Abstract

The conventional ways of recognizing faces always assume the possession and heavily relies on extensive and representative datasets, but that is not the case in most real-world situations where more often than not, a very limited or even only single sample per person (SSPP) is available which ultimately rendering most face recognition systems to fail severely. This paper proposes a development of face recognition based on a combination of traditional eigenface with artificial neural network (ANN), having the face recognition performance boosted by the classification of discriminant vectors learned from a set of generic samples. The discriminant vectors representing intra-subject and inter-subject variations are learned based on similarities of pairs of generic samples which then used to classify novel intra-subject pairs and inter-subject pairs from probe set and corresponding gallery set. After that, the resulting classification is used to recognize faces by combining it with the expressive ability of eigenface via a voting procedure. The proposed method when tested with FERET and YALE datasets suggests that in face recognition within the SSPP constraints, the performance of the proposed method is better than some state-of-the-art methods.

Keywords: Face recognition; Eigenface; Similarity matching; Single sample; Generic learning

1. Introduction

There are several ways to recognize a person from another person. Face, fingerprint, DNA, gait and iris are among biometrics properties that are widely used for person recognition. However, face recognition is the prominent approach due the non-expensive implementation and non-obtrusive nature of the image acquisition which may occur without voluntary subject participation [1, 2]. However, as the face recognition applications expand, it raises another set of challenge entirely due to the difficulty to collect face images. The solution to this problem is to reduce the number of face down to one sample per person. But, many face recognition techniques heavily rely on the number of training samples available. Nevertheless, as a result of using only one training sample per person, most of the established face recognition techniques such as Fisherface [3] and Eigenface [4] will suffer serious performance drop and might as well failed to work completely. In fact we have tested several popular methods such as Eigenface, Support Vector Machines (SVM) [5], ANN and Fisherface by randomly selecting 50 persons from FERET [6] dataset to see the degree of performance drop suffered caused by SSPP problem. The drop is measured by finding the difference of the performances between single sample per person and nine samples per person. ANN yields 35.6% reduction in performance, SVM is at 10.4%, while Eigenface (PCA) has dropped 16.8% and Fisherface yields 20.5% drop. This result revealed that the difficulty posed by SSPP is in entirety caused by the
existence of too many variables but very limited observations when identifying each facial images of a person taken under various views and across different illuminations based on just single full-frontal image. SSPP can be traced back to being considered in the early work of Poggio and colleagues by generating virtual samples [7, 8]. There are several other work in literatures which attempt to analyze SSPP issue such as a review in [9], or to solve it for instance by using probabilistic matching and motion estimation [10-12]; learning the subspace that represented each individual by representing the face subspace with Self-organizing Maps (SOM) [13]; using Principal Component Analysis (PCA) to interpret facial features and synthesize realistic frontal face images [14, 15]; employs feature selection process on the extracted eigenfaces [16]; using Linear Discriminant Analysis (LDA) based on Fisherface [17] and components [18]; and Projection-Combined PCA \((PC^TA)\) by acquiring more information from the original face via combining the original image with its projection map [19, 20].

The PCA and Fisher’s LDA are two of the most prominent subspace approach in feature extraction process in face recognition. While Fisherface produces a discriminant features, Eigenface on the other hand represent face in expressive features of low dimensional subspace with maximum data variance. LDA based algorithm is commonly held as superior than PCA based ones, since LDA has the low-dimensional representation of the objects optimized whereas PCA offers simply object reconstruction [3, 21]. However when there is a limitation on the number of samples available, LDA is not be readily applied and requires measure of the class separability [22], since the within class scatter matrix, \(S_w\) cannot be estimated.

2. Generic Learning and Discriminant Vectors

Since SSPP problem mainly lies in the learning part of the recognition, it might as well be related to the fundamental of machine learning in broad. According to [23] human did not learn just any particular skills - they also learned the bias such that they learn how to generalize and assume, enabling human to learn and generalize from fewer examples. In fact, other report such as [24] have stated based on psychological experiment results, for the case of face recognition even one sample is sufficient for human to learn from single image and generalize the novel views. It is almost clear that human do learn some biases that help in achieving such brilliant feat. Related to this concept, it also has been shown that using prior knowledge a recognition system trained on only a single example for each class can perform better than that of the one using thousands of training examples per character [25].

The underlying assumptions in generic learning in the context of face recognition is both the intra-subject variations \(\Omega_I\) (corresponding variations between same individual) and the inter-subject variations \(\Omega_E\) (corresponding variations between different individual) are exhibited in a similar fashion, and thus can be approximated by estimation of both variations from a generic large and representative population. The discriminant vector \(\Delta D\) which characterizes \(\Omega_I\) and \(\Omega_E\) is given simply as the intensity difference between pair of images \(x_i\) and \(x_1\) in Eq. (1) [26]:

\[
\Delta D = \text{norm} \left[ x_1 - x_2 \right]
\]

Even though the above formulation in Eq. (1) is simple and straight-forward, \(\Delta D\) is a very high dimensional vector and having large degree of freedom due to the fact that \(\Omega_I\) can be contributed by several factors such as illumination variations, poses, facial accessories, and facial expressions while \(\Omega_E\) is contributed simply by the difference in subject identity. We can exploit the ability of ANN in learning nonlinear relationships to learn prior knowledge of \(\Delta D\) from sufficiently large generic samples consisting inter-subject faces and intra-subject faces. Based on this idea, we propose the use of image similarities between pair of face images in a generic training set trained to learn the \(\Delta D\), in which each person has one or more than one training sample. The approach of using image similarities for face recognition is readily proven to be viable in [26] where the probability density functions describing \(\Delta D\) for two mutually exclusive \(\Omega_I\) and \(\Omega_E\) between two facial images are obtained from training data using an eigenspace density estimation technique. The probability density is then used to compute a similarity measure based on the a posteriori probability of membership in the intra-subject class, which is used to rank matches in the database. Similarly, in this paper we used the biased information of \(\Delta D\) to discriminate intra-subject sample from inter-subject sample and thus reducing the number of possible candidates to a person. The candidates are further ranked according to the confidence score for intra-subject class produced by the PCA and eigenface method in order to select the best match from an SSPP gallery. The underlying reason why we use the candidates ranking approach, which is also recommended in [6], is since only limited samples are available, top one recognition rate is non-satisfactory for a realistic application [16].

3. Face Recognition Framework and PCA Overview

For the purpose of SSPP face recognition in this paper, we propose an appearance-based faced recognition which
processes the face image holistically using PCA and eigenface, while utilizing biased knowledge from a trained ANN. We use a generic set \( G \) consisting of two classes of images; \( X_G^I \) consisting \( N_I \) set of paired intra-subject images and \( X_G^E \) consisting \( N_E \) set of paired inter-subject images. The generic sets images can be denoted as \( X_G = \{ x_G^k, k = 1, 2, \ldots, (N_E + N_I) \} \) where \( x_G^k \) is \( k \)-th pair of inter-subject or intra-subject face image. This generic training set can be collected from any face dataset that exhibits variations similar to the variations of faces to be tested. Another set called gallery images, \( G \) consisting only single fully frontal image (with neutral face expression preferred) per person where the person in gallery set strictly does not belong to the persons in generic set such that \( G \notin G \). Images in gallery images can be denoted as \( X_G = \{ x_G^k, k = 1, 2, \ldots, N_G \} \) where \( x_G^k \) is face image of \( k \)-th person. Another set of images which is specifically for testing is called probe set \( P \), where it consists of novel images of same person as in gallery set following the closed universe scheme [6]. Images in probe images can be denoted as \( X_P = \{ x_P^k, k = 1, 2, \ldots, N_P \} \) where \( x_P^k \) is \( k \)-th novel view in probe set. The total number of samples used in these 3 sets follows that identities in probe sets should be equal to number of identities in gallery sets, \( N_P = N_G \), then \( N_P > N_G, (N_E + N_I) \gg N_G \), and based on experimental evidence, optimum generic set size should follow \( N_E = 2N_I \). Any value higher does not produce significant increase in classification performance that justifies the inevitable longer training time.

Each face image in all sets is represented by a column vector \( x \), where \( x \in [0,1] \) with the vector length, \( L \) equal to \( w \times h \) where \( w \) is the width and \( h \) is the height of the image. Each image is in grayscale and scaled to the size of 60 pixels \( \times 60 \) pixels. The images are then pre-processed with the histogram equalization to reduce the variations in illumination. The face recognition (FR) system later will attempt to recognize pair of all face images from probe set \( P \) and gallery set \( G \) to determine whether they are intra-subject pair \( P_I \) or inter-subject pair \( P_E \) and later assigns confidence scores to them based on the framework illustrated in Fig. 1.

In this work, we use PCA for both dimensionality reduction of \( \Delta D \) and for computing the principle components of sample covariance matrix based on eigenface method. The covariance matrix \( C \) of vectors \( \Delta D \) is given by Eq. (2), following the formulation in [4].

\[
C = \sum_{k=1}^{N} (\Delta D_k - \mu) (\Delta D_k - \mu)^T
\]

where \( \mu = \frac{1}{N} \sum_{k=1}^{N} \Delta D_k \) is the mean of all discriminant vectors. Similarly, for determining the covariance matrix for calculating eigenfaces, the discriminant vector in Eq. (2) is replaced by image intensity \( I_k \). The maximum generated signatures or features are denoted as \( M \). The traditional eigenface cannot work properly in SSPP condition because the first \( n \) generated eigenfaces does not only contain \( \Omega_I \) but also \( \Omega_E \) which could affect classification. Therefore we use supervised ANN with scaled conjugate gradient algorithm [27] and train it with populations of generic discriminant vectors \( \Delta D \) in order to discriminate between intra-subject faces and inter-subject faces using the a priori obtained from paired images in the generic set. Output neurons of ANN will produce higher output response \( \phi \) to intra-subject face pairs as opposed to inter-subject face pairs, i.e. when the input face pairs have the vector \( \Delta D_k \in \Omega_I \) and \( \Delta D_k \notin \Omega_E \). The sum of \( \phi \) is compared against a fixed threshold \( T \) which is pre-determined prior the training process to classify \( \Delta D \). The classification is carried out
as in Eq. (3) while the assigned ANN confidence score $\psi$ can be determined from Eq. (4). This classification process by assigning $\psi$ is based on image similarities matching (classification of intra-subject pair $P_I$ from inter-subject pair $P_E$).

\[
\Delta D_{k,i} \in \begin{cases} 
\Omega_I & \text{if } \sum \phi_{k,i} < T \\
\Omega_E & \text{if } \sum \phi_{k,i} \geq T 
\end{cases} \tag{3}
\]

\[
\psi_{k,i} = \sum \phi_{k,i} \tag{4}
\]

On the other hand, in the eigenface approach, new faces from probe test set is recognized by projecting them into the subspace spanned by the eigenfaces of gallery images and then classify the new faces by comparing their positions in eigenspace with the positions of faces from gallery images [4], which in this case we use the Euclidean distance. Then the eigenface confidence scores $\varphi$ and total confidence score $S$, can be formulated from Eq. (5) and Eq. (6) respectively.

\[
\varphi_{k,i} = \left( \frac{1}{d_{k,i}} \right), k = 1,2,...,N_P; i = 1,2,...,N_G \tag{5}
\]

\[
S_{k,i} = m\varphi_{k,i} + n\psi_{k,i} \tag{6}
\]

where $d_{k,i}$ is the Euclidean distance between features $f$ of probe image $k$ and gallery image $i$, where $f < M$. Voting coefficients $n$ and $m$ is used to manipulate the voting proportion between the scores of $\varphi$ and $\psi$. If $n = 0$ or $m \gg n$ the FR will be reverted back to only using Eigenface method. We set the scores in such a way that $\max(\varphi) = \max(\psi)$ and $\min(\varphi) = \min(\psi)$. The total number of confidence scores assigned to the gallery images, $\sum \varphi + \sum \psi$ is equal to $2(N_P N_G)$. The ranking is then assigned to each candidate image $i$ in gallery based on the probability of it bearing the same identity $P$ to a single image $k$ in probe set given by the total confidence score $S$. Higher total confidence score indicates higher probability thus will result in higher rank of a gallery image. The voting process is illustrated in Fig.2.

Fig.2. Process of voting top-match gallery images according to confidence scores

4. Experiment Setup

For experiments carried out in this paper, we use FERET and Extended Yale Face Database B (YALE) [28] datasets and have the images in these datasets partitioned into a few smaller sets that suit the proposed method. All the images are passed to the face detector based on Haar Cascades [29], to crop the face images. We did not implement any type of further dataset manipulation such as masking or normalization based on eye/nose/mouth positions in order to illustrate the capability of bias knowledge learned by ANN from generic samples. The FERET database is then partitioned into 7 generic sets for training and 3 probe sets for testing. Among the 7 generic sets, FERET Set A-E contain intra-subject and inter-subject training pairs of only frontal images, with several variations such as illumination, facial expressions, scales and time of photo taken between them (duplicates photo – dup1 and dup2 [6]). The only difference between them is the number of images in each set. Meanwhile, the other two, FERET Set E and F contain different number of training images that covers a wider range of variations and also include diverse poses. The Probe Set frontal contains only frontal images exhibiting similar variations to FERET Set A-E, while Probe Set mixed A and mixed B are much more complex probe sets than frontal, containing not only images with variations exhibited in Probe Set frontal but also images with different poses. The difference between Probe Set mixed A and mixed B is the number of images in each set.

Meanwhile the YALE dataset is partitioned into YALE Set and Probe Set YALE, containing images affected by heavy illumination variations from different angle of light sources and also exhibiting different poses. In order to form a gallery set for each probe set, for each person in probe sets, we randomly selected one frontal image and put them in the gallery set thus automatically making the number of images in the matching gallery set equal to the number of identities in probe sets $P_p$. The parameters describing all the generic and probe sets used in the experiment such as variations types and
numbers of images in each set are summarized in Table 1. The number of facial features, \( f \) used in all experiments for PCA projection follows our finding that optimal result obtained when \( f = 300 \).

### Table 1. Parameters of FERET and YALE face image sets used in the experiment

| Image Set | Type of Variations Exhibited | \( \Sigma \mathcal{P} \) | Number of Images | Intra-subject Pairs | Inter-subject Pairs |
|-----------|-----------------------------|----------------|----------------|-------------------|-------------------|
| FERET Set A | Illumination, Expression, Time, Scales | 699 | 2090 | 1391 | 2782 |
| FERET Set B | Illumination, Expression, Time, Scales | 550 | 1684 | 1135 | 2270 |
| FERET Set C | Illumination, Expression, Time, Scales | 480 | 1219 | 740 | 1480 |
| FERET Set D | Illumination, Expression, Time, Scales | 330 | 838 | 509 | 1018 |
| FERET Set E | Illumination, Expression, Time, Scales | 100 | 295 | 195 | 390 |
| FERET Set F | Illumination, Expression, Time, Scales, Poses | 700 | 3934 | 3234 | 6468 |
| FERET Set G | Illumination, Expression, Time, Scales, Poses | 500 | 2245 | 1745 | 3490 |
| YALE Set | Diff. light sources direction, Heavy Illumination, Poses | 20 | 1261 | 1241 | 2482 |
| Probe Set frontal | Illumination, Expression, Time, Scales | 200 | 239 | 239 | 712 |
| Probe Set mixed A | Illumination, Expression, Time, Scales, Poses | 200 | 1578 | 1578 | 4716 |
| Probe Set mixed B | Illumination, Expression, Time, Scales, Poses | 200 | 774 | 774 | 2309 |
| Probe Set YALE | Diff. light sources direction, Heavy Illumination, Poses | 18 | 1134 | 1134 | 3213 |

5. Results and Discussion

The performance of the proposed method is compared to the Eigenface (PCA) method [4], ANN [27], the (PC)\(^2\)A method as proposed in [20], SVM [5] and Fisherface [3]. Before we do the performance comparison, first we illustrate the effect of generic sample selection on the matching of image pairs based on exhibited variations by measuring the correct classification of intra-subject pairs \( P_I \) and inter-subject pairs \( P_E \) as well as the face recognition performance in all probes sets. As shown in Table 2, for Probe Set frontal, the almost all generic sets give more than 0.9 correct classification measured in True Acceptance Rate (TAR) on intra-subject pairs \( P_I \) and inter-subject pairs \( P_E \) but FERET Set A gives the best performance in face recognition, measured in rank-3 match. Since the variations exhibited in Probe Set frontal is also subset of variations in generic sets FERET Set F and G, the result met our expectation that both generic sets gives similar performance with FERET Set A on face recognition. In terms of effect of number of images in generic sets, since FERET Set A possessed a higher number of images, it gives slightly higher recognition rate than FERET Set B, C, D and E. Still, it is interesting to note that, FERET Set E which contains a much smaller number of images is able to a give a good performance on face recognition. It demonstrates that the performance does not only dependent on the number of generic images, but also on the representativeness of the variations in generic sets to the variations in probe set used. Another significant finding is that even though YALE Set is entirely a different set from Probe Set frontal, it does contain some variations exhibited in that probe set thus it is able to give quite a good result.

### Table 2. Rate of classification of intra-subject pair \( P_I \) and inter-subject pair \( P_E \), and Rank-3 FR rate for FERET Probe Set frontal using different generic sets

| Results | FERET Set A | FERET Set B | FERET Set C | FERET Set D | FERET Set E | FERET Set F | FERET Set G | YALE Set |
|---------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|----------|
| \( P_I \) TAR | 0.9540 | 0.9414 | 0.9623 | 0.9079 | 0.9582 | 0.9707 | 0.9665 | 0.9247 |
| \( P_E \) TAR | 0.9031 | 0.9143 | 0.8834 | 0.9256 | 0.8933 | 0.9284 | 0.9087 | 0.7879 |
| Rank-3 FR | 0.8912 | 0.8870 | 0.8786 | 0.8901 | 0.8876 | 0.8912 | 0.8912 | 0.8219 |

The previous statement also proves another point which is when the discrimination in generic sets is not representative enough, the expressiveness in eigenface method could correct the final face recognition result via the voting process. This effect of representativeness of generic set is further proven in experiment using more challenging probes sets as shown in Table 3.
Table 3. Rate of classification of intra-subject pair $P_I$ and inter-subject pair $P_E$, and rank-3 FR rate for FERET Probe Set mixed A, FERET Probe Set mixed B and Probe Set YALE using different generic sets

| Generic Set | Probe Set mixed A | Probe Set mixed B | Probe Set YALE |
|-------------|------------------|------------------|----------------|
|             | $P_I$ TAR | $P_E$ TAR | Rank-3 FR | $P_I$ TAR | $P_E$ TAR | Rank-3 FR | $P_I$ TAR | $P_E$ TAR | Rank-3 FR |
| FERET Set A | 0.5337   | 0.9156   | 0.4637   | 0.5124   | 0.9597   | 0.4222   | 0.2787   | 0.9925   | 0.6084   |
| FERET Set F | 0.7713   | 0.8472   | 0.4873   | 0.8353   | 0.9115   | 0.4705   | 0.3448   | 0.9595   | 0.6199   |
| FERET Set G | 0.7135   | 0.8579   | 0.4860   | 0.8248   | 0.8760   | 0.4601   | 0.3377   | 0.9639   | 0.6225   |
| YALE Set    | 0.6353   | 0.6979   | 0.4104   | 0.6719   | 0.7984   | 0.3699   | 0.7716   | 0.8522   | 0.7663   |

According to Fig. 3, for Probe Set mixed A and Probe Set mixed B the best classification of $P_I$ and $P_E$ is obtained from using FERET Set F, which despite exhibiting similar variations as in FERET Set G, it has a slightly higher number of images. Meanwhile, for FERET Set A and YALE set which does not contain all the variations exhibited in Probe Set mixed A and Probe Set mixed B, the results are lower. Additionally, it is also noted that for Probe Set YALE, the best result is obtained from generic YALE set since the other generic sets do not contain the variations exhibited in Probe Set YALE.

Fig.3. Result of face recognition using our proposed method as compared to several state-of-the-art methods when tested using the four probes sets (a) Probe Set frontal, (b) Probe Set mixed A, (c) Probe Set mixed B and (d) Probe Set YALE. Voting coefficients $m = 1$ and $n = 0.5$ are used for all probes sets.
Using the best performing generic samples, we compare the face recognition performance with other methods. As the result in Fig.3 implies, our prosed method outperforms other tested methods in all probes sets. In the test using rather simple probe, the Probe Set frontal, our method slightly outperforms other methods especially in earlier rank match, including Rank-1 match. In fact, for all four probes sets, our method gives the best Rank-1 match as compared to other methods tested.

For instance we can see in Fig.3 that as the complexity of the probe set grows, the difference in performance becomes greater, where in Probe Set mixed A, starting after Rank-15, our method is constantly 5%-6% better than Eigenface and SVM method, while it is more than 20% better than (PC)²A and Fisherface. In Probe Set mixed B, starting from Rank-10, our method gives around 15% better performance than Eigenface and SVM, and 25% more than (PC)²A and Fisherface. The same case also observed in Probe Set YALE, where in Rank-1, the result is 7% higher than Eigenface, and the difference in performance increases rapidly as the rank increases.

Eigenface method is proven to be able to project faces for reconstruction from a low dimensional basis but discriminatively that is not optimal. Thus the discrimination yielded beforehand by the discriminant vector is the main factor of the good performance produced by this method. The main advantage of this method as opposed to other methods, especially the one that does not use generic learning is that there is no need for online estimation of any additional vectors or parameters from the probe and gallery set. The features from probe and gallery set do not modify the bias knowledge and are used only when obtaining the discriminant vectors during the actual recognition phase, i.e. the system only need to have the knowledge on the generic sets and then adapt the biased knowledge on the collected test images. However, as shown by the results, the face recognition performance is highly dependent on the representativeness of variations in the generic sets. For a good result, the generic set should represent most of the variations exhibited by the test images. Thus, it is important to define the generic set carefully according to the environment where the test images are collected. As the training of the generic set is done off-line, it is possible to obtain a personalized and representative generic set tailored to the need of test environment. Also as an added flexibility to this method is there is explicitly no need for recalculation of discriminant vectors even though new samples are added in gallery set, given that the new samples are collected in the similar environment with the current samples.

It is also important to mention that we find it imperative to investigate the relationship between generic sample size and large-scale probe set, i.e. the number of generic samples needed to sufficiently accommodate variations of a specific large number of test images in multiple environments. In future, we will also try to produce a comprehensive experimentation on the relationship of the voting coefficients and the performance of the face recognition in a wider variations scale. This combination or fusion method also should be explored further to determine its viability with other methods other than eigenface alone.

6. Conclusion

In this paper we introduced a face recognition method based on fusion of ANN and eigenface by learning the discriminant vectors from generic sets when only a single sample per person is available. The discriminant vectors learnt can be used to decide whether a pair of facial images exhibit intra-subject or inter-subject variations. The resulting classification is further enhanced by eigenface method based on voting process. It has been shown by experimentation using FERET and YALE datasets that the proposed method performs better than the traditional eigenface method, as well as several other state-of-the-art methods in single sample per person scenario.

Acknowledgements

Portions of the research in this paper use the FERET database of facial images collected under the FERET program. The authors would like to thank the FERET Technical Agent, the US National Institute of Standards and Technology (NIST) for providing the FERET database. Some portions of the research in this paper also use the Extended Yale Face Database B.

References

[1] A. K. Jain, "Technology: Biometric Recognition," Nature, vol. 449, pp. 38-40, 2007.
[2] S. Z. Li, Ed., Encyclopedia of Biometrics. New York: Springer Science+Business Media, 2009.
[3] P. Belhumeur, J. Hespanha, and D. Kriegman, "Eigenfaces vs. Fisherfaces: recognition using class specific linear projection," IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 19, pp. 711-720, 1997.
[4] M. Turk and A. Pentland, "Eigenfaces for recognition," Journal of Cognitive Neuroscience, vol. 3, pp. 71-86, 1991.
[5] C. Chih-Chung and L. Chih-Jen, "LIBSVM: A library for support vector machines," ACM Transactions on Intelligent Systems and Technology, vol. 2, pp. 27:1-27:27, 2011.
[6] P. J. Phillips, M. Hyeonjoon, S. A. Rizvi, and P. J. Rauss, "The FERET evaluation methodology for face-recognition algorithms," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 22, pp. 1090-1104, 2000.

[7] D. Beymer and T. Poggio, "Face Recognition From One Example View," in *Computer Vision, 1995. Proceedings., Fifth International Conference on*, Cambridge, MA, USA, 1995, pp. 500-507.

[8] T. Poggio and T. Vetter, "Recognition and structure from one 2D model view: Observations on prototypes, object classes and symmetries," MIT1992.

[9] T. Xiaoyang, C. Songcan, and Z. F. Z. Zhi-hua, "Face Recognition from a Single Image per Person: A Survey," *Pattern Recognition*, vol. 39, pp. 1725-1745, 2006.

[10] A. M. Martinez, "Recognizing imprecisely localized, partially occluded, and expression variant faces from a single sample per class," *IEEE Transactions on Machine Intelligence*, vol. 24, pp. 748-763, 2002.

[11] A. M. Martinez, "Matching expression variant faces," *Vision Research*, vol. 43, pp. 1047-1060, 2003.

[12] A. M. Martinez, "Recognizing expression variant faces from a single sample image per class," in *Computer Vision and Pattern Recognition, 2003. Proceedings. 2003 IEEE Computer Society Conference on*, 2003, pp. I-353-I-358 vol.1.

[13] X. Tan, S. C. Chen, Z.-H. Zhou, and F. Z. F, "Recognizing partially occluded, expression variant faces from single training image per person with SOM and soft kNN ensemble," *IEEE Transactions on Neural Networks*, vol. 16, pp. 875-886, 2005.

[14] T. Shan, B. C. Lovell, and S. Chen, "Face Recognition Robust to Head Pose from One Sample Image," presented at the Proceedings of the 18th International Conference on Pattern Recognition (ICPR'06), 2006.

[15] T. Shan, S. Chen, C. Sanderson, and B. C. Lovell, "Towards robust face recognition for Intelligent-CCTV based surveillance using one gallery image " presented at the Advanced Video and Signal Based Surveillance, 2007. AVSS 2007. IEEE Conference on London, 2007.

[16] J. Wang, K. N. Plataniotis, and A. N. Venetsanopoulos, "Selecting discriminant eigenfaces for face recognition," *Pattern Recognition Letters*, vol. 26, pp. 1470–1482, 2005.

[17] S. C. Chen, J. Liu, and Z.-H. Zhou, "Making FLDA applicable to face recognition with one sample per person," *Pattern Recognition*, vol. 37, pp. 1553 – 1555, 2004.

[18] H. Jian, P. C. Yuen, C. Wen-Sheng, and J. H. Lai, "Component-based LDA method for face recognition with one training sample," in *Analysis and Modeling of Faces and Gestures, 2003. AMFG 2003. IEEE International Workshop on*, 2003, pp. 120-126.

[19] S. Chen, D. Zhang, and Z.-H. Zhou, "Enhanced (PC)2A for face recognition with one training image per person," *Pattern Recognition Letters*, vol. 25, pp. 1173-1181, 2004.

[20] J. Wu and Z.-H. Zhou, "Face Recognition with one training image per person," *Pattern Recognition Letters*, vol. 23, pp. 1711-1719, 2002.

[21] J. Lu, K. N. Plataniotis, and A. N. Venetsanopoulos, "Face recognition using LDA-based algorithms," *IEEE Trans. Neural Networks*, vol. 14, pp. 195–200, 2003.

[22] S. Yu, S. Shiguang, C. Xilin, and G. Wen, "Adaptive generic learning for face recognition from a single sample per person," in *Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on*, 2010, pp. 2699-2706.

[23] S. Thrun, "Learning To Learn: Introduction," in *Learning to learn*, ed: Kluwer Academic Publishers, 1996.

[24] Y. Moses, S. Ullman, and S. Edelman, "Generalization to novel images in upright and inverted faces," *Perception*, vol. 25, pp. 443-462, 1994.

[25] E. G. Miller, N. E. Matsakis, and P. A. Viola, "Learning from One Example Through Shared Densities on Transforms," presented at the Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2000.

[26] B. Moghaddam, W. Wahid, and A. Pentland, "Beyond Eigenfaces: Probabilistic Matching for Face Recognition," presented at the The 3rd IEEE Int1 Conference on Automatic Face & Gesture Recognition, Nara, Japan, 1999.

[27] M. F. Møller, "A scaled conjugate gradient algorithm for fast supervised learning," *NEURAL NETWORKS*, vol. 6, pp. 525–533, 1993.

[28] A. S. Georgiades, P. N. Belhumeur, and D. J. Kriegman, "From few to many: illumination cone models for face recognition under variable lighting and pose," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 23, pp. 643-660, 2001.

[29] P. Viola and M. Jones, "Rapid object detection using a boosted cascade of simple features," in *Computer Vision and Pattern Recognition, 2001. CVPR 2001. Proceedings of the 2001 IEEE Computer Society Conference on*, 2001, pp. I-511-I-518 vol.1.