Fifth IFSA World Congress (1993), 903-906

Lipreading using The Fuzzy Degree of Similiarity

Kenji KUROSU*, Tadayoshi FURUYA**, Shigeru TAKEUCHI*, and Mitsuru SOEDA**

* Kyushu Institute of Technology
Kawazu, Iizuka-shi, Fukuoka-ken, 820, JAPAN

** Kitakyushu College of Technology
5-20-1, Shii, Kokura-minami-ku, Kitakyushu, 803, JAPAN

Abstract

Lipreading through visual processing techniques help provide some useful systems for the hearing impaired to learn communication assistance. This paper proposes a method to understand spoken words by using visual images taken by a camera with a video-digitizer. The image is processed to obtain the contours of lip, which is approximated into a hexagon. The pattern lists, consisting of lengths and angles of hexagon, are compared and computed to get the fuzzy similarity between two lists. By similarity matching, the mouth shape is recognized as the one which has the pronounced voice. Some experiments, exemplified by recognition of the Japanese vowels, are given to show feasibilities of this method.

1. Introduction

Lipreading is one of communication forms to understand a speaker without hearing; rather, with watching movements of lips, jaws and facial expressions. But it takes a long time to acquire the ability of lipreading. One of reasons for difficulties of learning lipreading is that the Japanese language has more than one hundred syllables with merely about fifteen mouth shapes to be distinguished easily[1]. These special features in Japanese lead to the reality, where the same mouth shape exists for too many pronounced syllables and the listeners are required to have high-level techniques of understanding. Many papers are reported for lipreading, some of which aim to realize systems to understand spoken words by recognizing visual images of lip movements[2],[3]. Some of researchers developed the systems to help people to learn lipreading or teach it[4]. Also some papers are reported to construct supplement systems for acoustic speech recognition[5]. Besides their purposes, there are many methods employed for recognition and visual data processing to realize lipreading, all of that stimulate engineers and researchers into studying varieties of implementation methods toward more feasible lipreading[6]. We proposed lipreading methods by visual image processing with pattern classification[7], by Fuzzy logic[8], by using an X-Y tracker[9], and by using a neural network[10]. Though many aspects and problems were discussed, none of their works are easy to be implemented because of speed, memory capacities, unrealistic settings, or poor recognition rates. Therefore, the lipreading must be investigated to produce more feasible and simpler systems. We have proposed a method[11] related to pattern recognition, better performance in the frame and shape recognition by a fuzzy similarity matching. This paper reports a method, combined with pentagon approximation and the fuzzy similarity, which intends to be technically realizable and effective. Many examples of vowel recognition through mouth shape are presented to show better performance.

2. Image processing

2.1 Images of mouth shapes

Some pictures taken by a full color video-digitizer are processed to obtain the mouth shapes of a subject in the system developed as shown in Fig.1. The subject, fixed in front of a camera, is illuminated by a white lamp, where his lips are colored by a lip stick. The camera gets the front views of his face, and they are digitized into a personal computer. Memorized video data 640 × 400 pixels are, first, compressed into 160 × 100 pixels, all of which are filtered to remove noise and excessive information by averaging the 3 by 3 surrounding points. After obtaining 160 × 100 pixels filtered data with 256 brightness levels, nonlinear transformation is applied to get binary black and white data. A nonlinear adaptive threshold method, whose algorithm determines the optimal threshold and binary pixels automatically, is used to obtain binary data to distinguish the lip shapes against the other backgrounds. After binary transformation, the contours of lips or mouth are estimated by the following procedures.

(1) From the contours of the image, some closed contours are left if the trace of the contour comes back to the original starting points.

(2) If the area of the closed contour is too small, the closed contour is erased.

(3) In the remained contours, the contour placed at the bottom is estimated as the one for mouth shapes.

By this algorithm illustrated in Fig.2, the mouth shape can be obtained.

2.2 Hexagon approximation

The characteristic values of the mouth shape is necessary to distinguish mouth shapes depending on the pronounced voices. From the contours of the mouth shape, 6 characteristic points are searched by an algorithm. The 6 points, using highest, lowest, edges, left, right, etc. as shown in Fig.3, are determined to approximate the mouth shapes as a hexagon. The characteristic pattern lists for the hexagon are represented as

$$Sh: (h1, h2, h3, h4, h5, h6)$$
$$Sk: (k1, k2, k3, k4, k5, k6) \qquad (1)$$
$$Sty: (ty)$$

where the list Sh has elements of 6 side lengths of the hexagon, the list Sk has 6 angles, and Sty's element is the ratio between vertical and horizontal lengths. The lists of these characteristic values of the hexagon are compared and computed to get similarity values between them.

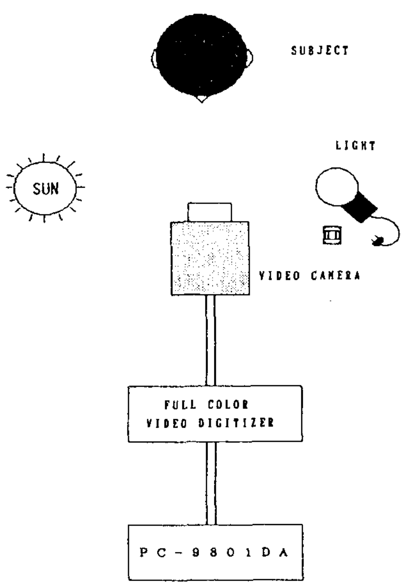

Fig.1 Experimental apparatus

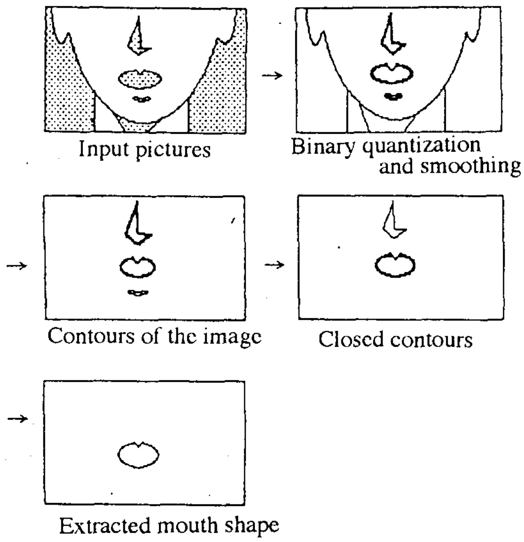

Fig.2 Extraction of a mouth shape

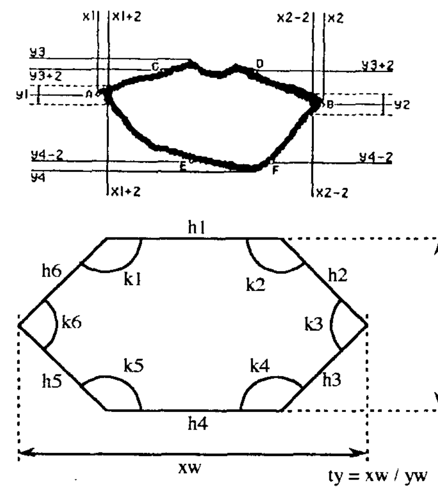

Fig.3 Hexagonal approximation

3. The method for vowel recognition

3.1 Distinction of the mouth shape

To distinguish vowels by using symbolic characteristic lists of mouth shape approximated with a hexagon, the following pattern recognition method is applied. The basic procedure of the pattern recognition consists of five processes as shown in Fig.4. In this figure, the most essential and important part is the process to get the characteristic lists and to evaluate the similarity degree between the standard pattern and extracted characteristic pattern.

Using the list of these lengths and angles of the hexagon defined in 2.2, the similarity between two patterns, standard pattern and the pattern of mouth shapes detected by a video camera, is evaluated. In this case, it is necessary to give consideration to the errors in observing mouth shape and approximating by a hexagon. So we use a fuzzy degree of similarity based on the fuzzy theory idea to distinguish vowels more exactly.

3.2 The fuzzy degree of similarity

(1) The degree of similarity between two numbers

In order to evaluate the degree of similarity between two characteristic lists, it is necessary to obtain the degree of each element of the lists. In this research, it is assumed that all the elements of the characteristic list are fuzzy numbers. The variable A is fuzzy number that has a triangular membership function $\mu a(x)$ as shown in Fig.5, and the length of the base of this membership triangle is equal to an absolute value of A. Assuming the all numbers are fuzzy numbers as the variable A, the similarity degree S(A,B) between two numbers, A and B, is defined as follows.

$$S(A, B) = \mathrm{Max}\{\mathrm{Min}\{\mu a(x), \mu b(x)\}\} \quad (2)$$

Here μa and μb are the membership function of A and B respectively. Fig.5 shows S(A, B). In this figure, the distance S is the degree of similarity between two numbers, where $0 \leq S \leq 1$.

(2) Comparison of the characteristic lists

Using the degree of similarity between two fuzzy numbers, similarity between two characteristic lists is evaluated simply by comparing elements of the lists. Let X_n and Y_n represent the n-th element of the two lists, denoted as $(X_1, X_2, \dots, X_n)$ and $(Y_1, Y_2, \dots, Y_n)$ respectively; Let S_i denote the degree of similarity between X_i and Y_i. S_i can be calculated easily using equation (2). Then the degree of similarity between two lists is defined as follows.

$$S_L = \underset{1 \leq i \leq n}{\mathrm{Min}}\{S_i\} \quad (3)$$

(3) Similarity degree of the characteristic lists

The mouth shape is approximated with a hexagon, so each of the shape patterns is symbolized as the line length and turning angular degree that is represented as X_i and A_i, where X_i is the length of i-th line and A_i is the turning angle from the i-th line to (i+1)-th line. If the polygon has n edges, the shape patterns can be determined by using

$(X_i, A_i) \quad i=1,2,\dots,n$

Any polygon, which consists of n lines and n angles, can be expressed as 2n symbolic sequence such as

$$(X_1, A_1, X_2, A_2, \dots, X_n, A_n) \quad (4)$$

These sequences are the characteristic lists of the mouth shape. Let $L_1=(X_1, A_1, X_2, A_2, \dots, X_n, A_n)$ denote a characteristic list; Let $L_2=(Y_1, B_1, Y_2, B_2, \dots, Y_n, B_n)$ denote the other characteristic list. Because the sequences $(X_i, A_i, \dots, X_n, A_n, X_1, A_1, \dots, X_{i-1}, A_{i-1})$ represent the same shape as (4), whose position is changed.

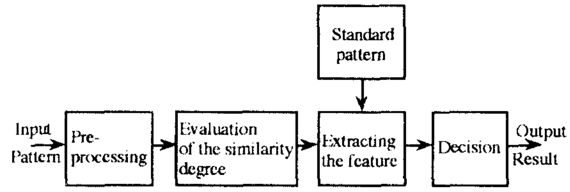

Fig.4 The process of the pattern recognition

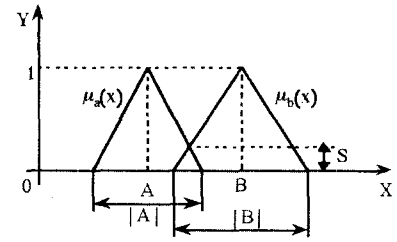

Fig.5 The similarity degree of two fuzzy numbers

And so, when we compare a characteristic list L_1 with the other list L_2, we must consider n lists concerning L_1 as follows;

L_1: $L_{11}=(X_1, A_1, X_2, A_2, \dots, X_n, A_n)$

$L_{12}=(X_2, A_2, X_3, A_3, \dots, X_n, A_n, X_1, A_1)$

...

$L_{1n}=(X_n, A_n, X_1, A_1, \dots, X_{n-1}, A_{n-1})$

The degree of similarity between the list L_{1k} and L_2 is denoted as S_{Lk} (k=1,2, ... ,n), which is calculated using equation (3). Then the similarity degree of two characteristic patterns, L_1 and L_2, of mouth shape is defined as following equation,

$$S_F = \underset{1 \leq k \leq n}{\mathrm{Max}}\{S_{Lk}\} \quad (5)$$

3.3 The method for recognizing the vowels

As stated in the previous sections, the shape is symbolized as the side lengths and inner angles of the hexagon to make the characteristic lists. Then two kinds of the characteristic lists - one is the list whose components are the lengths and the angles and the other is the list whose components are the ratios of the width to the height of the mouth - are used to calculate the fuzzy degree of similarity between the mouth patterns and the standard pattern to recognize the spoken vowels.

The procedures for discrimination of vowels are as follows. The extracted pattern is compared to the standard pattern to recognize the spoken vowels by using the fuzzy similarity degree. The degree of similarity of the ratio of length and width is calculated using the equation (2). Because the mouth shape is approximated by a hexagon, There are 6 kinds of the characteristic lists concerning the side lengths and the angles as in equation (4). So the degree of the similarity of the lengths and angles is obtained by using equation (5). In each case, 2 points are given in the pattern which value of the similarity degree is the largest, and 1 point is given in the pattern which similarity degree has the second value. Finally, the pattern, whose total of points is the largest, is selected and regarded as the pattern of the spoken vowel.

Thus the spoken vowels are determined from the mouth shape.

4. Experimental results

The images of a subject, who pronounced the Japanese 5 vowels, A, I, U, E and O, were taken into a personal computer. The four sets of samples taken 4 times from the same subject were processed into the characteristic pattern lists. The same experiment was carried out for another subject to get one set of pattern. An example of the pattern computation is shown in Fig.6. The standard pattern for 5 vowels were computed by averaging 3 sets of the values chosen from the subject. With these 5 sets of vowel patterns, recognition by matching was experimented as follows.

Experiment 1: Recognition between the standard pattern and the 3 patterns, which are used for constructing the standard pattern. Experiment 2: Recognition between the standard pattern and the fourth pattern, which is exclussive set to form the standard pattern. Experiment 3: Recognition between the standard pattern and the patterns of another subject.

The results are shown in Table1, where O corresponds to the correct recognition and × corresponds to wrong recognition.

5. Conclusion

It shown that the spoken Japanese vowels can be recognized by the proposed lipreading meth-od, where the mouth shapes are approximated as hexagons, and the fuzzy similarity method is applied to distinguish the mouth shapes. The results reveal that the perfect recognition is possible when the standard pattern are composes of the same person's mouth shapes. To improve recognition rates in many cases, construction of standard patterns must be investigated more theoretically. In the future, it is necessary to improve the method as performing lipreading faster for extraction of mouth shapes and pattern matching.

References

1)Fukuda,Y., S.Hiki: Characteristics of the mouth shape in the production of Japanese stroboscopic observation, J.Acoust.Soc.Japan, (E), 3, 2, pp.75-91(1982)
2)Mase, K., A.Pentland: Automatic lipreading by optical flow analysis, Trans. Inst. Electron. Inf. Commun. Eng., D-II, J73D-II, 6, pp.796-803(1990)
3)Okada, K., C.Ohira, H.Nakamura: A method for lipreading, Trans. Inst. Electron. Inf. Commun. Eng., D-II, J72D-II, 9, pp.1532-1583(1989)
4)Gulliiams,I., A.Segui: Interactive videodisc for teaching and evaluating lipreading, Proceedings of An. Intern. Conf. of IEEE Eng. in Medicine and Biology Soc., Cat.88Ch2566-8, pp.1551-1552, 4(1988)
5)Pentajan, E.D.: Automatic lipreading to enhance speech recognition, Proceedings of CVPR'85, Cat.85Ch2145-1, pp.40-47(1985)
6)Watanabe,T.: Machine lip reading of two concatenated vowels, Trans. JPN. Soc. Mech. Eng. (C), 55-509, pp.134-138(1989)
7)Matsuoka,K., T.Furuya, K.Kurosu: Speech Recognition by image processing of lip movements, Trans. of SICE, 22, 2, pp.67-74(1986)
8)Kurosu,K., T.Furuya, S.Takeuchi: Lip reading by fuzzy logic, Proc. of 2nd Congress of IFSA(1987)
9)Furuya,T., M.Soeda, K.Kurosu, S.Tamura: Speech recognition with lip movement data using an X-Y tracker, Trans. of SICE, 27, 8, pp.958-965(1991)
10)Kurosu,K., T.Furuya, etal.: Lipreading by a neural net, Proc. of 8th European Annual Conf. on Human Decision Making and Manual Control, pp.329-339 (1989)
11)Weijing, Z., T.Furuya, K.Kurosu: A list processing method on Composition and decomposition of figure patterns,Tran. of SICE, 26, 1, pp.81-86(1990)

Table1 The recognition results

Input	Experiment 1	Experiment 2	Experiment 3
A	O	O	× O
I	O	O	O
U	O	O	O
E	O	O	O
O	O	O	× U

O Success × Fail

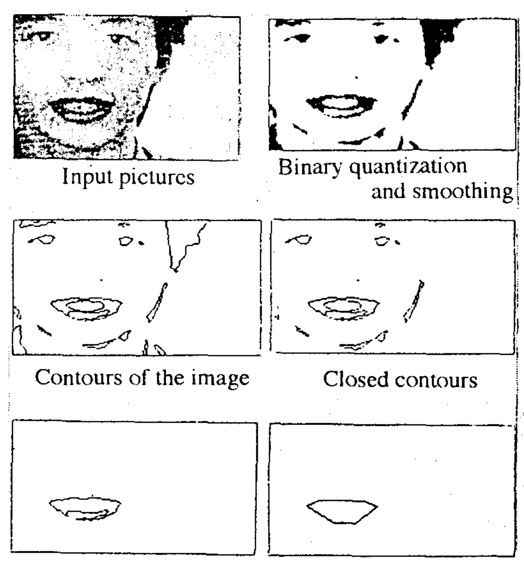

Fig.6 Example of the pattern computation