{in} / 2 x_0 & , V_{in} \in [0, x_0] \\ \pi/2 & , V_{in} \in [x_0, x_1] \\ \pi \pi/(255-V_{in})/2(255-x_1) & , V_{in} \in [x_1, 255] \end{cases} \]

Where the gray scale of the input pixel \( V_{in} \) is between 0 and 255; \( V_{out} \) is the gray scale of the output pixel; \( a \) is a constant 0.5; \( x_0 \) and \( x_1 \) are the two threshold values of the two ends. This method could be used in color or black and white images. When it is used in color images, the Red, Green, and Blue color data can be processed one by one.

Figure 2 compares the bright and dark parts of the images by traditional Gamma and adaptive Gamma. The adaptive Gamma process Figure 2. (c) has better performance in comparing the bright and dark parts of images than the traditional Gamma Figure 2 (a) Gamma=0.6 and (b) Gamma=1.4.

3.2 RGB and YCbCr Color space transformation

Skin color is an important recognition feature of the human face area. Different age and race causes different skin color. Regardless, brightness is a major factor to influence the skin color. When normalizing the brightness of skin images, the detection error of facial skin could be minimized and the skin portion could be detected by the integrity of skin color clustering. A suitable color model should be found to get rid of the brightness factor in a testing image (Rafael &. Richard, 2002; Lay et al, 2012).

For an RGB color image, the brightness value of each pixel in an image is composed of the values of RGB. Hence normalization of the brightness value of RGB ingredients will decrease the effect of brightness in the process of face skin detection. The normalization model of a RGB color image is as Eq. (3) (Yang et al., 2004).

\[ \begin{align*} 
    r &= R (R + G + B) \\
    g &= G (R + G + B) \\
    b &= B (R + G + B) \\
    \end{align*} \]

where \( r + g + b = 1 \)

\[ \begin{cases} 
    \text{Original} & \text{Corrected} \\
    \text{Original} & \text{Corrected} \\
    \text{Original} & \text{Corrected} \\
    \text{(a) } \gamma=0.6 & \text{(b) } \gamma=1.4 & \text{(c) Adaptive Gamma} 
\end{cases} \]
YCbCr is a reasonable color representing in color space. The image colors are stated by color intensity Y, as well as Cb and Cr which represent the blue- and red-difference. The YCbCr model is adapted by the image contraction format of MPEG and JPEG. This model is good for the needs of processing digital images. From Chai’s research (Chai &. Bouzerdoum, 2000), When observed, there is no big difference in human skin color while using YCbCr to grab the portion of skin color. The YCbCr normalization model is shown in Eq. (4):

\[
\begin{align*}
Y &= 0.299R + 0.587G + 0.114B \\
C_b &= -0.16875R - 0.33126G + 0.5B \\
C_r &= 0.5R - 0.41869G - 0.08131B
\end{align*}
\] (4)

3.3 The segmentation of skin portion

The normalization RGB model could decrease the effect of image brightness but it cannot decrease the effect of hue and saturation. Hence using both YCbCr and RGB models to segment the skin portion of the image would be a better method (Kong & Zhu, 2006). From experiments, the skin portion distribution of the RGB model is shown in Eq. (5). The face skin distribution map of r (red) and g (green) pixels are shown in Figure 3.

\[
\begin{align*}
0.36 &\leq r \leq 0.465 \\
0.28 &\leq g \leq 0.363
\end{align*}
\] (5)

Fig.3. The face skin distribution map of r and g pixels

\[
\begin{align*}
20 &\leq Y \leq 220 \\
100 &\leq C_b \leq 140 \\
120 &\leq C_r \leq 160
\end{align*}
\] (6)

Fig.4. The face skin distribution map of Y-Cb and Y-Cr pixels

The combination of YCbCr and RGB models could segment the skin portion of the face image. A test image processed by RGB and YCbCr created the binary output image as shown in Figure 5, where the skin portion of the face is clearly be segmented.

Fig.5. The segmentation of skin portion

3.4 Lip Color Segmentation

After the face skin has been segmented, the lip area is the next step to segment. To choose the suitable R/G range by RGB method (Hsu et al., 2002), the lip portion is easy to be grabbed as shown in Figure 6. From experiments, R/G range could be found and shown in Eq. (7).

\[
\begin{align*}
(R / G) &\geq 1.6 \\
(R / G) &\leq 2.0
\end{align*}
\] (7)

Fig.6. Lip color segmentation
3.5 The enhancement of lip feature

Most of the lip area consists of red rather than blue color. Hence, the lip area has higher Cr values and lower Cb values. Using this method, the lip area is easily enhanced and a clear lip shape can be obtained. Owing to the noise, some facial areas have shapes similar to that of the lips, which will cause false recognition. The inflation method is a good way to enhance the lip feature and decrease the recognition error. The algorithm is shown as Eq. 8 and the processing result of the lip feature and its enhancement are shown in Figures 7 and 8, respectively.

\[
\begin{cases}
\text{MouthMap} = C_r^2 \cdot (C_r^2 - \eta C_r / C_b)^2 \\
\eta = 0.95 \frac{1}{n} \sum C_r(x,y)^2
\end{cases}
\]

3.6 Region of Interest

In Figure 7(a), the vertical and horizontal projection values were used to locate the face and lip position. The statistical amount of vertical projection is a two-dimensional figure whose vertical values are the accumulated amount of every vertical pixel and whose horizontal axis values are the horizontal axis values of the testing image. The statistical values of horizontal projection are the accumulated amount of every horizontal pixel and the vertical axis values are the vertical axis values of the test image. Figure 7(b) shows the face skin pixel measurement. Figure 7(c) shows the lip pixel measurement. Figure 8 shows the located results of face and lip.

4. Lip Recognition

4.1 Feature Extraction – Principal Component Analysis (PCA)

Principal Component Analysis (PCA) is a method to convert all of the original variables to be some independent linear set of variables. Those independent linear sets of variables possess the most information in the original data called the principal components. PCA involves the calculation of the eigenvalue decomposition of a data covariance matrix or singular value decomposition of a data matrix (Shaw, 2003). PCA is the simplest method to analyze the true eigenvector-based multivariate. Generally, its operation can be thought of as revealing the internal structure of the data which can be the best way to explain the variance in the data. If a multivariate data set is a set of coordinates in a high-dimensional data space, PCA supplies the user with a lower-dimensional picture, a “shadow” of this object when viewed from its most informative viewpoint. The steps of feature extraction for PCA are described as follows.

Step 1. Calculate the Covariance Matrix C. Convert the N samples of two-dimensional image data to be N plies of one-dimensional data arrays. The result is shown as Figure 9.

\[
\begin{align*}
\text{Sample 1} & \quad \text{Sample 2} \\
\vdots & \quad \vdots \\
\text{Sample M} & \quad \text{Column vectors}
\end{align*}
\]
\[
\mu = \frac{1}{M} \sum_{i=1}^{M} X_i
\]  
(9)

Calculate the Difference vector D shown in Eq. (10).

\[
D_i = Y_i - \mu, \quad i = 1, ..., M
\]  
(10)

Plug Eq. (9) and Eq. (10) into Eq. (11) then the Covariance Matrix \( C \) can be completed.

\[
C = \frac{1}{M} \sum_{i=1}^{M} D_i D_i^T, \quad A = [D_1, D_2, ..., D_M]
\]  
(11)

Step 2. Calculate the eigenvalues and unit eigenvectors of the Covariance Matrix.

Step 3. Sort the eigenvalues as \( \lambda_1 \cdot \lambda_2 \cdot \lambda_3 ... \lambda_p \).

Step 4. Solve the mapping eigenvectors \( a_1, a_2, a_3 ... a_M \), where \( ai' ai = 1 \) and \( ai' aj = 0 \) then

\[
y_1 = a_1'x = a_{11}x_1 + a_{12}x_2 + ... + a_{1M}x_M, \quad \text{The 1st PCA}
\]

\[
y_2 = a_2'x = a_{21}x_1 + a_{22}x_2 + ... + a_{2M}x_M, \quad \text{The 2nd PCA}
\]

\[
\vdots \quad \vdots \quad \vdots
\]

\[
y_n = a_M'x = a_{M1}x_1 + a_{M2}x_2 + ... + a_{MM}x_M, \quad \text{The Mth PCA}
\]

The vector \( ai' \) represents the main axis of the first principle analysis. The coefficients of vector \( ai' \) are the eigenvalues which is needed, and are shown in Figure 10. Table 1 is the PCA data of the word “Mnn”.

| Table 1. PCA data of the word Mnn |
|-----------------------------------|
| 1.000000 | 1.000000 | 1.000000 | 1.000000 | 1.000000 | 1.000000 | 1.000000 | 1.000000 | 1.000000 |
| 0.988465 | 0.989733 | 0.979960 | 0.987274 | 0.982495 | 0.985544 | 0.989411 | 0.986827 | 0.991693 |
| 0.965993 | 0.966024 | 0.946911 | 0.951119 | 0.959848 | 0.960736 | 0.974371 | 0.967332 | 0.971497 |
| 0.955517 | 0.942149 | 0.940567 | 0.950070 | 0.947873 | 0.952070 | 0.960691 | 0.966180 | 0.969298 |
| 0.938144 | 0.926898 | 0.921842 | 0.926625 | 0.934878 | 0.954841 | 0.963025 | 0.958906 | 0.956065 |
| 0.916205 | 0.919092 | 0.905074 | 0.915891 | 0.934411 | 0.948669 | 0.943636 | 0.927511 | 0.935132 |
| 0.898193 | 0.904571 | 0.901210 | 0.903152 | 0.925188 | 0.906282 | 0.878200 | 0.870469 | 0.901911 |
| 0.880519 | 0.892669 | 0.895148 | 0.895577 | 0.883987 | 0.858942 | 0.843125 | 0.830140 | 0.848427 |
| 0.857510 | 0.857490 | 0.850796 | 0.852115 | 0.854378 | 0.859440 | 0.863735 | 0.842495 | 0.817587 |
| 0.840525 | 0.809539 | 0.799399 | 0.834493 | 0.863153 | 0.892678 | 0.901185 | 0.876191 | 0.825236 |
| 0.813399 | 0.791857 | 0.825334 | 0.868358 | 0.899661 | 0.899984 | 0.897049 | 0.876973 | 0.834125 |
| 0.821610 | 0.816803 | 0.857039 | 0.894822 | 0.898299 | 0.885110 | 0.877259 | 0.858822 | 0.825329 |
| 0.828828 | 0.832946 | 0.867202 | 0.893897 | 0.884953 | 0.884556 | 0.866993 | 0.841938 | 0.817999 |
| 0.823841 | 0.826564 | 0.840332 | 0.869683 | 0.862388 | 0.867127 | 0.853605 | 0.828179 | 0.793817 |
| 0.813333 | 0.811716 | 0.819394 | 0.856001 | 0.847578 | 0.839010 | 0.844226 | 0.820572 | 0.774378 |
| 0.798731 | 0.796999 | 0.823257 | 0.841060 | 0.842567 | 0.855489 | 0.866062 | 0.855332 | 0.785186 |
| 0.798884 | 0.779335 | 0.817649 | 0.840040 | 0.837814 | 0.873744 | 0.879381 | 0.861338 | 0.792813 |
| 0.786479 | 0.758362 | 0.796957 | 0.851833 | 0.832109 | 0.851884 | 0.857037 | 0.840377 | 0.772261 |
| 0.775156 | 0.737276 | 0.775855 | 0.818029 | 0.809667 | 0.823033 | 0.834136 | 0.808425 | 0.747077 |
| 0.764871 | 0.729425 | 0.752822 | 0.780546 | 0.787288 | 0.799391 | 0.798345 | 0.778842 | 0.734709 |

4.2 Feature extraction - The lip shape changing rate

The steps of feature extraction for lip shape changing rate are described as follows.

Step 1. Find the length and width of the lips. The lip image was processed by edge detection algorithm to find the contour the lips. The result is shown as Figure 11.

Fig.11. The height and width of a testing lip

Step 2. Calculate the lip changing rate \( V \) with Eq. (12) as its algorithm.

\[
V_i = (W_{i+1} H_{i+1}) - (W_i H_i), \quad i = 1, 2, ..., n - 1
\]  
(12)
where V is Mouth’s Variable Rate.
W is Mouth’s Width.
H is Mouth’s Height.
I is the image order.

Mouth changing rate of five words sample is shown as Figure12.

**Fig.12.** Mouth changing rate of five sample words

---

### 5. Neural Network

The SOM Neural Network (SOMNN) algorithm is used to transform an input signal vector of arbitrary dimension into one- or two-dimension discretion maps which display the important statistical characteristics of the input vector. After an input vector is calculated by SOMNN, a best matching or winning neuron is found in the output map. It shows that similar input vectors activate the selected neuron and its neighbors simultaneously. This indicates that the similar characteristics of neurons will group together. The input vector, representing the set of input signals, is denoted by Eq. (13):

$$ x = [x_1, x_2, ..., x_p] $$  

There is a need to use the following formula, Eq. (14), to calculate the distance between the vector of the output layers and that of the input layers. The best-matching criterion is equivalent to the minimum Euclidean distance between vectors (Kohonen,1990).

$$ |x_p - w_{p}^{\text{wij}}| = [x_p - w_{p}^{\text{wij}}] [x_p - w_{p}^{\text{wij}}] = \sum_{p} [x_p - w_{p}^{\text{wij}}]^2 $$  

where

- $x_p$ : The input vector of the $p$th element;
- $w_{p}^{\text{wij}}$ : The weight vector between input and output layers;
- $||x_p - w_{p}^{\text{wij}}||$ : Euclidean distance between vectors.

### 5.1 The Topological Graph of the Neural Network

The output layer neuron displays in the output space with significant topological structure based on the features of input vector. The topological structure of output layers is allowed to respond to all distribution relations of input values. Therefore, this network is called a self-organizing characteristic mapping network.

This mapping graph is also called a topology. Each input vector would map to a coordinate of the topological graph. The relationship or similarity of any two input vectors could be calculated from the distance of two output coordinates in the topological graph. In the learning stage, input vectors of the similar character would gradually become closer. This means that similar input vector neighbors’ distance would decrease to a certain degree.

### 6. Experimental Results

In the experiment, hence used 62 words and 10 consecutive lip images per word as the database to test the implemented system performance by 55 users. Each image has 200 PCA data and 10 lip changing rate data. Hence, a word has 210 floating point data to represent all 10 image features.

The Self-Organizing Map is the neural network method to find the dynamic image variation of the lip shape changing. Table 2 showing the topological graph of 62 words in the database reveals the SOM topological graph whose coordinates were clustered together for the same word. Lip language recognition results are shown in Table 3. Table 3 shows the lip reading recognition rate of 62 tested Mandarin words with 55 users. The average correct rate is 85%.

### 7. Conclusion

In this study, the proposed system grabbed consecutive changing lip images and found the features of each image by using PCA algorithm and Neural Net technology. The PCA method could decrease the dimensions of raw data allowing the image data to be condensed without losing its major features. The Neural Net clustered similar image features together for reorganization. The reorganization rate of the implemented system highly depends on the lip shape. From the experimental results, the researchers discovered that the major problem in automatic lip reading is to recognize the lip shape sets when possessing the word with the similar pronunciation voice. This is the limitation of all lip reading recognition systems. However, combining the lip reading system to cluster those same-sound words could be a good supplementary tool for an optical character recognition system as a novel input device of a computer system. Additionally, only 62 words were used to analyze the lip reading recognition. The further study needs to enlarge the words-pool to make the research more comprehensive.
Table 2. The SOM topological graph of 62 words for each ten images

| data/image | 1. You (Ni) | 2. I (Wo) | 3. He (Ta) | 4. Own (Yu) | 5. Big (Da) | 6. Father (Pa) | 7. Mother (Ma) | 8. Ice (Bin) | 61. Mmm (N) | 62. Good (Hou) |
|------------|-------------|-----------|------------|-------------|-------------|---------------|---------------|-------------|-------------|---------------|
| 1          | (6, 9)      | (9, 3)    | (9, 3)     | (8, 2)      | (5, 8)      | (4, 6)        | (2, 4)        | (3, 3)      | (6, 3)      | (4, 0)        |
| 2          | (1, 9)      | (9, 2)    | (9, 2)     | (8, 1)      | (5, 7)      | (4, 4)        | (1, 2)        | (0, 6)      | (1, 3)      | (9, 0)        |
| 3          | (8, 9)      | (9, 2)    | (9, 2)     | (0, 3)      | (5, 7)      | (3, 4)        | (1, 2)        | (0, 7)      | (5, 2)      | (3, 0)        |
| 4          | (9, 9)      | (9, 0)    | (9, 0)     | (8, 0)      | (5, 5)      | (3, 8)        | (9, 4)        | (8, 6)      | (5, 4)      | (0, 1)        |
| 5          | (7, 9)      | (7, 0)    | (7, 0)     | (6, 0)      | (7, 5)      | (7, 8)        | (7, 0)        | (4, 4)      | (5, 4)      | (0, 1)        |
| 6          | (9, 9)      | (0, 4)    | (0, 4)     | (8, 1)      | (0, 9)      | (2, 4)        | (4, 4)        | (4, 0)      | (1, 5)      | (6, 2)        |
| 7          | (5, 8)      | (9, 6)    | (9, 6)     | (4, 5)      | (5, 2)      | (3, 6)        | (5, 7)        | (6, 0)      | (4, 3)      | (0, 0)        |
| 8          | (9, 7)      | (0, 7)    | (0, 7)     | (0, 6)      | (0, 3)      | (2, 6)        | (7, 6)        | (7, 2)      | (7, 5)      | (3, 1)        |
| 9          | (0, 9)      | (5, 6)    | (5, 6)     | (8, 5)      | (9, 2)      | (8, 9)        | (1, 6)        | (6, 5)      | (2, 2)      | (8, 0)        |
| 10         | (9, 8)      | (0, 3)    | (0, 3)     | (0, 2)      | (0, 8)      | (0, 3)        | (3, 5)        | (3, 3)      | (6, 8)      | (1, 0)        |

Table 3. The lip reading recognition rate of 62 tested Mandarin words for 55 users

| Word Result | 1. You (Ni) | 2. I (Wo) | 3. He (Ta) | 4. Own (Yeou) | 5. Big (Da) | 6. Father (Pa) | 7. Mother (Ma) | 8. Ice (Bin) | 9. Sweet (Tien) | 61. Mmm (N) | 62. Good (Hou) |
|-------------|-------------|-----------|------------|---------------|-------------|---------------|---------------|-------------|----------------|-------------|---------------|
| Tests       | 100         | 100       | 100        | 100           | 100         | 100           | 100           | 100         | 71             | 100         | 100           |
| Correct     | 82          | 98        | 90         | 92            | 93          | 89            | 90            | 92          | 78             | 71          | 91            |
| Wrong       | 18          | 2         | 10         | 8             | 7           | 11            | 10            | 8           | 22             | 29          | 9             |
| Recognition rate | 82%          | 98%        | 90%        | 92%           | 93%         | 89%           | 90%           | 92%         | 78%            | 71%         | 91%           |

8. Acknowledgement

The author would like to show appreciation that this work was supported by the National Science Council of Taiwan, ROC under Grant No. NSC-96-2221-E-167-026-MY3 and NSC 99-2410-H-167-005-MY2.

9. References

1. Bauman N (2003) Speechreading. http://www. hearingloss-help.com/articles/ speechreading.htm.
2. Chai D and Bouzerdoum A (2000) A Bayesian approach to skin color classification in YCbCr color space, TENCON 2000, Proceedings, pp.421-424.
3. Chiang C C, Tai W K, Yang M T, Huang Y T and Huang C J (2003) A novel method for detecting lips, eyes and faces in real time, Real-Time Imaging, Vol.9, pp.277-287.
4. Gonzalez R C and Woods R E (2002) Digital Image Processing, *Prentice Hall.*

5. Hsu R L (2002) Abdel-Mottaleb. M, Jain. A.K., Face detection in color images, Pattern Analysis and Machine Intelligence, *IEEE Transactions on*, Vol.5, No.24, pp.696-706.

6. Kohonen T (1990) The Self-organizing map, *Proc. IEEE*, Vol.9, No.78, pp.1481-1480.

7. Lin C S, Tseng HW, Lay Y L, Chen CT and Yeh M S (2012) Design and reliability analysis of image type head gesture interface in distributed control system, *Indian Journal of Science and Technology*, Vol.5, No.8, pp.3150-3156.

8. Lin C S, Yang H J, Lay Y L and Yang S W (2011) Design and evaluation of a public installation of an eye-gaze system, *Assistive Technology*, Vol.23, pp.187-198.

9. Lay Y L, Yang H J, Lin CS, Chen WY (2012) 3D Face Recognition by Shadow Moiré, *Optics & Laser Technology*, Vol.44, No.1, pp.148-152.

10. Yang J, Fu Z, Tan T and Hu W (2004) Skin color detection using multiple cues, *Pattern Recognition*, ICPR 2004. Proceedings of the 17th International Conference on, pp.632-635.

11. Kong W and Zhu S (2006) A New Method of Single Color Face Detection Based on Skin Model and Gaussian Distribution, *Proceedings of the 6th World Congress on Intelligent Control and Automation*.

12. Shaw PJA (2003) Multivariate statistics for the Environmental Sciences, *Hodder-Arnold*.

13. Yao W J, Liang Y L and Du M H (2010) A real-time lip localization and tacking for lip reading, *Advanced Computer Theory and Engineering*, Vol.6, pp.363-366.