J. of the Chosun Natural Science
Vol. 5, No. 2 (2012) pp. 120 – 126

Gesture Recognition Using Higher Correlation Feature Information and PCA

Jong-Min Kim[1†] and Kee-Jun Lee[2]

Abstract

This paper describes the algorithm that lowers the dimension, maintains the gesture recognition and significantly reduces the eigenspace configuration time by combining the higher correlation feature information and Principle Component Analysis. Since the suggested method doesn't require a lot of computation than the method using existing geometric information or stereo image, the fact that it is very suitable for building the real-time system has been proved through the experiment. In addition, since the existing point to point method which is a simple distance calculation has many errors, in this paper to improve recognition rate the recognition error could be reduced by using several successive input images as a unit of recognition with K-Nearest Neighbor which is the improved Class to Class method.

Key words : Eigen Space, PCA(Principal Component Analysis), HLAF(Higher order Local Auto Correlation Features)

1. Introduction

Men provide various information using non-verbal means such as gestures and facial expressions. If it is possible to analyze these non-verbal communication means, it would be possible to build a natural and intelligent interface between man and computer. For analyzing human motions, existing 2D systems have many limitations in motion[1]. In order to supplement such a shortcoming, many researches to acquire 3D information and analyze motions using this information are in progress recently[2]. However, mathematically calculated 3D information contains lots of errors. Also, to use the non-verbal communication as VR interface, accurate and fast motion analysis methods are needed.

The gesture recognition can be divided into 2D gesture recognition and 3D gesture recognition largely depending on the target gesture. Since 2D gesture recognition is only possible to recognize the section of gesture, the gestures are mostly with flat faces[3,4]. In contrast, 3D gesture recognition is the method, which is possible to recognize every views of gesture. Since it needs to recognize every view, this is a problem which is not solved well compared to 2D gesture recognition and it is approached in various method[5].

In this paper, after converting input images into silhouettes images through preprocessing, input images are projected to the lower dimensional vector space, that is parametric eigenspace which can express the features of gesture appearance by a statistical technique called PCA (Principal Component Analysis) for features extracted through HLAF (Higher Order Local Autocorrelation Features). Each gesture is presented as a trajectory of consecutive points sequentially, and the gesture is recognized by comparing the trajectory pre-learned and trajectory of input image. In this paper, features using silhouette image and HLAF were used to improve configuration speed of eigenspace. Also, in the existing point to point method an incorrect matching that recognizes as an other gesture occurs frequently in the recognition process in case of matching the projected input image with model image, even though the actual gesture image is succeeded to match in the gesture space that projects several gestures. Therefore, in this paper the recognition error could be reduced using several consecutive input images using K-Nearest Neighbor which is improved Class to Class method to improve the recognition rate.

[1]Computer Science and Statistic Graduate School, Chosun University, Gwangju, 501-709, Korea
[2]Department of Health Education & Information, Gwangju Health College, Gwangju, 506-701, Korea

[†]Corresponding author : mrjjoung@hotmail.com

(Received : May 13, 2012, Revised : June 20, 2012, Accepted : June 24, 2012)

2. Preprocessing

2.1. Background Removal

The image sequence obtained through camera is the one that obtained in a simple background and in the image obtained from general environment, it includes a lot of gesture (background) that doesn't need for gesture recognition. However, since the gesture region (front view) is needed for gesture recognition, it is necessary to separate the background and body areas first and to do this the background model should be created first. However, because the brightness of the light is not constant and changes often, all are not equal and it is difficult to obtain stable background model even if the same background is taken with the same camera in a certain period of time.

In this paper, after obtaining the background image I_t for a certain period of time T_i by measuring the changes in the brightness of the background due to light changes and considering the time factor (t), pixel value $P_{\max}(x)$ was determined when the light was brightest and pixel value $P_{\min}(x)$ was determined when the light was darkest by analyzing each pixel (x) existing in the following image region R. The difference between these two pixel value $D(x)$ is the threshold of brightness that can appear with the change of light. By using these 3 factors, the Background Model will be configured. Such information is shown in equations (1-4).

$$BM = \{P_{\max}(x), P_{\min}(x), D(x)\}_{x \in R} \tag{1}$$

$$P_{\max}(x) = MaxI_t(x), (1 \le t \le T_i) \tag{2}$$

$$P_{\max}(x) = MinI_t(x), (1 \le t \le T_i) \tag{3}$$

$$D(x) = P_{\max}(x) - P_{\min}(x) \tag{4}$$

Once the background model is made, the binary image $B(x)$ will have the maximum value (255) if the difference value which is obtained from differential operation between the brightest pixel value $P_{\max}(x)$ of input image $I(x)$ and the darkest pixel value $P_{\min}(x)$ is larger than threshold $D(x)$ and otherwise it will have the minimum value (0).

$$B(x) = \begin{cases} 255 & if|P_{\max}(x) - I(x)|or|P_{\min}(x) - I(x)| > D(x) \\ 0 & otherwise \end{cases} \tag{5}$$

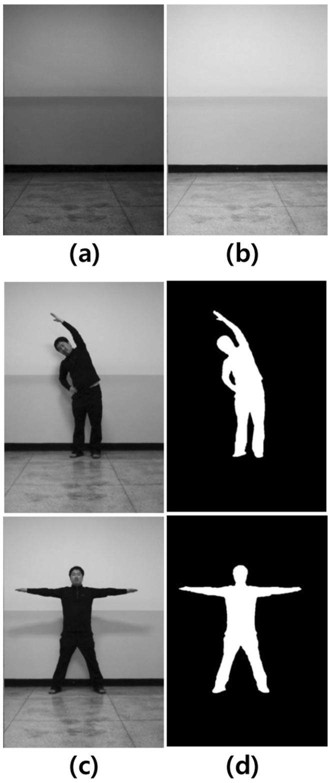

Fig. 1. (a) The minimum brightness of the image (b) The maximum brightness of the image (c) Input Image (d) Silhouette image extracted using the background model parameters of input image.

The equation (5) becomes the basis for separating the region that has the difference in the motion of gesture's pose changes ignoring the brightness differences that can be caused by the light. In the binary image obtained from the result of equation (2), due to changes in lighting which is off from the threshold of brightness value set from background model, it is classified as foreground region even though it is background, so small point of 1 pixel can be included. Therefore, to remove this noisy, Morphological operation was used. A single erosion operation was performed and the dilation operator was used to restore because the reduction of gesture was occurred at this time.

2.1. Higher Order Local Auto-correlation Features Generation

If the eigenspace is configured using normalized images obtained through the method described in the previous section, computations for 76,800 dimensions are needed because the size of gesture image is 320 * 240. In this paper, by using the Higher Local Auto-Correlation Coefficient to reduce the size of dimension, a method to reduce the feature vector size, that is the size of dimension into 25 was proposed.

HLAF (Higher order local auto correlation features) are known as a function of shift-invariant. The function which expands these auto-correlation features further is the higher local auto-correlation function. When P is marked in the image area, the auto-correlation feature of dimension can be defined as an equation (6) if N is replaced by $a_1,.......a_n$.

$$x_i^N = (a_1,\ldots\ldots a_n) = \int_P f(\tau+a_1)\ldots f(\tau+a_N)d\tau \qquad (6)$$

In equation (6), $f(\tau)$ expresses the gray level on τ. Since the auto-correlation features obtained by combining P which has a large area have too many number of coefficients for practical application, it should be reduced through certain process. Therefore, at first the range of N dimension is limited to 2 dimension ($N = 0,1,2$). The 0 dimensional auto-correlation feature means the average of gray level in P area.

As shown in Fig. 2, 3×3 local mast pattern is replaced within a specified range. The value of the center pixel among these local mask patterns will be referred and the number of pattern is limited to 25. By locating local mask patterns which have these 25 patterns into images sequentially, 25 feature vectors can be calculated. Fig. 3 is showing 25 patterns and the symbol * in here means “don’t care”.

HLAF calculates total 25 feature vectors for image area P through the sum of reference pixel values of each mask pattern using 25 3×3 mask patterns. These feature vectors apparently have the shift-invariant characteristics.

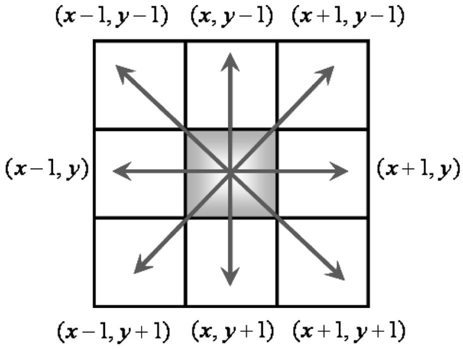

Fig. 2. 3×3 local mask pattern.

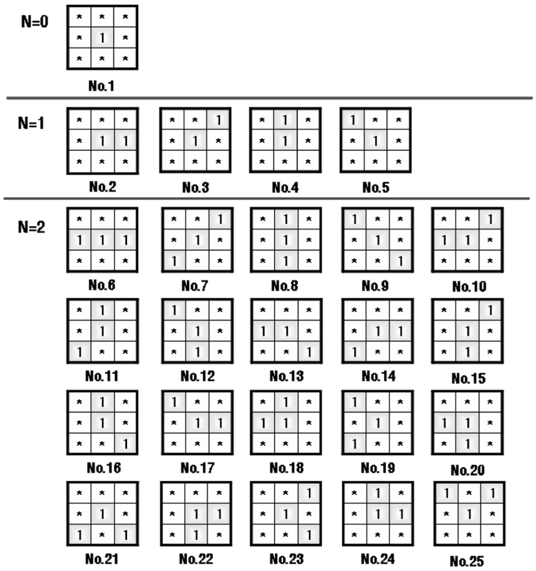

Fig. 3. 25 3×3 local mask patterns

The feature vector f^v obtained from Fig. 3 using local mask pattern has the feature vector of higher local auto feature defined as an equation (7).

$$f^v = f_1,\ldots\ldots,f_{25} \qquad (7)$$

If $I_{x,y}$ is the feature vector when 3×3 mask patterns are searched on x, y coordinates, the 0 dimension auto-correlation function of $I_{x,y}$ can be expressed as f_1. Therefore, each feature vector $f_1(i = 1, 2, 3,......,25)$ can be defined as an equation (8).

$$\begin{aligned} f_1 &= \sum_x\sum_y I(x,y) \\ f_2 &= \sum_x\sum_y I(x,y)I(x+1,y) \\ f_5 &= \sum_x\sum_y I(x,y)I(x-1,y-1) \\ f_{15} &= \sum_x\sum_y I(x,y)I(x,y+1)I(x+1,y-1) \\ f_{25} &= \sum_x\sum_y I(x,y)I(x-1,y-1)I(x+1,y-1) \end{aligned} \qquad (8)$$

The normalized gesture image from given input image generates 25 feature vectors containing gesture image using higher local correlation features. The gesture images obtained through gesture extraction algo-

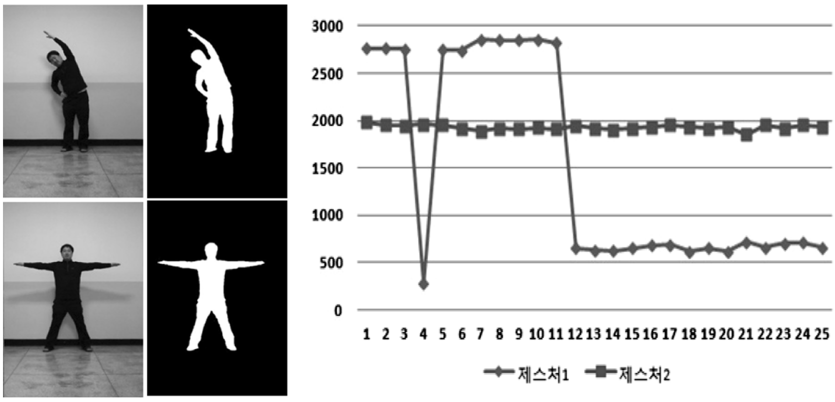

Fig. 4. Higher local correlation feature data of 2 gestures.

rithm generates feature vector by 3×3 mask pattern search. Extracted image which is binary image will be determined by the sum of the pixels for mask patterns. Fig. 4 is the data that extracts the higher local correlation feature from 2 gestures.

3. Space Generation Using PCA

Principal Component Analysis (PCA) can reduce the high dimensional input data set to low dimensional meaningful data set. PCA is also called as K-L transform and the K-L transform is the method that is aimed to reduce the dimension of image by maintaining the data distribution and characteristics in the feature region without using the class information. To calculate the eigenvector, if a set of M number of learning image with N (Row x Col) size is $X = [x_1, x_2, x_3, ..., x_n]$, the covariance matrices S that presents the difference can be defined as following. In here, r is average data.

$$S = \sum_{i=1}^{M} [x_i - r][x_i - r]^T, r = \frac{1}{M} \sum_{i=1}^{M} x_i \tag{9}$$

From the formula (5), the eigenvector and eigenvalue of S can be calculated using the following formula.

In here, e_i is the eigenvector and λ is the eigenvalue. In the next step, if the eigenvector e_i is arrayed in the large order of eigenvalues, the formula (11) that is consisted of p number of eigenvectors can be obtained.

$$W_{PCA} = [Xe_1, Xe_2, Xe_3, \ldots\ldots, Xe_p] \tag{11}$$

The eigenvector of p dimension obtained from the formula (11) has the eigenvalue for each image, so that this is called the feature vector. In here, the eigenvectors composed of orthogonal row is called the eigenface and the face image can be expressed with the linear combination of eigenface and feature vector obtained in here. However, the size of eigenvalue that each eigen vector has means the importance of that eigenvector, so the important eigenvector that specifies the eigenspace can be chosen by using the formula (12). Therefore, only the principal component vector that can represent many images without using every eigenvectors in eigenspace construction can be used.

$$\frac{\sum_{i=1}^{k} \lambda_i}{\sum_{i=1}^{N} \lambda_i} \geq T_1 \tag{12}$$

In here, T_1 is the threshold that controls the number of Eigen vector and if the eigenvector that is used when

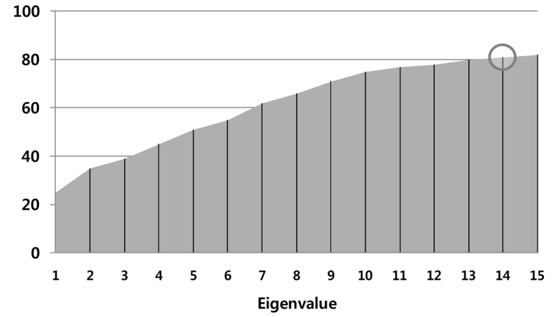

Fig. 5. Cumulative contribution depending on the number of eigenvalue.

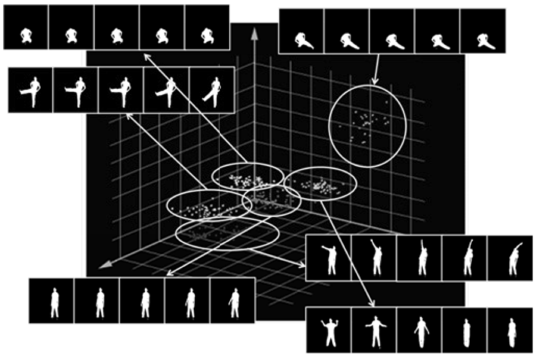

Fig. 6. The result of projecting 6 gestue image sequences to eigenspace.

evaluating the recognition and rotated pose is arrayed in the large order of eigenvalues to constructs the low dimensional space is shown in Fig. 5. In the study, k= 14 was used. The space constructed like this expresses as the face eigenspace.

The distribution within the eigenspace for each of 6 silhouette gesture poses is shown in Fig. 6.

4. Distance Evaluation and Gesture Recognition Using Improved K-NN

An incorrect matching that recognizes as other gesture occurs in case of matching the projected input image with model image (Point to Point), even though the actual gesture image is succeeded to match in the gesture image space that projects several gestures. To solve these problems, respective model images for several consecutive input images were used as a unit of recognition instead of matching of the single gesture image unit (Class to Class).

As shown in equation (13) and equation (14), K-Nearest Neighbor method was used as matching algorithm. Where, arg $S(M_j)=j$ is an operator that determines the number of model.

$$w=\frac{(\arg S(M_j)-Min(\arg S(M_j)))}{d(k-1)} \tag{13}$$

$$\frac{\sum\sum w(I_j-M_j)}{k} \tag{14}$$

By using the value obtained from the above equation (14), the recognition of model image and input image is determined. In this case, k = 14 where the cumulative ratio is more than 80% was used. Fig. 7 shows the matching method using the above method for consecutive input images and model images projected to the space.

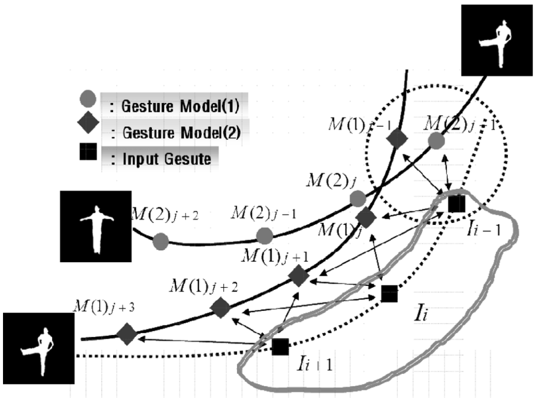

Fig. 7. Gesture image recognition using K-Nearest Neighbor.

As shown in Fig. 7, even though the image has the closest distance between input image and model image, there is a possibility that it is actually other gesture image. To solve this problem, the method suggested in this paper could know the consentaneity between trajectories which is the entire gesture image by performing the matching as consecutive several image units and the results obtained by using this could be used to evaluate the entire gesture image.

5. Experiment Results

The gesture image used in the test was detected by the silhouette as in Fig. 8.

The main problem in recognition using principal component analysis is that it has some difficulties to apply in real-time because the time that configures the eigenspace and speed that performs the gesture recognition are not fast. For that reason, to configure the eigenspace for gesture recognition, the gesture image taken by 640×480 was converted into 320×240 through size normalization. And, in this study silhouette image was created without using every gesture image data. Only with silhouette image of gesture, it has much of the gesture features so these features were used as input data for HALF. Also, the gesture silhouette image had 320×240 = 76,800 data and only 25 feature data among

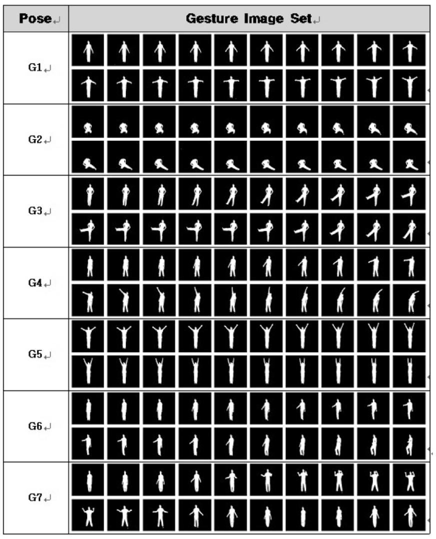

Fig. 8. Gesture image sequence using experiment

Table 1. PCA processing time and recognition rate using PCA and HLAF

Number of Model Image	PCA		PCA using HLAF	
	Training Time(sec)	Recognition Rate	Training Time(sec)	Recognition Rate
100	5.67	93.00%	0.016	89.00%
200	23.91	93.70%	0.017	89.70%
300	57.69	94.25%	0.047	90.00%
400	102.21	94.30%	0.094	90.50%
500	165.55	95.20%	0.150	90.70%
600	249.96	96.50%	0.214	91.50%
700	361.12	96.30%	0.287	91.90%
Average	138.02	94.75%	0.12	90.47%

Table 2. Successful matching rate according to the matching methods

Matching method	Matching failed	Incorrect matching	Matching succeeded
Existing Distance Calculation (Point to Point)	9.5 %	10 %	80.5 %
Improved k-Nearest Neighbor (Class to Class)	5.1 %	3.7 %	91.2 %

these data were extracted using HLAF. In other words, in the preprocess 76,800 data was reduced to 25 data.

Table 2 shows the analysis results for successful matching rate by each matching method. As shown in Table 2, the matching method using improved K-Nearest Neighbor has higher successful matching rate than the existing minimum distance matching method. Especially, it showed many improvement rate for incorrect matching.

6. Conclusion

In this paper, by combining HLAF and Principal Component Analysis the feature data obtained through HLAF from silhouette image which is the information of object form instead of using principal component analysis using existing object image was extracted and by utilizing this data in Principal Component Analysis, the method that lowers the dimension and maintains the recognition rate up to 90% and reduces the eigenspace configuration time significantly was suggested. It was proved through the experiment that the suggested method was easy to implement in the real world and was very well suited for real-time system establishment since it didn't require a lot of computations compared to the method using the existing geometric information or stereo images. However, there were difficulties in separating only object region in the complex background and in the future, it is plan to develop a more reliable object recognition algorithm by solving these problems.

References

[1] J.Aggarwal, Q. Cai, "Human Motion Analysis: A Review", Computer Vision and Image Understanding, Vol. 73, No. 3, pp. 428-440, 1999.

[2] Chistopher Wren, Ail Azarbayejani, Trevor Draeerl, and Alex Pentland. Pfinder:Real-time tracking of the human body. In Photonics East, SPIE Proceedings Vol. 2615 Bellingham,WA,1995.SPIE.

[3] I. B. Ozer, T. Lu, and W. Wolf, "Design of areal-time gesture recognition system: high performance through algorithms and software," *IEEE, Signal Processing Magazine*, Vol. 22, No.3, pp. 57-64, 2005.

[4] R. Bowden, R. Mitchell and M. Sarhadi, "Non-Linear Statistical Models for the 3D Reconstruction of Human Pose and Motion from Monocular Image Sequences," Image and Vision Computing, Vol. 18, No. 9, pp. 729-737, 2000.

[5] Rafael C. Gonzalez, Richard E. Woods, "Digital

Image Processing 2/E," Prentice Hall, pp. 519-532, 2001.

[6] Liang Xiao, Zhihui Wei, and Huizhong Wu, "Robust Orientation Diffusion Via PCA Method and Application to Image Super-Resolution Reconstruction", Lecture Notes in Computer Science LNCS.4688 pp. 726-735, 2007.

[7] J.-M. Kim, M.-K. Song, "Three Dimensional Gesture Recognition Using PCA of Stereo Images and Modified Matching Algorithm", IEEE Fuzzy Systems and Knowledge Discovery Vol. 4, pp. 116-120, Oct. 2008.