The Thirteenth International Symposium on Artificial Life and Robotics 2008(AROB 13th '08),
B-Con Plaza, Beppu, Oita, Japan, January 31-February 2, 2008

Face Recognition under varying illumination using Mahalanobis Self-organizing Map

Saleh Aly †, Naoyuki Tsuruta‡, and Rin-Ichiro Taniguchi†

† Department of Intelligent Systems, Kyushu University
744, Motooka, Nishi-ku, Fukuoka, 819-0395
‡ Department of Electronics Engineering and Computer Science, Fukuoka University
8-19-1, Nanakuma, Jonan, Fukuoka 814-0180
(Tel: 81-92-802-3580; Fax: 81-92-802-3579)
Email: † {aly, rin} @limu.is.kyushu-u.ac.jp, ‡ tsuruta@tl.fukuoka-u.ac.jp.co.jp

Abstract: We present an appearance-based method for face recognition and evaluate it with respect to robustness against illumination changes. Self-organizing map (SOM) used to transform the high dimensional face image into low dimensional topological space. However, the original learning algorithm of SOM uses Euclidean distance to measure similarity between input and codebook images, which is very sensitive to illumination changes. In this paper, we present Mahalanobis SOM, which uses Mahalanobis distance instead of the original Euclidean distance. The effectiveness of the proposed method is demonstrated by the experiments on Yale B and CMU-PIE face databases.

Keywords: self-organizing map, face recognition, Mahalanobis distance, illumination variation

I. INTRODUCTION

Important practical applications of automatic face recognition have made it a very popular research area in the last three decades [1]. In spite of the expanding research in the field of face recognition, many problems are still unsolved. Recently, more researchers focus on robust face recognition, which is invariant to pose, expression, illumination variations. Illumination variation is still a challenging problem in the research area of face recognition especially for appearance-based methods [2]. The same person can appear much differently under varying lighting conditions. Varieties of approaches have been proposed to solve the problem, and they can be classified into three categories: preprocessing and normalization [3],[4],[5], invariant feature extraction [6], and face modeling [7].

Self-organizing map (SOM) [8] is a well-known and quite widely used model that learns the distribution of patterns in an unsupervised neural network. A pattern is projected from an input space to a position in the map-where the information is coded as the location of the activated neuron. The SOM is unlike most classification or clustering techniques in that it provides a topological ordering of data. Similarity in the input space is preserved in the output space. The topological preservation of the SOM process makes it especially useful in the classification of data, which includes a large number of classes.

SOM has been previously used to solve face recognition problem. In [9, 10], the authors partitioned the face images into blocks that are subsequently used for the training of SOM. The image blocks are then encoded in a low-dimensional SOM space; however, the performance of their methods has not tested against illumination changes.

This paper focuses on the appropriate metric choice for face recognition using SOM. The distance most commonly used in SOM is the Euclidean distance that considers each observation dimension with the same significance whatever the observation distribution inside classes. Clearly, if the data set variances are not uniformly shared out among input dimensions, Mahalanobis distance is more appropriate than Euclidean distance.

The remainder of the paper is organized as follows: In section 2, the learning algorithm of self-organizing map neural network is presented. Section 3 mainly describes metric choice. Mahalanobis distance and Mahalanobis SOM learning algorithm are shown in section 4. In the following three sections, Face databases, experimental results, and conclusions are given.

II. Self-organizing map

In this section, we give a brief description of the SOM algorithm. SOM defines a mapping from an input space $\Re^n$ onto a topologically ordered set of nodes, usually in a lower dimensional space. A reference vector in the input space, $m_i \equiv [\mu_{i1}, \mu_{i2}, \ldots, \mu_{in}]^T \in \Re^n$, is assigned to each node in the SOM. In each training step, the following two steps are repeated for each input sample x.

(1) Find the best matching neuron c using a similarity measure between the input and all the map's neurons. This step name is winner-take-all (WTA), where c is the desired winner and should satisfy:

$$\|x - m_c\| = \min_i \|x - m_i\| \qquad (1)$$

(2) Update the weight vector of the winner c and also all its topological neighborhood in the map towards the prevailing input according to the rule:

$$m_i(t+1) = m_i(t) + h_{ci}(t)[x - m_i(t)] \qquad (2)$$

$$h_{ci}(t) = \alpha(t).\exp(-\|r_c - r_i\| / 2\sigma^2(t)) \qquad (3)$$

where $h_{ci}(t)$ is the neighborhood kernel function around the *winner c* at time t, $\alpha(t)$ is the learning rate and is decreased gradually toward zero and $\sigma^2(t)$ is a factor used to control the width of the neighborhood kernel.

The SOM codebook has the following characteristics:

(1) The probability Distribution Function (PDF) of the codebook is a good approximation for the PDF of the training data.

(2) The topographic order of the training data is preserved in the codebook, even if the dimensionality of the SOM is smaller than that of training data.

The second characteristic means that similar facial features are mapped to nearby positions in the feature map. This ordering takes place automatically without external supervision based on only the internal relations in the structure of the input patterns and the coordination of the neuron activities through the lateral connections among the neurons.

III. Metric choice

The original contribution of this paper lies in the use of Mahalanobis distance in the learning and recognition phases of SOM. The distance D_E *in* an N-dimensional Euclidean space between two patterns x and y is given in its generalized form with the following equation:

$$D_E(x,y) = (x-y)^T A^{-1}(x-y) \qquad (4)$$

where A is NxN symmetric and positive semidefinite matrix. The following three forms of the Matrix A may occur.

(1) If A is the identity matrix, it induces the traditional Euclidean distance.

(a) (b) (c)

Fig. 1. (a) and (b) represent the face image for the same person under different illumination condition while face image (c) represent another person

(2) If A is a diagonal matrix, it induces the weighted Euclidean distance.

(3) If A is the variance-covariance matrix of the data, it induces the Mahalanobis distance.

Figure 1 illustrates how the Euclidean distance measure is very sensitive to lighting changes. There are three face images, images (a), (b) represent the same person under different lighting condition, while face image (c) represent another person. A reasonable image metric should present smaller distance between Fig. 1a and 1b than that of Fig. 1b and 1c. However, the Euclidean distance gives counter intuition result. Computing the Euclidean distance yields $D_E(a,b)=121$ and $D_E(b,c)=58$. While Computing the Mahalanobis distance yields $D_M(a,b)=88$ and $D_M(b,c)=828$. *The* pair of the same subject has larger Euclidean distance; this phenomenon is caused by the fact that the Euclidean distance does not take into account the correlation among pixels. The traditional Euclidean distance is only a summation of the pixel-wise intensity differences and, consequently, small perturbation may result in a large Euclidean distance.

IV. Mahalanobis SOM

1. Mahalanobis distance

One important distance metric that can take care of hyperelipsoidal-shape clusters is Mahalanobis distance. Practical implementation of the Mahalanobis distance can be done by replacing the matrix A with the average of the within-class covariance matrix for each class, or by the total covariance matrix of all training data.

$$D_M(x,y) = (x-y)^T \Sigma^{-1}(x-y) \qquad (5)$$

One of the main issues associated with incorporating Mahalanobis distance in SOM is: if the number of patterns small compared to the input dimensionality, then the covariance matrix is singular and it is difficult to compute its inverse. To solve this problem, we first apply Principal component analysis (PCA) algorithm

[11] to reduce the dimensionality of the image, then using Euclidean distance in the new projected data.

2. The whitening transform

It is easy to show that Mahalanobis distance between two vectors in the input space is equivalent to the Euclidean distance between them in a transformed ("whitened") space, i.e.

$$D_M(x, y) = D_E(u, v) \tag{6}$$

And u, v can be obtained by the whitening transform

$$u = \Lambda^{-1/2} \phi X \tag{7}$$

$$v = \Lambda^{-1/2} \phi Y \tag{8}$$

Where $\Lambda = diag(\lambda_1, \lambda_2, ..., \lambda_d)$ and ϕ are the eigenvalues and the eigenvectors of the covariance matrix. Note that the "whitening" transform is a scaled principal component analysis (PCA).

3. Mahalanobis SOM

Equation (6) indicates that the computation of the Mahalanobis distance can be decomposed into two steps: (i) projecting the input space into the whitened space and (ii) computing the Euclidean distance in the whitened space. This two-step computation can directly embedded in the learning algorithm of SOM by preprocessing the input images using whitening transform and applying the original SOM learning algorithm in the new whitened space.

V. Face database

In this paper, the Yale face database B and the CMU PIE face database are both used to evaluate the proposed approach.

1. Yale Face database B

Yale face database B [7] is publicly available for studying pose and illumination problem in face recognition. There are 10 individuals under 64 different lighting conditions for 9 poses. Since this work mainly deals with the illumination problem, frontal images under varying lighting conditions is selected. The images are divided into five subsets according to the angle that the light source direction makes with the camera axis: Subset 1 (up to 12^0), Subset 2 (up to 25^0), Subset 3 (up to 50^0), Subset 4 (up to 77^0), and Subset 5 (up to 90^0). Face images are all cropped and aligned in accordance with [7].

2. CMU PIE Face Database

CMU-PIE face database [12] is available for studying pose, illumination, and expression problems in face recognition. There are 68 individuals under 43 different lighting and 3 different facial expressions for 13 poses. In our experiment, only frontal images under varying lighting conditions are selected, those which includes the images under 21 different directional flashes. Similar to Yale B face database, the images are divided into four subsets according to the angle that the light source direction makes with the camera axis

VI. Experimental results

In this experiment, subset 1 is used as the training set (gallery images) and other subsets are used for testing (probe images), all images are rescaled to the size of 48x48 in order to reduce the computation time.

It has been shown that images of an object under varying illumination conditions lie in a convex cone, called illumination cone, formed in the image space when the superposition property of illumination is assumed [7]. This cone can be approximated by 3-d hyperplane. However, previously proposed appearance based methods [2-6] do not sufficiently take into account that an images of an object under varying illumination conditions lie in a convex cone. Based on these observations, we propose an appearance-based method that utilizes properties of illumination cone in advance. We choose the shape of the map to be a cylindrical shape; moreover, the map dimension is chosen to be 3-dimensional.

To evaluate the validity of the proposed algorithm, we compared the performance of Mahalanobis SOM against 3 different approaches namely, correlation [13], Eigenfaces [11], and fisherfaces [6], furthermore, the accuracy of the original SOM is reported as shown in figures 2,3. In all the previously mentioned feature extraction methods, the accuracy calculated using nearest neighbor classifier based on the Euclidean distance between feature vectors. The number of features in each method is decided as follows: for the correlation method, the total number of pixels (48x48) represents the feature vector of the face image. While in Eigenfaces method, we select the eigenvectors that represent the 99% present of the energy of the eigenvalues. In Fisherfaces, the number of features is restricted by the number of classes. For SOM-based methods, each dimension in the map is considered as one feature; therefore, a 3-dimensional map represents

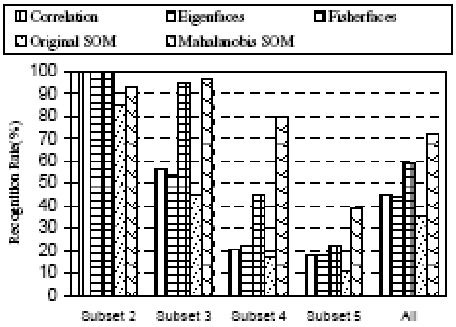

Fig. 2. Recognition performance for Yale B face database comparisons of different methods

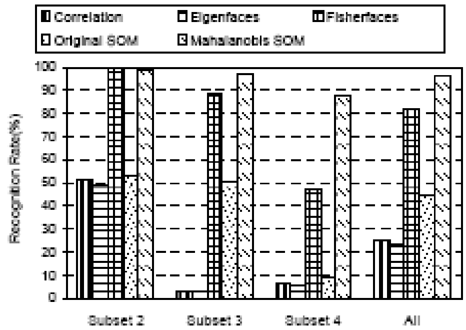

Fig.3. Recognition performance for CMU-PIE face database comparisons of different methods

each image with a vector contains three integer values. The feature vector in the SOM-based methods indicates the index of the winner neuron in the map.

It is clear from the results shown in Figures 2, 3 that the proposed method outperforms all other methods. In addition, the dimension of the feature vector is very small in comparisons with other appearance-based methods.

Selecting the appropriate distance metric seem promising rather than combining various normalization methods like histogram equalization and gamma correction with SOM [5].

Conclusion

The experimental results on Yale B and CMU-PIE face databases reveal a number of interesting points,

(1) The original SOM performance highly degrades when the test sample lighting condition is different from the training samples condition.

(2) Applying Mahalanobis distance for Face recognition using SOM highly improves its performance.

(3) Selecting the appropriate distance metric in SOM gives a good result than preprocessing images using histogram equalization or Gamma correction methods.

REFERENCES

[1] Zhao W. Y., Chellappa R., Rosenfeld A., and Phillips P. J.. "Face recognition: A literature survey", ACM Comput. Surv., pp.399-458, 2000.

[2] Li S. Z., Jain A. K., "Handbook of Face Recognition", Springer, 2005.

[3] Du B., Shan S., Qing L.; Gao W., "Empirical comparisons of several preprocessing methods for illumination insensitive face recognition", IEEE International conference on Acoustic, speech, and signal processing, vol.2,pii/981-ii/984, 2005.

[4] Chen W., Er M. J., and S. Wu," Illumination Compensation and Normalization for Robust Face Recognition Using Discrete Cosine Transform in Logarithm domain", IEEE Trans. On Systems, Man, and Cypernetics-part B: Cypernetics, 36(2):458-466, 2006.

[5] Aly S., Sagheer A.,Tsuruta N., and Tanigushi R. "Face Recognition across illumination" The 12th International Symposium on Artificial Life and Robotics (AROB), 2007.

[6] P.N. Belhumeur, J. Hespanha and D.J. Kriegman, "Eigenfaces vs. Fisherfaces: Recognition Using Class Specific Linear Projection," IEEE PAMI, 19(7):711-720, 1997.

[7] A. Georghiades, D. Kriegman, and P. Belhumeur." From few to many: Generative models for recognition under variable pose and illumination" IEEE Trans. on Pattern Analysis and Machine Intelligence, 23(6):630-660, 2001.

[8] T. Kohonen, "Self-Organizing Maps", 2nd edition, Springer, 1997.

[9] S. Lawarance, C. Giles, A. Tsoi, A. Back, "Face Recognition: A Convolutional Neural Network Approach", IEEE Trans. NN, 8(1):98-113, 1997.

[10] X. Tan, J. Liu, S. Chen, F. Zhang, " Recognizing Partially Ocluded, Expresion Variant Faces From Single Training Image per person With SOM and Soft k-NN Ensemble" IEEE Trans. NN, 16(4):875-886, 2005.

[11] M. Turk and A. Pentland, " Face recognition using Eigenfaces", In Proc. IEEE CVPR'91, pp.586-591, 1991.

[12] Sim T, Baker, S.; Bsat, M. "The CMU pose, illumination, and expression database", IEEE Trans. PAMI., 25(12): 1615-1618, 2003

[13] Brunelli R. and Poggio T., "Face Recognition: features versus templates", IEEE Trans. PAMI, 15(10):1042-1052,1993.