Proceedings of the Federated Conference on
Computer Science and Information Systems pp. 869–874

DOI: 10.15439/2015F398
ACSIS, Vol. 5

Minimum Variance Method to Obtain the Best Shot in Video for Face Recognition

Kazuo Ohzeki
Graduate School of Engineering and Science, Shibaura Institute of Technology 3-7-5 Toyosu, Koutou-ku, Tokyo
Email: ohzeki@shibaura-it.ac.jp

Ryota Aoyama
Graduate School of Engineering and Science, Shibaura Institute of Technology 3-7-5 Toyosu, Koutou-ku, Tokyo
Email: ma15001@shibaura-it.ac.jp

Yutaka Hirakawa
Graduate School of Engineering and Science, Shibaura Institute of Technology 3-7-5 Toyosu, Koutou-ku, Tokyo
Email: hirakawa@shibaura-it.ac.jp

***Abstract*—This paper describes a face recognition algorithm using feature points of face parts, which is classified as a feature-based method. As recognition performance depends on the combination of adopted feature points, we utilize all reliable feature points effectively. From moving video input, well-conditioned face images with a frontal direction and without facial expression are extracted. To select such well-conditioned images, an iteratively minimizing variance method is used with variable input face images. This iteration drastically brings convergence to the minimum variance of 1 for a quarter to an eighth of all data, which means 3.75-7.5 Hz by frequency on average. Also, the maximum interval, which is the worst case, between the two values with minimum deviation is about 0.8 seconds for the tested feature point sample.**

I. Introduction

THERE are two major methods for face recognition [1]; one is the feature-based method which uses feature points at the endpoints of facial parts. The other is the holistic method, which processes the whole face region without decomposing regions into feature points. The former method has become unpopular because it is difficult to detect accurate feature points automatically. The latter method is now popular. However, Kathryn Bonnen and Anil K. Jain have recently proposed the effectiveness of a component-based method which uses facial parts in detail, rather than globally recognizing face information [2]. The holistic method conventionally analyzes the whole face region at once to achieve robust recognition. The component-based method utilizes separate regions of the eyebrows, eyes, nose and mouth, and performs dedicated recognition for each separate region, then integrates the results. The component-based method implies that it is now possible to detect facial parts before face recognition. The performance of the component-based method is better than that of the single holistic face recognition method.

The performance of face recognition has improved recently [3], and the results of various contests have been reported [4-6]. In those reports, recognition of frontal faces was considered as an easy task and more complicated conditions with age changes and with expressions are now targeted. Then, the basic face recognition technology is left in popular development activity. The recognition rate is now 97.5% for the frontal face without expressions [7], and it is still a difficult problem to realize 100% recognition for the frontal face without expressions.

One of the disadvantages of feature-based methods using feature points of face parts is that it is difficult to detect feature points correctly [1]. On the other hand, many devices are presented in the holistic methods with additional cases, such as faces with a non-frontal direction, under bad lighting conditions, and with facial expressions. However, for all cases, the recognition rate of the face is up to 99.90%, with a false acceptance ratio (FAR) of 1%. It is now difficult to reduce the FAR value since the recognition rate is at the upper limit.

In this paper, we consider the feature-based method because it provides high-precision results if the feature point detection works well and if its work successfully provides digital precision, while the holistic method views a whole face image with rough ambiguity at most. Recognition improves for a well-conditioned image from the frontal direction and without facial expression. We use moving video as the input and automatically extract the best shot from the frontal direction and without expression, and use the best shot images for registering the face image and matching input and registered face images. To obtain the best shot from an input video sequence, there are two methods using feature points. One involves maximizing the distance between feature points while viewing the input sequence. The other involves minimizing the variance of the distance between feature points. In this paper, the latter method will be described.

One of the methods using feature points of 3D presented by Drira et al. [8] performs 99.2% utilizing distances of 3D curves for faces without expressions. Another method using 3D mesh and the distance function in a wavelet-transformed domain under bad lighting conditions presented by Toderici [9] outperforms the 2D case. However, the recognition ratio is not so high. Guillaumin et al. presented an improvement using a learning method utilizing the nearest neighbor method to feature point distances [10].

978-83-60810-66-8/$25.00

II. Previous Methods And Ideas For Improvement

In contests for face recognition, the recognition ratio for still pictures is reported to be 0.92 for rank 1 [4]. Further contests with age changes and with expressions are reported [5-6]. Also, most of these studies and reports settle some values of FAR (False Acceptance Ratio) or FRR (False Rejection Ratio), real distinguishing ability is not realized. From these results, the number of people that the face recognition system can distinguish without using any other information such as ID numbers is several hundred to a thousand. The so-called face-pass that permits a person to pass an authentication gate with only a face identification system without using an ID card does not make a service for more than thousand people. To improve the basic distinguishing ability of the number of people recognized and increase it above a thousand is the object of this paper.

To improve the recognition ratio, it is important to enhance the processing methods in the first stage of the system to acquire larger distances of feature values between people, not to incorporate additional conditions of age changes, expressions lighting conditions etc.

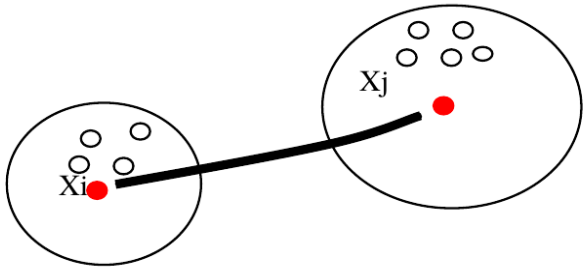

Fig.1 Feature points with the same name neighborhood[10].

III. Face Recognition Using Feature Point Distance

A. Proposed system[11][12]

Fig.2 shows the proposed face recognition system. Using input N frames, the best shot frame is selected by the method to be described in the following section. The best shot frame is registered in a database before recognition. At recognition, the best shot frame is selected and is verified using data in the database. The feature points of face parts are detected from the input face image. The detection is performed using software developed by Milborrow et al. [13] [14]. The software is based on the Active Shape Model and detects 77 feature points on the frontal face image. After detection of the feature points, two compensation operations of rotation and scale normalization are performed.

A(1) Rotation compensation[12]

The rotation compensation is to rotate at an angleθ which is a slope between the edges of both eyes. The rotation operation matrix shown in (1) is applied to the (x,y) coordinates of all feature points.

$$\begin{bmatrix} X \\ Y \end{bmatrix} = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} \quad (1)$$

After this rotation compensation, both eyes are aligned horizontally.

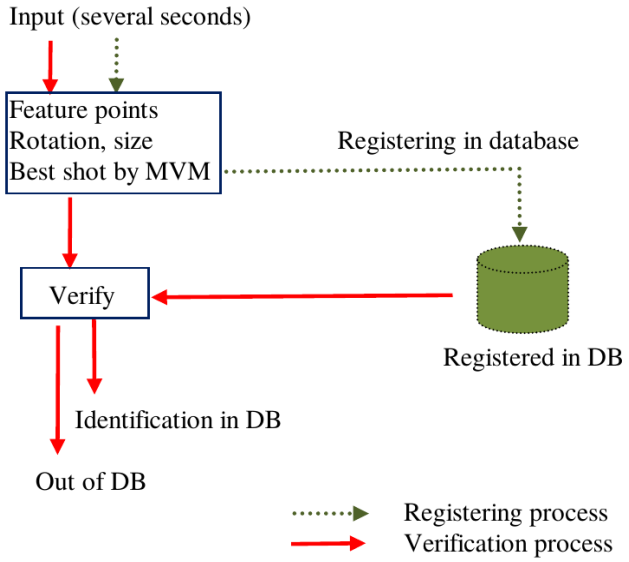

Fig.2 Face recognition system

A(2) Size and position normalization

Next, normalization according to the size is performed. This operation is applied to all feature points at the same rate as changing the size of a specified face part to a fixed size. In actual, let NF be

NF=sc/(X coordinate of the left edge of the left eye – X coordinate of the right edge of the right eye) (2)

Then, multiply this NF to all X and Y coordinates.

Position normalization is applied by moving all points in parallel after the size normalization. Let the input X and Y coordinate values of the first feature point be (In(0),In(1)) and the registered values be (Reg(0),Reg(1)), then the parallel movement value for X is Reg(0)-In(0), and for Y is Reg(1)-In(1)).

B. Best Shot Detection methods from video

An advantage of using video for face recognition is that it can utilize well-conditioned data and discard poorly conditioned data. Therefore, video provides temporal continuity, so classification information from several frames can be combined to improve recognition performance [15]. Also, tracking of detected facial regions is possible and the system can be expanded to carry out facial expression detection [16].

Two methods of obtaining the best shot are considered. One is maximizing the length between the two feature points. Another is removing the value with greatest deviation to minimize variance.

B (1) Maximizing Length Method

A moving head in the input video shows a three- dimensional rotation pattern. Figure 3 is a face with a frontal direction and without a facial expression, which is the best shot. We will detect this kind of best shot image from all varying

data. As for the Y and X axes rotations, the distances between feature points can be reduced by three-dimensional displacement and there is no data for compensation from the single camera environment. The Maximizing Length Method involves selecting a frame in which the distance between feature points is the maximum in all data.

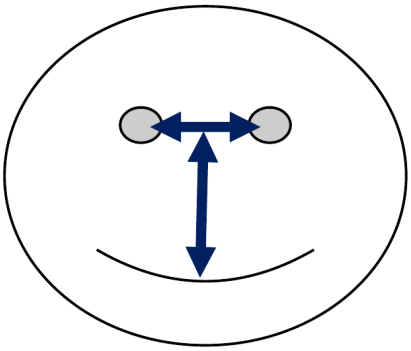

Fig.3 Regular face with frontal direction and without facial expression.

B (2) Minimizing Variance Method

The Minimizing Variance Method is used to remove irregular data from the total data to obtain concentrated data with smaller variance. The method actually removes iteratively the largest value distant from the temporal average value and makes a new data set. Figure 4(a)-(d) shows the iterative removal status for the case of a pre-obtained time sequence data set of feature points. The asterisks in Figure 4(a)-(c) indicate the largest value at each stage. A removed value is replaced by a dotted line. From Figure 4(a) to (c), two values are removed one after another, and the third largest value is marked in Figure 4(c). The same processes are repeated to reach Figure 4(d), which shows a temporally small variance as if the iteration may converge to a fixed average with the smallest variance.

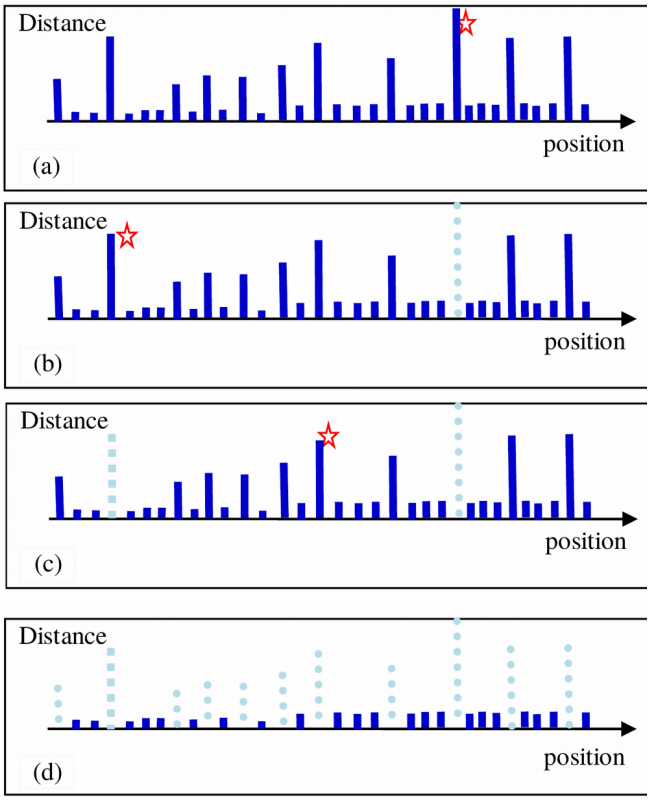

Fig. 4 Variance is reduced by removing the value with the greatest deviation.

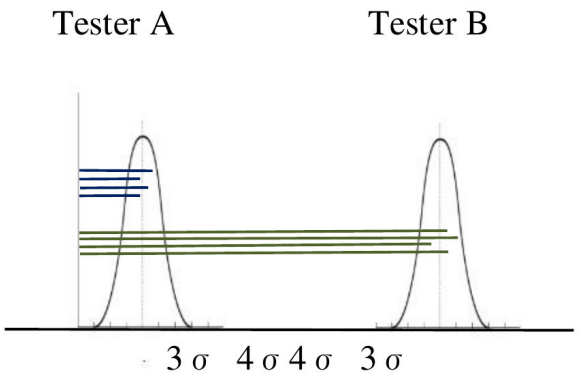

Fig.5 Distribution of two different persons.

C Estimation of the number of distinguishable people

To distinguish one person from all the other persons using the detected feature points, the values should be clearly different to each other. To reduce the FAR and FRR, the detected feature point values should have meaningful distances. Here, we try to estimate the distance distribution for a larger number of testers from a smaller number of testers. The distance between two feature points on a face is a feature value. Let us consider a distribution of this feature values that are length data. We assume this distribution is normal with mean m and standard deviation σ.

Fig.5 shows a one-dimensional distribution for a feature value of a distance between two values. The distribution is made from the data of many testers. When this distribution is normal, the value of a sample taken from the data set randomly is considered to form a random sequence with this distribution.

Fig. 6 shows an example. This shows the minimum distance among data taken one after another from data with a distribution of mean m and standard deviation σ.When this minimum value goes below a threshold, i.e. 1, we cannot distinguish at least one person among the data using the feature value. When the minimum value remains larger than the threshold, we can distinguish all people in the taken samples. Fig.6 shows the case of m=16, σ=4, 8, 16, 32, 64. The horizontal axis is the number of samples taken. This graph shows that the minimum value goes below the threshold for more than 5-20 samples.

IV. EXPERIMENTS

Four experimental items are carried out in the following sections. The Minimizing Variance Method (MVM) described in III B(2) is introduced in experiments to automati-

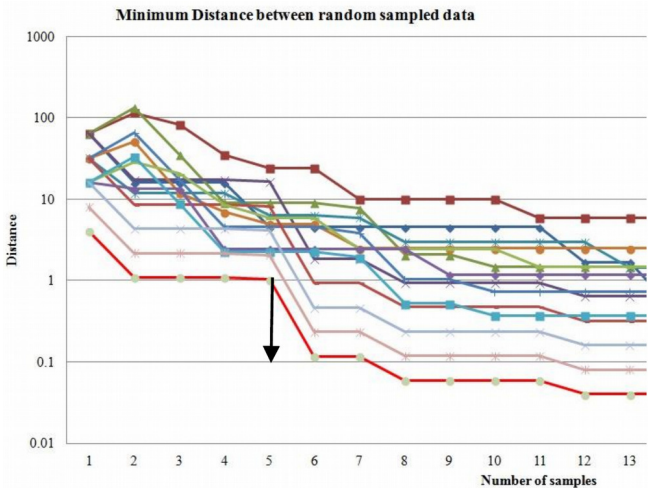

Fig. 6 Minimum distance among sampled random data. m=16, σ=64,32,16,8,4. This graph is based on a model generated by random samples from the normal distribution. Five values can be sampled keeping the distance values larger than 1 for all cases, which will be used in a later section, to allow taking samples five times. The width of windows is restricted to 13.

cally obtain frontal face images without expressions from input moving video of a person speaking for face recognition.

A. Reducing variance by MVM

First, by adapting MVM described in III B(2) to use with a recorded part of a video with a length of about one minute, the convergence status of variances is investigated. The number of pieces of distances between feature points is the combination of two out of 77 feature points, that is $_{77}C_2=2926$, where the total number of feature points is 77. For all this length data, let an initial value of the number of frames according to the time direction be "n". The algorithm involves the following three steps;

a. Obtain the average length for time direction with the total number of the lengths of n.
b. Remove the largest value that is distant from the average value.
c. We get a new set of length data in which the number of lengths is subtracted by 1. (n=n-1)

This process is repeated until we get n=1.

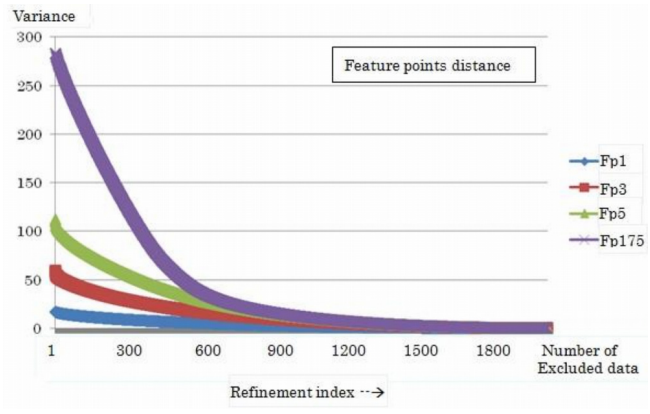

Fig.7 Variances of four distances between feature points converge as deleting iterations.

Figure 7 shows four data which were randomly chosen from 2926 results with graphs of variances as the process above proceeded. The variances at the position of 1000 times of iteration, which is about half the number of total iterations are greatly reduced. Also, at the position of 1500 times, which is about a quarter of the total number, the variance is almost below 1-2. Other results show nearly the same tendency, as shown in the graphs.

B. Variation by the positions of feature points

Figure 8 shows all 2926 distance values between feature points with well converging status according to the number of iterations. The numbers of iterations are from 500 to 1935, and their results are displayed overlapped. Figure 8(a) shows variances with the iteration numbers of 500 and above. Fig. 8(b) shows 1000 and above, Fig. 8(c) shows 1500 and above. These Figures show the variances reduce to small values after 1500 iteration. And after 1750, almost all variance data go below 2, and half of them seems to be below 1. A quarter to an eighth of all data seem to be below 1, which the input data are converge to stable values for register and recognition.

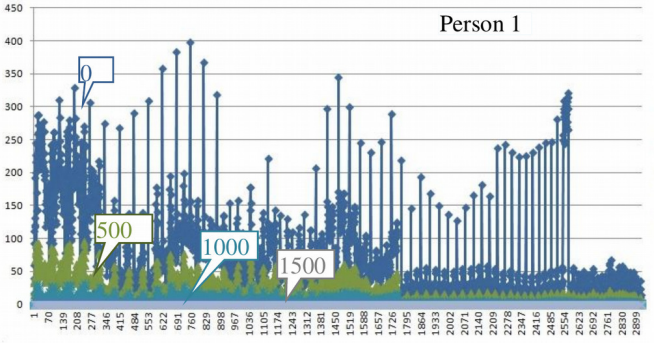

Fig.8 (a) Variances of distances between feature points converge as the number of iterations increases from 500 to 1935.

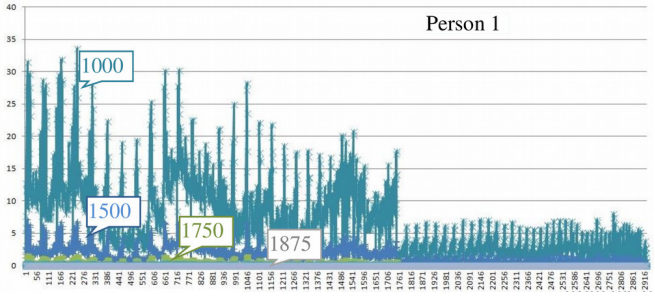

Fig.8 (b) Partial variances of distances between feature points converge as the number of iterations increases 1500, 1750, 1875, 1935. Enhanced in the vertical direction.

C. Maximum interval between minimum variance data

After adapting the MVM to the length data, a quarter to an eighth of the resulting data forms a set with a small variance of 1-2. This quarter means 3.75-7.5 Hertz in the time

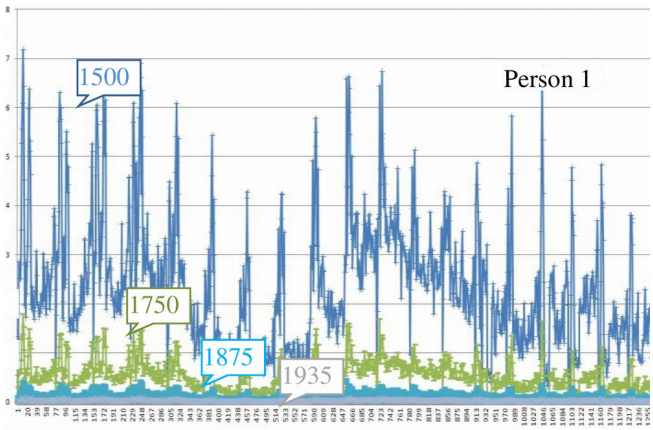

Fig.8 (c) Variances of distances between feature points converge as the number of iterations increases from 1500 to 1935. For the 1750 case, most feature values are below 2 and half of them are below 1.

direction because the original data was 30 Hertz. Face recognition can work at about 5Hz in the best conditions on average. But the value of 5Hz is average and we must consider the possible worst case. Then, the maximum interval between the converged data with small variances, which shows the worst case for detecting the best shot, is searched for. Figure 9 shows the maximum interval between the best shot frames vs. thresholds of deviation between the average value and the length value. The average values are the final values obtained by experiment III-A.

D Estimation of the number of distinguishable people

For identification by face recognition, the number of distinguishable people will be estimated. The total number of feature values is 2926 as the pieces of distances between feature points. It is important to select feature values whose distance value is large enough to distinguish people. Fig. 10 shows a hundred feature values from the first number. The number of feature values that have a distance larger than 5 is 45. The distance value larger than 4 can be assumed to be the case of a standard deviation larger than 4 in Fig.6. From the curves of Fig.6, the number of valid samples to be distinguishable is 5. From this result, and if we take four more feature values (in total five feature values) we can find the number of distinguishable people is $5^5=3125$ [17]. If we take one more feature value it will be $5^6=15625$.

We can estimate that there are at least five feature values from the different face parts of an eye, an eye brow, a mouth, a nose, and a contour, though the total 2926 data points may have correlations and are not independent of each other.

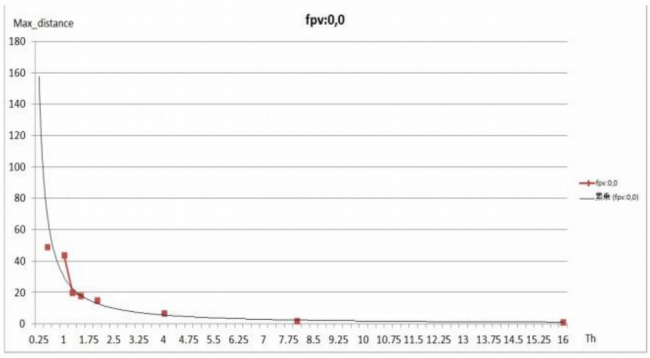

Fig.9 The maximum interval between the best shot frames vs. thresholds of deviation between the average value and length value. Dots are measured intervals. The value belongs to the best shot if it is smaller than the threshold.

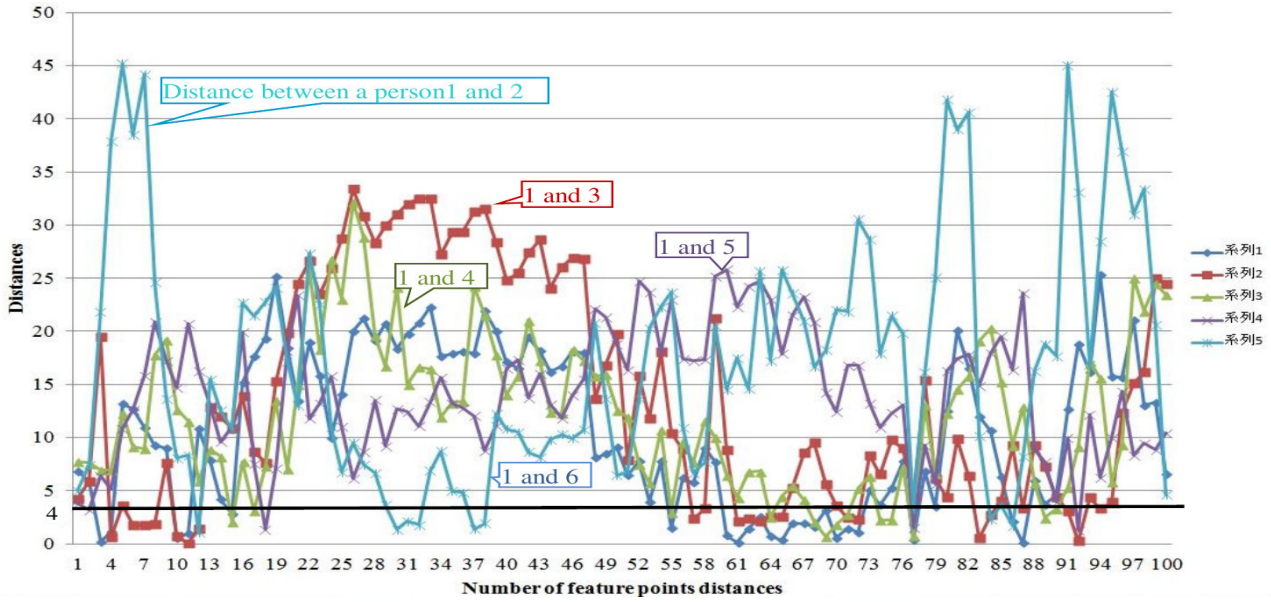

Fig.10 Five distance values obtained by six testers. The horizontal axis is the numbering of 100 feature values out of 2926. The feature values that have a distance of more than 4 are good distinguishable features for recognition.

V. CONCLUSIONS AND FUTURE WORK

A new method of extracting the best shot from moving video input for face recognition is proposed. The best shot is obtained at the average probability of 1/4 to 1/8, which means 3.75-7.5 Hz in the time direction on average. The best shot can be obtained for all 2926 combinations of feature points. The maximum interval between the best shots is 0.8 second. According to Fig6, feature values larger than 4 can be used five times when sampling people. According to Fig.10, more than five feature values can be used. The number of distinguishable people can be estimated up to 3125 based on normal distribution. If we take six features, it implies the features distribute within 4σ.

In the future, how to select more reliable feature values for 5σ will be studied.

REFERENCES

[1] Rabia Jafri and Hamid R Arabnia, "A Survey of Face Recognition Techniques", *Journal of information Processing Systems* Volume: 5, No: 2, pp. 41-68, 2009.

[2] Bonnen, K. Klare, B.F. Jain, A.K., "Component- Based Representation in Automated Face Recognition", *IEEE Transactions on Information Forensics and Security,* Vol.8, No.1 pp.239-253, Jan. 2013.

[3] JCB 2014 Conference report http://www.ijcb2014.org/IJCB%20Conference/IJCB14_Conference_Report.pdf

[4] Patrick J. Grother; George W. Quinn; P J. Phillips, "Report on the Evaluation of 2D Still-Image Face Recognition Algorithms", *NIST Interagency/Internal Report (NISTIR)* – 7709 June, 2010.

[5] M. Ngan and P. Grother, "Face Recognition Vendor Test (FRVT) Performance of Automated Age Estimation Algorithms", *NIST Interagency Report* 7995 Mar 2014.

[6] Patrick Grother Mei Ngan, "Face Recognition Vendor Test (FRVT) Performance of Face Identification Algorithms" *NIST Interagency Report* 8009 May 2014.

[7] Carl Gohringer, "Advances in Face Recognition Technology and its Application in Airports", *Allevate Limited*. Pp.1-10., 17 Jul, 2012.

[8] Drira, H. Ben Amor, B. ; Srivastava, A. ; Daoudi, M. ; Slama, R.," 3D Face Recognition under Expressions, Occlusions, and Pose Variations", *IEEE Transactions on Pattern Analysis and Machine Intelligence,* Vol. 35, ,Issue: 9, pp.2270 – 2283 Feb. 2013.

[9] Toderici, G.; Passalis, G. ; Zafeiriou, S. ; Tzimiropoulos, G., "Bidirectional relighting for 3D-aided 2D face recognition", *Proc. of IEEE Conference on Computer Vision and Pattern Recognition (CVPR),* pp. 2721-2728, 2010 June 2010

[10] Guillaumin, M. ; Verbeek, J. ; Schmid, C.," Is that you? Metric Learning Approaches for Face Identification", *Proc. of IEEE 12th International Conference on Computer Vision,* 2009 pp. 498–505, Sept. 2009

[11] Kazuo Ohzeki, YuanYu Wei, Yutaka Hirakawa, Toru Sugimoto, "Authentication System using Encrypted Discrete Biometrics Data", *Proceedings of TRUST 2014 Greece Springer LNCS* 8564 pp.210-211 June 30-July2 2014.

[12]]Kazuo Ohzeki, Masahiro Takatsuka, Masaaki Kajihara, Yutaka Hirakawa, Kiyotsugu Sato, "On the False Rejection Ratio of Face Recognition Based on Automatic Detected Feature Points", *Proc. international workshops on "Pattern Recognition and Image Understanding"OGRW-9* Mo.3-1, ogrw2014_024_Ohzeki.pdf Dec.2014.

[13] Stephen Milborrow, Fred Nicolls, "Locating Facial Features with an Extended Active Shape Model", *Proceeding of ECCV Part IV* pp.504-513, Springer-Verlag Berlin, Heidelberg 2008

[14] S. Milborrow and F. Nicolls, "Active Shape Models with SIFT Descriptors and MARS", *International Conference on Computer Vision Theory and Applications (VISAPP)* pp.380-387. 2014

[15] Howell and H. Buxton, "Towards unconstrained face recognition from image sequences," in *Proceedings of the Second IEEE International Conference on Automatic Face and Gesture Recognition,* 1996, pp.224-229.

[16] L. Torres, "Is there any hope for face recognition?" in *Proc. of the 5th International Workshop on Image Analysis for Multimedia Interactive Services (WIAMIS 2004)*. Lisboa, Portugal, 2004.

[17] Yasuko Tanaka, Eigo Miyazaki, and Kazuo Ohzeki, "Feature Point Analysis Using Facial Parts for Face Recognition", *National Convention, D-12-36* Institute of Electronics, Information, and Communication Engineers Mar. 2011 (in Japanese)