Estimating the Face Direction for the Human Interface

Takeshi SHIGEMURA* Makoto MURAYAMA*
Hitoshi HONGO** Kunihito KATO* Kazuhiko YAMAMOTO*

*Department of Information Science,
Faculty of Engineering, Gifu University
1-1 Yanagido, Gifu 501-1193, JAPAN,
takeshi@yam.info.gifu-u.ac.jp
TEL&FAX:+81-58-293-2753

**Hypermedia Research Center,
SANYO Electric Co., Ltd.
180 Ohmori, Anpachi-gun, Anpachi-cho,
Gifu 503-0195, JAPAN
hongo@gf.hm.rd.sanyo.co.jp
TEL:+81-584-64-6594
FAX:+81-584-64-3943

Abstract

To improve natural conversation between human and computer, we aim at applying face direction detection to human interface. In this paper, we propose a face direction detection method using higher order local autocorrelation features. Higher order local autocorrelation feature is shift-invariant. First, the Prewitt operator generates an edge image from an original image. Next, the higher order local autocorrelation features are extracted from the binarized edge image. Before estimating the face direction, the dictionaries using face direction detection are organized by the images of various angle faces. We experimented with our method using the face directional dictionaries. As a result, we could estimate the face direction roughly.

1. Introduction

Recently, the importance of man-machine interface between humans and computers is rising. This research field is particularly focus on user-friendliness and a comfort. There is a lot of research in recognition face, expression recognition, face direction recognition and so on. These methods have been applied to various applications of human interface. We assume that the detection of human's attention direction is useful for human interface. For example, a window can be moved forward automatically by a system that detects the user's attention. Generally speaking, a human turns his/her face his/her direction of attention for acquire information. Thus, we detect the face direction instead of the direction of the attention when humans look at objects. Natural communication between humans and computers is possible with this method. If it is possible to estimate face, direction, it could apply to human interfaces. For example, we can move camera where the human looks at, and acquire the information what he/she is looking at.

There are many methods of detecting face direction. The estimated face direction from the center of gravity of skin color and hair region was proposed[1]. But this method was not effective when the hair region had asymmetry. It was proposed that an input face be compared with a face model for parts extraction[2], but real-time processing might be difficult.

In this paper, we propose a method that estimates face direction from a camera image using higher order local autocorrelation features. We extract higher order local autocorrelation features from a facial image, and make various angle dictionaries for linear discriminant analysis. Then our algorithm estimates the face direction by comparing with each angle dictionary.

2. Estimated method of the face direction

We use the discriminant methods[3][4] to estimate the face direction using higher order local autocorrelation features that can recognize variable objects. First, our method learns features in various facial image directions, and then estimates face direction of the input facial image. It can acquire features quickly because of a mask pattern. The discrimination of recognition object category also used linear discriminant analysis. This method can compress the number of features, and can quickly process at the recognition phase.

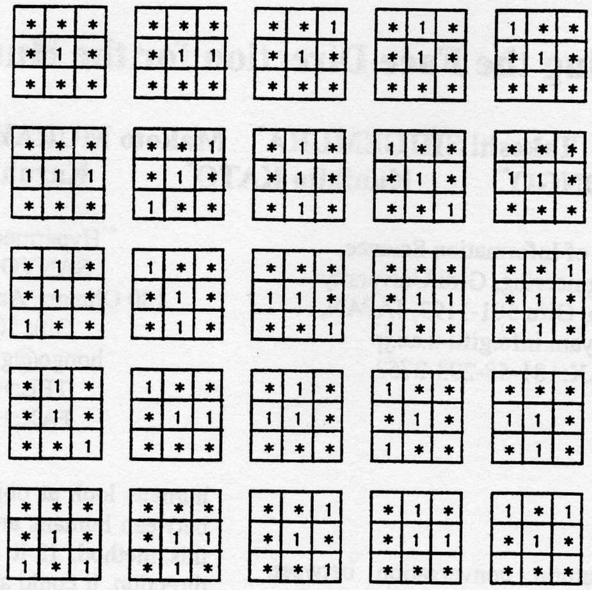


Figure 1: Local mask patterns for the higher order local autocorrelation features.

2.1 Higher order local autocorrelation features

It is well known that the autocorrelation function is shift-invariant. Higher order local autocorrelation functions is extended to a higher order. Let a region in an image be denoted by P. Then the N th-order autocorrelation functions with N displacements $a_1, \dots, a_N$ is define by

$$x_f^N(a_1,\dots,a_N) = \int_P f(r)f(r+a_1)\cdots f(r+a_N)dr, \quad (1)$$

where $f(r)$ denotes gray-level at the position r.

Since the number of these autocorrelation functions obtained by the combination of the displacements over the region P is enormous, we must reduce them for practical application. First, we restrict the order N up to two (N=0,1,2). It is seen that the 0th-order autocorrelation is just the averaged gray-level of the region P.

We also restrict the range of displacements to within a local 3×3 region, the center of which is the reference point. By eliminating the displacements that are equivalent to the shift in scanning, the number of the patterns of displacements is reduced to 25. Figure 1 shows the patterns, where the symbol "*" represents "don't care".

Thus the higher order local autocorrelation features are obtained by scanning the image over P with the 25 local 3×3 masks and by computing the sums of products of the corresponding pixels.

The features extracted from the image of the highest resolution includes only very local and detail information. We suppose that global information is often useful for the face direction estimation.

The extracted features from images with several resolutions are utilized. In this paper, we use 75 features with three types of resolution, such as, original, half, and quarter[5]. We guess that available features contain low frequency ingredients to estimate the face direction. A set of the higher order local autocorrelation features extracted from each of the image in the pyramidal structure includes both detailed and global information on facial image. Thus our method can extract useful primitive features the whole.

2.2 Linear discriminant analysis

Linear discriminant analysis is that the optimal coefficient matrix is given by solving the eigenvalue problem. The optimal coefficient matrix generates the best new features from the primitive features, that is higher order local autocorrelation features. Let the number of classes be denoted by K. The dimension of discriminant space N is bounded by min(K–1, M), where M is the dimension of higher order local

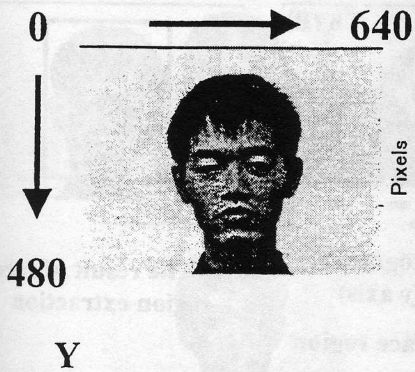
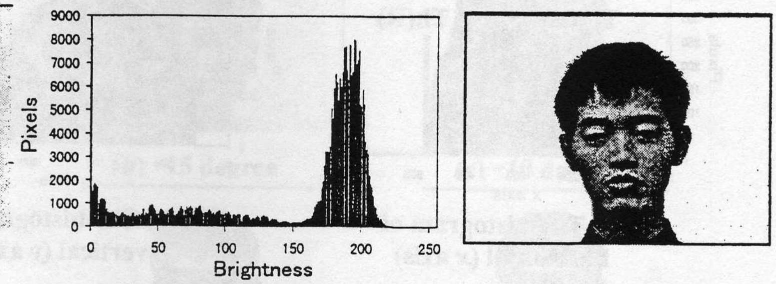


Figure 2: Original image



(a) Brightness histogram

(b) The image removed background

Figure 3: The removed background processing

autocorrelation features. The dimension M is 75. To discriminate unknown images, we extracted the higher order local autocorrelation from the images. The new features are given by linear mapping the primitive features using the coefficient matrix. The estimation result determines the class whose average vector is the closest to the new features.

3. Feature extraction method

It is necessary to extract features of the face region to estimate face direction from an input image correctly. Because we had the better results in the other face recognition experiments based on the higher order local autocorrelation features[6], we clipped the region which includes face from the original image. We extracted the face region using the brightness histogram.

3.1 Face region extraction

The first, the brightness histogram is made from a gray-scale original image, 640×480 pixels, with the background removed.

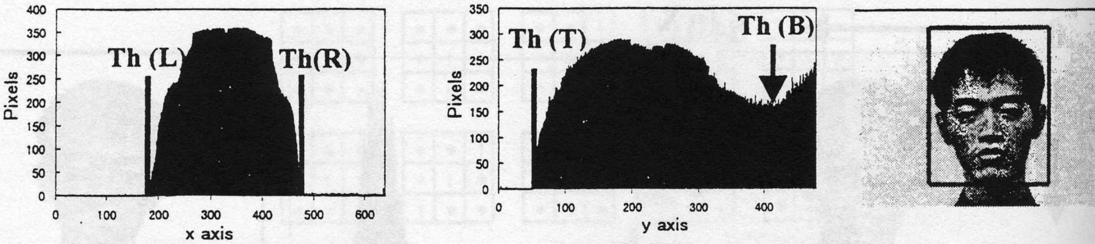
While the background region is plain and larger than the face region, we assume that the background has the most pixels in the brightness histogram. We make the brightness histograms and remove the background. The background is the concentration value that has the most pixel value in the brightness histogram. It is removed by eliminating the peak concentration value. Figure 2 shows the original image. Figure 3 shows the brightness histogram and the image of the background removed. Next, we make a projection histogram of each axis, horizontal(x) and vertical(y). The face region is determined by threshold processing. Figure 4 shows the projection histogram of the horizontal and vertical axes, and the extracted face region. The processing of neck is performed by scanning to right from the maximum pixel point in the Figure **4(b)**, and by searching the minimum pixel point. The size of face region varies. We calculated the size of the nine face images looking in different direction. We confirmed that there was little different size in size of image.

3.2 Features extraction algorithm

The features for estimating face direction are extracted from the face region given **3.1.** The features extraction algorithm is as follows:

1) Creating the different resolution images such as 1/2 and 1/4 from the original face region.
2) Making the edge images by Prewitt operator.
3) Binarizing above images by threshold processing.
4) Extracting the higher order local autocorrelation features from three images.

The dictionary is organized from the higher order local autocorrelation features to estimate face direction using linear discriminant analysis. The features are also extracted from the unknown image.



(a) The histogram of horizontal (x axis)

(b) The histogram of vertical (y axis)

(c) The result of face region extraction

Figure 4: The clipped face region

4. Preparatory experiment

4.1 The purpose of the experiment and the training data

We performed a preparatory experiment to confirm the credibility of the training data. The training data makes the dictionary to estimate the face direction. We investigated how dispersion influences the estimation of face direction for the wrong training data.

Figure 5 shows examples of the training data. We recorded the face in a room with fluorescent lights. The distance between the camera and the subject should be about two meters. We had a subject face the indexes. We prepared 60 facial images for training data in each direction. In this experiment, nine directions including - 20, - 15, - 10, - 5, 0, 5, 10, 15, and 20 degrees were selected.

4.2 Experimental method and result

First, we used the face image captured using a video camera as described above. In the sixty face angle images, the thirty images are used for the training data, and the others are used for the unknown data. The distance between the subject and the camera was constant. Table 1 shows the experimental result of the extracted face region and the whole region. We decide that the answer was the class of dictionary that was calculated minimum distance from the input image.

For the result, we could confirm that the recognition rate of the case of the extracted face region is the higher than the whole region. We used the extracted face region. Figure 6 shows the average and dispersion of the thirty unknown data image. The horizontal axis is the training face angles, the vertical axis is the distance between the input data and the classes. The results indicate that the plus-minus 20 degrees have a larger difference of distance than the other classes. Although the others are small, it could recognize the correct class because of the small class dispersion. We confirmed the reliability of the training data for every five angle. We also confirmed that the movement of the face was continuous.

Table 1: The change of the recognition rate

The feature extraction region	Original image (%)	Face region (%)
Training data	100.0	100.0
Unknown data	78.9	84.9

5. Estimation experiment of face direction

5.1 Experimental method

We experimented with the recognition method by using the training data which are confirmed reliability by the preparatory experiment of section 4.

The training data is the same as the image used in the preparatory experiment. The total number of images in the training data is 270 consisting of thirty images in nine directions. The experimental data was taken by rotating the body of a subject sitting on a chair from −20 degrees to 20 degrees horizontally. The number of images in the experimental data was 180.

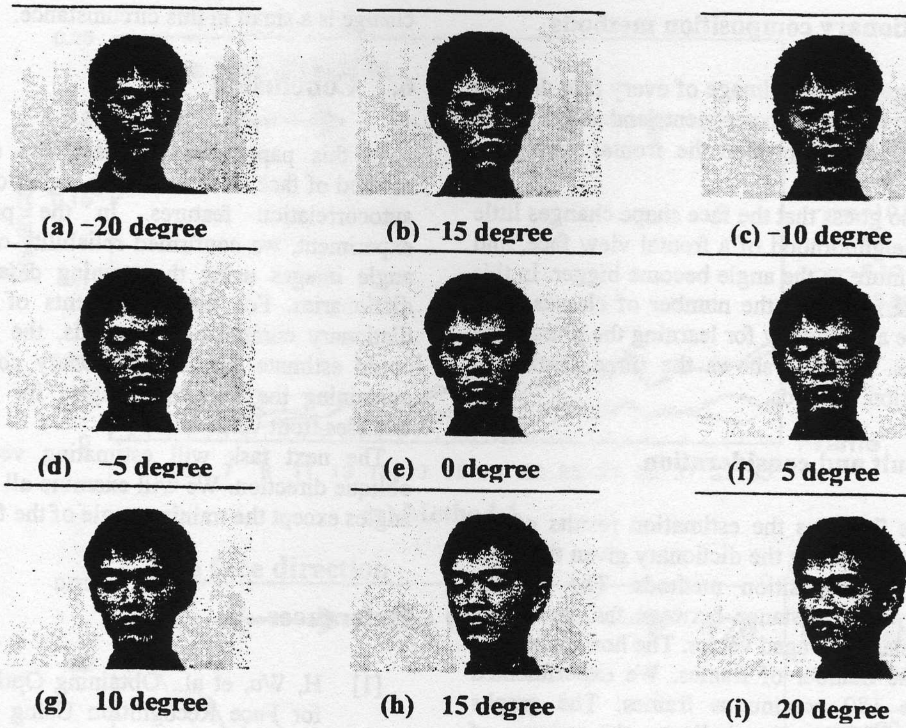


Figure 5 : Examples of the training data

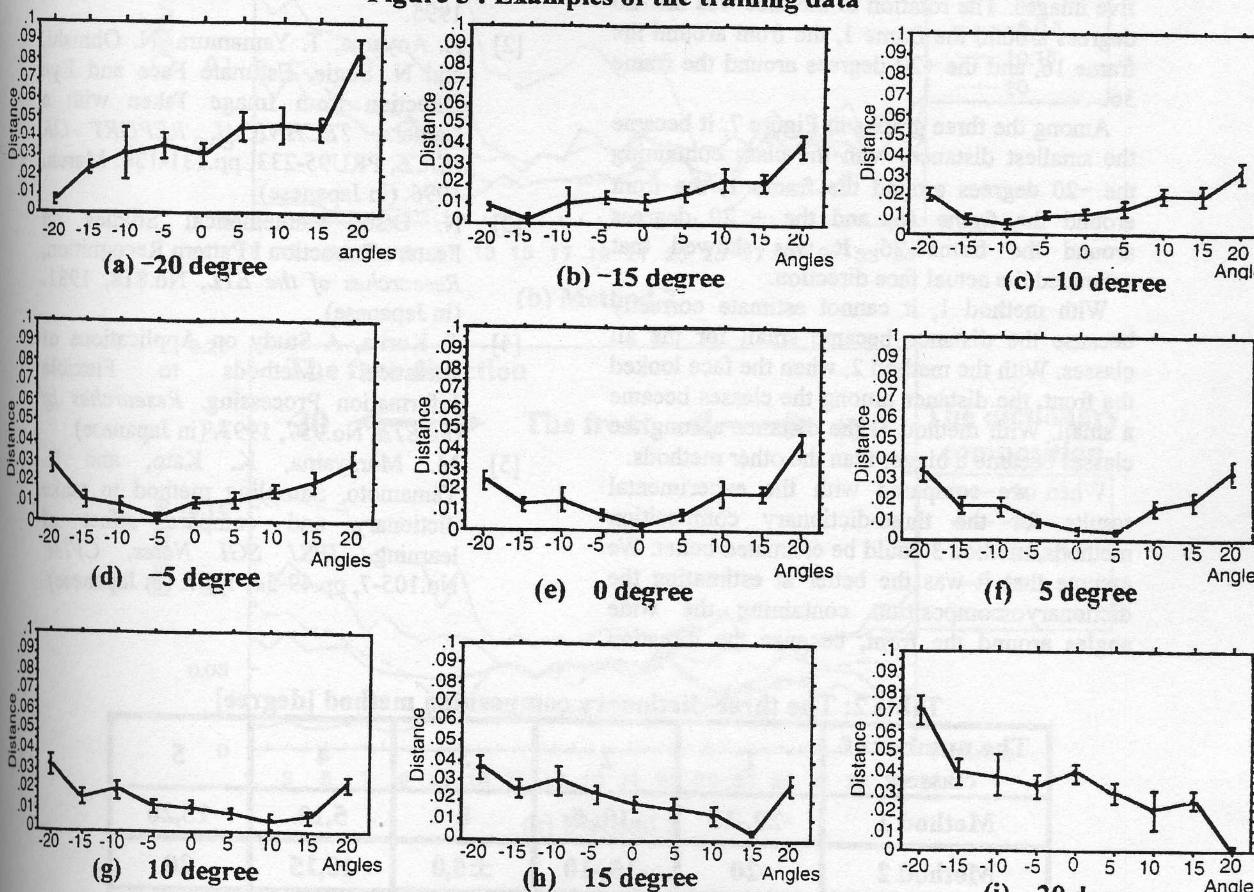


Figure 6 : The average and dispersion of the distance each angle for the preparatory experiment

5.2 Dictionary composition methods

It learned the face image of every five degrees for the preparatory experiment, and didn't make much difference between the frontal view face and ±5 degrees.

We also guess that the face shape changes little for the neighborhood of a frontal view face, and changes more as the angle become bigger. In this paper, we restricted the number of classes to 5, and made a dictionary for learning the plural face directions. Table 2 shows the three dictionary composition methods

5.3 Result and consideration

Figure 7 shows the estimation results of the face direction using the dictionary given the three dictionary composition methods. The vertical axis shows the distance between the input data and the class averaged vector. The horizontal axis shows the number of frames. We experimented with the 180 continuous frames. The graphs show the distance that indicates the average of five images. The rotation of the face was the -20 degrees around the frame 1, the front around the frame 16, and the +20 degrees around the frame 36.

Among the three graphs in Figure 7, it became the smallest distance, with the class containing the −20 degrees around the frame 1, the front around the frame 16, and the + 20 degrees around the frame 26. It was showed that estimated the actual face direction.

With method 1, it cannot estimate correctly because the distance became small for the all classes. With the method 2, when the face looked the front, the distance among the classes became a small. With method 3, the distance among the classes became a bigger than the other methods.

When we compared with the experimental results for the three-dictionary composition methods, method 3 could be estimated better. We assume that it was the better at estimating the dictionary composition containing the wide angles around the front, because the direction change is a small in this circumstance.

6. Conclusion

In this paper, we proposed the estimation method of face direction using higher order local autocorrelation features. In the preparatory experiment, we confirmed reliability of the face angle images using the training data to make dictionaries. For the experiments of the three dictionary composition methods, the method 3 could estimate the best dictionary composition containing the expanded angles for dictionary includes front view face.

The next task will estimating vertical and oblique direction. We will examine all other face angles except the training angle of the face.

References

[1] H, Wu, et al., Obtaining Optical Image for Face Recognition Using an Active Camera, *ACCV '95*, pp. III-708-712, 1995.

[2] K. Aoyama, T. Yamamura, N. Ohnishi, and N. Sugie, Estimate Face and Eye Direction from Image Taken with a Camera, *TECHNICAL REPORT OF IEICE*, PRU95-233, pp.131-136, March, 1996. (in Japanese)

[3] N. Otsu, Mathematical Studies on Feature Extraction I Pattern Recognition, *Researches of the ETL*, No.818, 1981. (in Japanese)

[4] T. Kurita, A Study on Applications of Statistical Methods to Flexible Information Processing, *Researches of the ETL*, No.957, 1993. (in Japanese)

[5] M. Murayama, K. Kato, and K. Yamamoto, Sampling method to make dictionary and compare effect of learning, *IPSJ SGI Notes, CVIM*, No.105-7, pp.49-54, 1997. (in Japanese)

Table 2: The three-dictionary composition method [degree]

The number of classes	1	2	3	4	5
Method 1	-20,-15	-10,-5	0	5,10	15,20
Method 2	-20	-15,-10	±5,0	10,15	20
Method 3	-20	-15	±10,0	15	20

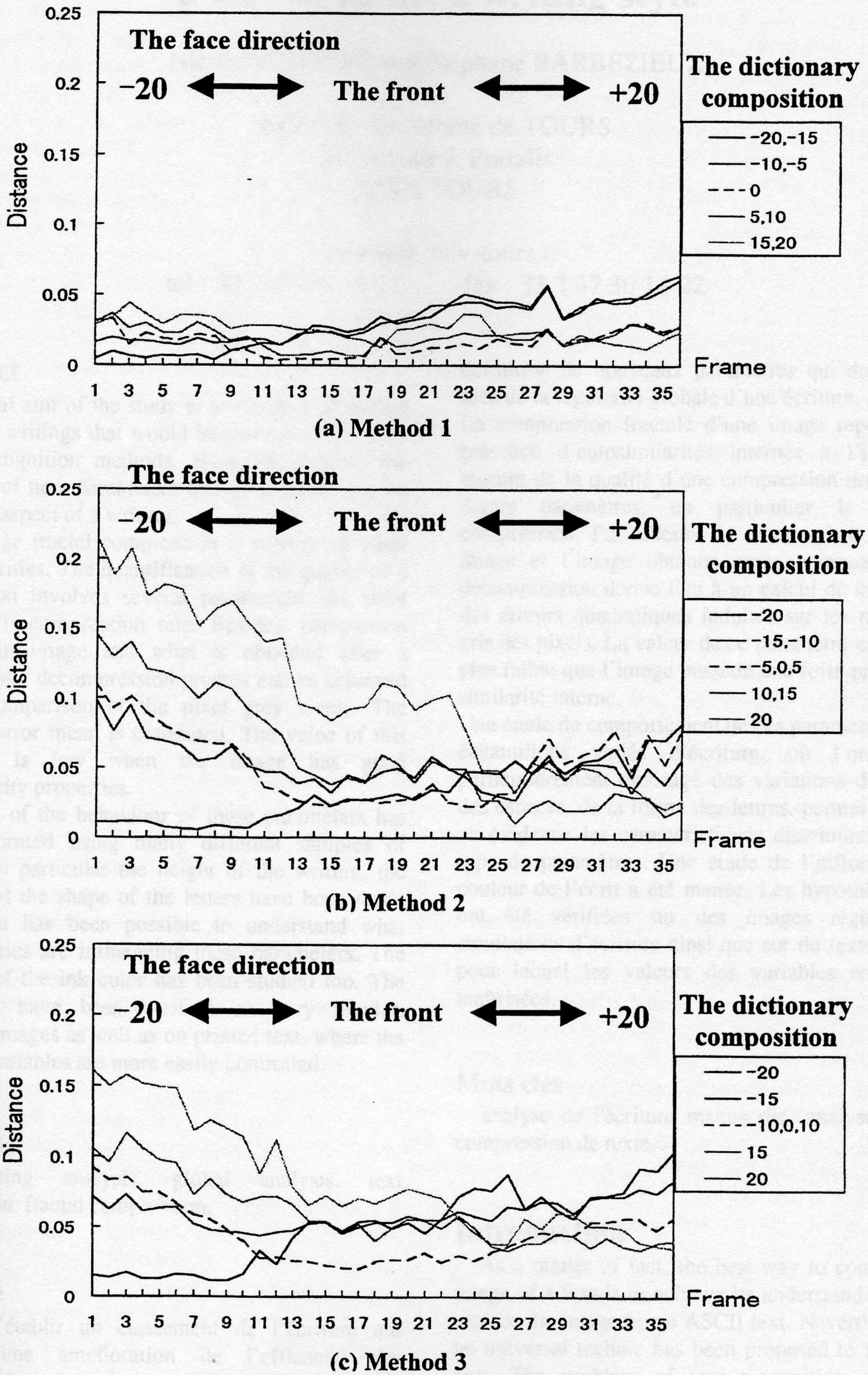


Figure 7: The distance between the each class and input image for the three dictionary composition