Proc. of the 2017 IEEE Region 10 Conference (TENCON), Malaysia, November 5-8, 2017

Robust Human Re-identification using Mean Shape Analysis of Face Images

Sarattha Kanchanapreechakorn, Worapan Kusakunniran
Faculty of Information and Communication Technology
Mahidol University
Salaya, Nakhon Pathom, Thailand
j.sarattha@hotmail.co.th, worapan.kun@mahidol.ac.th

***Abstract*—Human re-identification is an important component in many application domains especially the automatic surveillance system. This paper proposes a robust method to re-identify persons using their face shapes based on the Active Shape Model (ASM) and the Procrustes Shape Analysis (PSA). The ASM-based technique is used to extract landmark points of each face image, as the feature. Then, the Procrustes Distance (PD) is used to measure the similarity between two ASMs of any two face images. In addition, the trained ASMs of each subject can be grouped into clusters, using the PD-based K-mean clustering. Then, the Procrustes Mean Shape (PMS) is computed for each cluster using all belonging ASMs. In stead of using ASM of individual face image, the PMS is used as the representative face model. This process is performed to increase the robustness and reduce the number of models representing each subject. The proposed method is evaluated on the well-known face datasets and the real-world scenario of the security guard re-identification under the real environment. The experimental results and comprehensive comparisons show a very promising performance of the proposed method.**

***Index Terms*—Human re-identification, Active shape model, Procrustes shape analysis, Procrustes mean shape, Procrustes distance**

I. Introduction

Since the human face is a biometric feature that cannot be replicated easily, it is widely used for the human re-identification. It is a process of identifying a person using his/her face images, under a wide range of camera resolutions. The existing methods can be roughly divided into two categories which are the appearance-based method and the model-based method.

In the appearance-based method, it creates the face model using dominant appearances in the face image. [1] proposed the eigenface method for the face recognition. It applied the Principle Component analysis (PCA) on the training face samples to compute dominant eigenvectors which were used as the eigenfaces of each individual. [2] further developed the eigenface method by applying the super-resolution technique to improve the quality of face images captured by surveillance cameras. This is because the surveillance cameras had low-resolution. It was difficult to recognize human faces that were recorded from these cameras. [3] proposed the face recognition system based on the eigenface method, using the multi-view information of face images. [4] also proposed a method of face recognition based on the eigenface technique. The neural network was used as the recognition tool to achieve the higher accuracy, when compared with the traditional technique. However, it could not perform well for the testing images that had large variations from the training images, such as the variations in the lighting conditions. Thus, this method did not always work on the application that have to be used outdoor.

In the model-based method, it uses the landmarks detected in the face images to construct the face models, which can be translated, rotated and scaled into the new target image. [5] introduced the new technique to locate the landmarks' positions in the face image, called Active Shape Model (ASM). The ASM is the statistical model which can be deformed to fit the human face in the new target image. [6] applied the ASM with local patches to define face's landmarks, in order to overcome the illumination variation problems. It also used the Support Vector Machine (SVM) as the classifier. [7] presented a new method for the 3D face recognition using ASM. It could recognize the face images in 3D and 2D ranges. [8] proposed the extension of the ASM, called the Extended Active Shape Model. The extension covered fitting more landmarks, selectively using two of landmark templates, adding noises to the training set, relaxing the shape model, trimming co-variance matrices by setting most entries to zero, and stacking two ASMs in series. This made the ASM more robust and provided the better recognition result, when compared with the conventional ASM.

Based on our literature review, it can be seen that the model-based method could perform better than the appearance-based method for the datasets containing high variations of individual face images and/or scene's environments such as outdoor scenes. This is because it relies only on the landmark points instead of the whole face image which is very sensitive to multiple sources of changes including light intensity changes. Therefore, this paper proposes a new method for the robust face recognition for the human re-identification, based on the ASM. To reduce the sensitivity and the amount of individual face models, the trained ASMs of each subject are grouped into the clusters using the PD-based K-mean clustering. Then, the PMS [9] is calculated from all ASMs in each cluster to generate the face model. That is, the number of face models for each subject depends on the desired number of clusters. Then, the PD [10] is used in the similarity measurement.

Therefore, there are three main contributions in this paper.

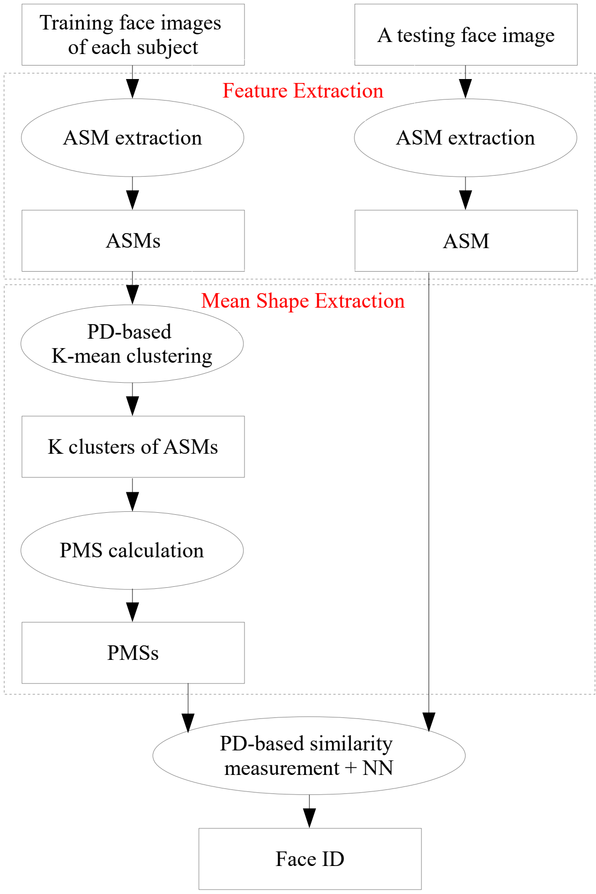

Fig. 1. The proposed framework.

First, ASM and PD are explicitly combined to solve the human re-identification using face images. Second, the PD-based clustering is used to reduce the number of face models for each subject and to also make the models more robust. Third, PMS is applied to extract the mean shape of face images for each individual.

The rest of this paper is organized as follows. The proposed method is explained in Section II. Then, the experimental results are illustrated in Section III and the conclusions are drawn in Section IV.

II. The Proposed Method

Fig. 1 shows the framework of the proposed method. In this figure, rectangles represent inputs/outputs, while ellipses represent processing steps. In the training phase, the ASM of each training face image is extracted. For each subject, the ASMs of his/her training face images are grouped into K clusters, using the PD-based K-mean clustering. Then, the PMS is calculated as the mean shape of all ASMs in each cluster, which will be used as the face model. That is, there are K PMSs (i.e. face models) of each subject. In the testing phase, the ASM of the testing face image is also extracted. Then, the PD is applied to calculate the similarity between the ASM of the testing face image and each trained PMS. Finally, the Nearest Neighbor (NN) is used as the classification tool.

The rational behind the proposed method is that, to achieve the high accuracy, the sufficient number of training images for each subject must be provided. This usually results in a large number of trained face models (i.e. ASMs), when having many subjects in the dataset. It leads to a high computation time for comparing the ASM of the testing face image against all ASMs of the training face images, which may not be suitable for a real-time application. Thus, as mentioned above, the proposed clustering technique is applied to decrease the total number of trained face models, by having only K models (i.e. PMSs) which can be pre-desired to be a small number. It also can reduce the sensitivity of the face model by grouping the similar face images together. They are then represented by one mean shape.

A. ASM

In this paper, the ASM is used as the main representation of the face image. It is the statistical shape model of the object that can translate, rotate and scale to define the shape in the target image [5]. It is used to model the key shapes in the face image including face shape, eyes, nose, mouth and eyebrows, which will be later used for the re-identification. The ASM extraction relies heavily on the pre-trained face models which were constructed using the set of annotated face images by human experts. Each model is the set of landmarks which can describe the typical face information over the training annotated samples.

Then, the ASM is extracted by best fitting (i.e. regarding to translation, rotation and scaling) the models to match the face information in the given image. The best fit is then used as the ASM which is the representation of one face image as mentioned. The ASM is in the form of the set of landmarks/key points $\{(x_1, y_1), (x_2, y_2), ..., (x_n, y_n)\}$, where n is the total number of points. To simplify the explanation of PSA in the later sections, the ASM is also referred as the shape configuration labeled by Z.

B. PD

In this paper, PD is used to measure the dissimilarity between two shape configurations (i.e. ASMs) of human faces. It is a minimal distance between the two shape configurations (Z_1 and Z_2) by considering translation, rotation and scaling as below [11][12].

$$Z_1 = \alpha 1_n + \beta Z_2, \quad \beta = |\beta| e^{j\angle\beta} \tag{1}$$

where α, $|\beta|$ and $\angle\beta$ represents translation, scaling and rotation of Z_2 to best match Z_1 respectively. Thus, the cost function for calculating the minimal distance is given below.

$$\min_{\alpha,\beta} ||Z_1 - \alpha 1_n - \beta Z_2||^2 \tag{2}$$

As mentioned above, the ASM is referred as the shape configuration (Z) which is the set of key points, where each key point is described using a complex number [9].

$$Z = \{z_i\}_{i=1}^{n} \tag{3}$$

where n is the total number of key points, $z_i = (x_i - x_c) + j(y_i - y_c)$, (x_i, y_i) is the coordinate of the key point i, and (x_c, y_c) is the coordinate of the shape centroid. Now, it can be seen that the centers of the two shape configurations are moved to the same coordinate. That is, the translation (α) is not needed. Then, the parameters in the equations (1) and (2) can be simplified [13] as below.

$$\alpha = 0, \quad \beta = \frac{|Z_1^* Z_2|^2}{||Z_2||^2} \tag{4}$$

where the superscript $*$ is the complex conjugation transpose, and β represents the similarity degree between the two shape configurations. Therefore, the PD is the inverse of the similarity which is calculated as below.

$$PD = 1 - \frac{|Z_1^* Z_2|^2}{||Z_1||^2||Z_2||^2} \tag{5}$$

The value of PD is between 0 and 1. The small value of PD represents the high degree of similarity between the two shapes. Then, the nearest neighbor [14] based on PD is used as the classification scheme for the re-identification.

C. PD-based K-mean Clustering and Mean Shape Representation

The K-mean clustering is used to cluster the training samples (i.e. ASMs of the training face images) of each subject into K groups in order to reduce the sensitivity of individuals and also reduce the number of trained face models for each subject. The ASM will be put into the nearest cluster by comparing with the cluster's centroid. The PD is used for the distance calculation and the cluster's centroid is the mean shape of all ASMs in the cluster.

The PMS is used as the mean shape (Z_m) of the cluster. The cost function for calculating the PMS is as below, by minimizing the distance to all ASMs in the cluster.

$$\min_{\alpha_i, \beta_i} \sum_{i=1}^{N} ||Z_m - \alpha_i 1_n - \beta_i Z_i||^2, \quad \beta_i = |\beta_i| e^{j\angle\beta_i} \tag{6}$$

where N is the total number of ASMs in the cluster, Z_i is the shape configuration i, and α_i, $|\beta_i|$ and $\angle\beta_i$ represents translation, scaling and rotation of Z_i to best match Z_m respectively. Thus, the PMS is calculated as the dominant eigenvector of the complex sum of squares and product matrix (S_Z) [9][12][13].

$$S_Z = \sum_{i=1}^{N} \frac{Z_i Z_i^*}{Z_i^* Z_i} \tag{7}$$

Then, the PMS is used as the face model. Thus, each subject has K trained face models. Since the PMS and the ASM are in the same format of the shape configuration, they can be compared using PD. Later, in the testing phase, the ASM will be extracted from each testing face image and compared with the trained PMSs of each subject, in order to find the best match for the re-identification.

TABLE I
THE PERFORMANCE COMPARISONS BASED ON THE FACE RECOGNITION DATA: UNIVERSITY OF ESSEX 94-96.

Methods	Accuracy (%)			
	Face 94	Face 95	Face 96	Face 94-96
PCA [18]	72	70	71	-
LDA [18]	79	77	78	-
LBP [18]	86	80	84	-
Gabor [18]	93	90	93	-
HLDA [19]	-	-	92	-
LDA [19]	-	-	93	-
NDA [19]	-	-	94	-
SVM [19]	-	-	98	-
SVM + LDA [19]	-	-	98	-
SVM+NDA [19]	-	-	99	-
PCA+NCC (male) [20]	-	-	-	93
PCA+NCC (female) [20]	-	-	-	85
Eigenfaces [21][22]	-	-	-	90
SRDA [21][23]	-	-	-	92
DFBFR [21][24]	-	-	-	94
Enhanced ASM [21]	-	-	-	95
The proposed method	-	-	-	100

III. EXPERIMENTS

In this paper, the experiments are divided into two main parts. The first part is to evaluate the proposed method using the well-known published face databases including the Psychological Image Collection at Stirling (PICS): Pain expressions [15], the Face Recognition Data: University of Essex 94-96 [16], and the Japanese Female Facial Expression (JAFFE) [17]. Then, the comparisons with the existing methods are illustrated. The second part is to evaluate the proposed method in the real-world scenario of the security guard re-identification under the real environment in both day and night time. The number of clusters used in the experiments is five.

A. Face Recognition Data: University of Essex 94-96

This dataset contains 377 subjects, in which 20 images were captured for each individual. For each subject, the first 10 images were used as the training data and the rest of 10 images were used as the testing data. It consists of 3 subsets which are Face 94 containing 153 subjects, Face 95 containing 72 subjects, and Face 96 containing 152 subjects. Each image in the dataset is represented with the resolution 180×200 pixels of the frontal face. The performance comparisons are shown in the Table I.

Some methods were evaluated using only the subset of the dataset i.e. Face 94 or Face 95 or Face 96. However, for the last seven methods including the proposed method shown in the table, they were evaluated using all three subsets i.e. Face 94-96 which covers all individuals. It can be seen that the proposed method outperforms the other existing methods.

TABLE II
THE PERFORMANCE COMPARISONS BASED ON THE PSYCHOLOGICAL IMAGE COLLECTION AT STIRLING (PICS): PAIN EXPRESSIONS.

Methods	Accuracy (%)
PCA [18]	68
LDA [18]	73
LBP [18]	78
Gabor [18]	89
KPCA [25]	93
PNMF [25]	92
NCK-NMF [25]	94
NCK+PCA [25]	94
The proposed method	100

TABLE III
THE PERFORMANCE COMPARISONS BASED ON THE JAPANESE FEMALE FACIAL EXPRESSION (JAFFE).

Methods	Accuracy (%)
Discriminative SIFT feature [26]	95
SVM+NDA [19]	95
SVM+LDA [19]	95
NDA [19]	85
LDA [19]	85
HLDA [19]	85
PCA+FLD+ANN [27]	85
The proposed method	90

B. Psychological Image Collection at Stirling (PICS): Pain expressions

This dataset was collected from 23 subjects of both male and female with different facial expressions. There are 26 images for each subject. For each subject, the first 13 images were used as the training data and the rest of 13 images were used as the testing data. The performance comparisons are shown in the Table II. It can be seen that the proposed method can achieve the highest accuracy.

C. Japanese Female Facial Expression (JAFFE)

This dataset contains 10 Japanese female subjects with 23 face images for each individual. For each subject, the first 11 images were used as the training data and the rest of 12 images were used as the testing data. Each face image is in the grayscale format with the resolution of 256×256 pixels. The dataset contains the frontal faces with different facial expressions. The performance comparisons are shown in the Table III.

The proposed method can achieve the promising performance. Although it cannot achieve the highest accuracy, it can reduce the number of trained models per subject. In this experiment, since 11 images were used as the training data for each subject, thus the total number of models representing each individual could be 11. However, based on the proposed step in section II.C, the total number of models were reduced to be 5. This can be useful in the real application, in which a large number of training images per subject is needed to achieve the reliable performance.

D. Real-world Scenario

In this experiment, the proposed method is evaluated under the real environment (i.e. outdoor) by re-identifying the security guards in the guardhouse of the housing estate. The total number of the security guards is 11. They were asked to sign-in (i.e. 7.00 a.m.) and sign-out (i.e. 7.00 p.m.) for their works by using their faces in real-time. Therefore, in this experiment, the android application was created for capturing and re-identifying human faces in real-time.

In addition, there were 30 training face images of each security guard. Then, the proposed PD-based K-mean clustering were applied to cluster the training face images of each security guard into 5 groups. The PMS was constructed as the mean shape of each cluster, which was used as the face model. Thus, there were 5 trained face models for each security guard. In total, there were 55 face models for 11 security guards, which were uploaded together with the developed application into the android mobile phone. The mobile phone used in this experiment was requested to be low-cost, due to the feasibility in the practical use. Thus, reducing the number of trained models from 330 (i.e. 11 subjects $\times$ 30 training face images) to 55 (i.e. 11 subjects $\times$ 5 clusters) can significantly speed up the performance of the application in the re-identification stage.

The experiment took 3 days. The result was very promising. The accuracies were 95%, 100% and 100% in day-1, day-2 and day-3, respectively. When the security guards got used to how to use the application, the accuracy became higher. It can be seen that the proposed method worked well under both day time and night time with different lighting conditions.

IV. CONCLUSION

This paper proposes the robust human re-identification method based on the mean shape of the human faces, using the ASM and the PSA. Each subject is represented by K PMSs. This is obtained by clustering the training face images of each subject into K groups using the PD-based K-mean clustering. The similar face images are put into the same group. Then, the PMS is constructed for each cluster. This process provides the benefits of reducing the total number of face models of each subject and decreasing the sensitivity of each individual. Finally, the PD is used for calculating the dissimilarity between two shape configurations. Based on the experiments, it can be seen that the proposed method achieves the very promising performance and can outperform the existing methods in the literature reviews. In addition, the proposed method is shown to work efficiently in the real scenario of the security guard re-identification under the real environment of both day and night time.

ACKNOWLEDGMENT

This research project was partially supported by Faculty of Information and Communication Technology, Mahidol University. It was also supported by Waller Security Service Co., Ltd.

REFERENCES

[1] M. Turk and A. Pentland, "Eigenfaces for recognition," *Journal of cognitive neuroscience*, vol. 3, no. 1, pp. 71–86, 1991.

[2] B. K. Gunturk, A. U. Batur, Y. Altunbasak, M. H. Hayes, and R. M. Mersereau, "Eigenface-domain super-resolution for face recognition," *IEEE transactions on image processing*, vol. 12, no. 5, pp. 597–606, 2003.

[3] H. F. Liau, L.-M. Ang, and K. P. Seng, "A multiview face recognition system based on eigenface method," in *Information Technology, 2008. ITSim 2008. International Symposium on*, vol. 2. IEEE, 2008, pp. 1–5.

[4] M. Agarwal, H. Agrawal, N. Jain, and M. Kumar, "Face recognition using principle component analysis, eigenface and neural network," in *Signal Acquisition and Processing, 2010. ICSAP'10. International Conference on*. IEEE, 2010, pp. 310–314.

[5] T. Cootes, E. Baldock, and J. Graham, "An introduction to active shape models," *Image processing and analysis*, pp. 223–248, 2000.

[6] U. Prabhu and K. Seshadri, "Facial recognition using active shape models, local patches and support vector machines," *contrib. andrew. cmu. edu*, pp. 1–8, 2009.

[7] S.-J. Park and D.-W. Shin, "3d face recognition based on feature detection using active shape models," in *Control, Automation and Systems, 2008. ICCAS 2008. International Conference on*. IEEE, 2008, pp. 1881–1886.

[8] S. Milborrow and F. Nicolls, "Locating facial features with an extended active shape model," in *European conference on computer vision*. Springer, 2008, pp. 504–513.

[9] W. Kusakunniran, Q. Wu, J. Zhang, and H. Li, "Speed-invariant gait recognition based on procrustes shape analysis using higher-order shape configuration," in *Image Processing (ICIP), 2011 18th IEEE International Conference on*. IEEE, 2011, pp. 545–548.

[10] A. Mostayed, S. Kim, M. M. G. Mazumder, and S. J. Park, "Face recognition using 3d head scan data based on procrustes distance," in *Intelligent Engineering Systems, 2008. INES 2008. International Conference on*. IEEE, 2008, pp. 203–208.

[11] J. T. Kent, "New directions in shape analysis," *The art of statistical science*, vol. 115, 1992.

[12] L. Wang, T. Tan, W. Hu, and H. Ning, "Automatic gait recognition based on statistical shape analysis," *IEEE transactions on image processing*, vol. 12, no. 9, pp. 1120–1131, 2003.

[13] A. Ericsson, *Automatic shape modelling with applications in medical imaging*. Centre for Mathematical Sciences, Lund University, 2006, vol. 2006, no. 6.

[14] E. Fix and J. L. Hodges Jr, "Discriminatory analysis-nonparametric discrimination: consistency properties," DTIC Document, Tech. Rep., 1951.

[15] P. Hancock, "Psychological image collection at stirling," 2004.

[16] L. Spacek, "Face recognition data. university of essex," *Computer Vision Science Research Projects*, 2008.

[17] M. J. Lyons, S. Akamatsu, M. Kamachi, J. Gyoba, and J. Budynek, "The japanese female facial expression (jaffe) database," in *Proceedings of third international conference on automatic face and gesture recognition*, 1998, pp. 14–16.

[18] F. Ahmad, A. Najam, and Z. Ahmed, "Image-based face detection and recognition:" state of the art"," *arXiv preprint arXiv:1302.6379*, 2013.

[19] N. M. Khan, R. Ksantini, I. S. Ahmad, and B. Boufama, "A novel svm+nda model for classification with an application to face recognition," *Pattern Recognition*, vol. 45, no. 1, pp. 66–79, 2012.

[20] T. F. Karim, M. S. H. Lipu, M. L. Rahman, and F. Sultana, "Face recognition using pca-based method," in *Advanced Management Science (ICAMS), 2010 IEEE International Conference on*, vol. 3. IEEE, 2010, pp. 158–162.

[21] C. A. Behaine and J. Scharcanski, "Enhancing the performance of active shape models in face recognition applications," *IEEE Transactions on Instrumentation and Measurement*, vol. 61, no. 8, pp. 2330–2333, 2012.

[22] M. A. Turk and A. P. Pentland, "Face recognition using eigenfaces," in *Computer Vision and Pattern Recognition, 1991. Proceedings CVPR'91., IEEE Computer Society Conference on*. IEEE, 1991, pp. 586–591.

[23] D. Cai, X. He, and J. Han, "Srda: An efficient algorithm for large-scale discriminant analysis," *IEEE transactions on knowledge and data engineering*, vol. 20, no. 1, pp. 1–12, 2008.

[24] H. Demirel and G. Anbarjafari, "Data fusion boosted face recognition based on probability distribution functions in different colour channels," *EURASIP Journal on Advances in Signal Processing*, vol. 2009, p. 25, 2009.

[25] Y. Zhao, W. Chen, B. Pan, and B. Chen, "Non-negative compatible kernel construction for face recognition," in *Chinese Conference on Biometric Recognition*. Springer, 2015, pp. 19–26.

[26] A. Majumdar and R. K. Ward, "Discriminative sift features for face recognition," in *Electrical and Computer Engineering, 2009. CCECE'09. Canadian Conference on*. IEEE, 2009, pp. 27–30.

[27] P. Tsai and T. Jan, "Expression-invariant face recognition system using subspace model analysis," in *Systems, Man and Cybernetics, 2005 IEEE International Conference on*, vol. 2. IEEE, 2005, pp. 1712–1717.