WHOLE SPACE SUBCLASS DISCRIMINANT ANALYSIS FOR FACE RECOGNITION

Bappaditya Mandal, Liyuan Li, Vijay Chandrasekhar and Joo Hwee Lim

Email: {bmandal, lyli, vijay, joohwee}@i2r.a-star.edu.sg
Institute for Infocomm Research, A*STAR, Singapore

ABSTRACT

In this work, we propose to divide each class (a person) into subclasses using spatial partition trees which helps in better capturing the intra-personal variances arising from the appearances of the same individual. We perform a comprehensive analysis on within-class and within-subclass eigenspectrums of face images and propose a novel method of eigenspectrum modeling which extracts discriminative features of faces from both within-subclass and total or between-subclass scatter matrices. Effective low-dimensional face discriminative features are extracted for face recognition (FR) after performing discriminant evaluation in the entire eigenspace. Experimental results on popular face databases (AR, FERET) and the challenging unconstrained YouTube Face database show the superiority of our proposed approach on all three databases.

Index Terms— Feature extraction, discriminant analysis, subspace learning, face recognition.

1. INTRODUCTION

In multi-class classification, Fisher-Rao's linear discriminant analysis (LDA) minimizes Bayesian error when sample vectors of each class are generated from multivariate Normal distributions of same covariance matrix but different means (homoscedastic data [1]). However, for real-world face images, the classes have random (or heteroscedastic [1]) distributions, the variances are quite large and the data are of very high dimensionality. So the estimation of variances (using within-class and between-class scatter matrices in LDA) are limited to averages of all the possible variations among training samples. This limits the usages of LDA in face image data for FR, where large number of classes with high dimensional data are involved [2, 3]. Mixture discriminant analysis (MDA) [4] models each class as a mixture of Gaussians, rather than a single Gaussian as in LDA. MDA uses a clustering procedure to find subclass partitions of the data and then incorporate this information into the LDA criterion.

Subclass discriminant analysis (SDA) in [5] maximizes the distances between both the means of classes and the means of the subclasses. SDA emphasizes the role of class separability rather than the discriminant information in the within-subclass scatter matrix, hence it may not capture the crucial discriminant information in the within-subclass variances for FR [6]. Mixture subclass discriminant analysis (MSDA), an improvement over SDA is presented in [7]. In this approach, a subclass partitioning procedure along with a non-Gaussian criterion are used to derive the subclass division that optimizes the MSDA criterion, this has been extended to fractional MSDA and kernel MSDA in [8].

All the above approaches discard the null space of either within-class and within-subclass scatter matrices, which plays a very crucial role in the discriminant analysis of faces [6, 3, 2, 9]. Xudong *et al.* [10, 11] proposed an eigenfeature regularization method (ERE) which partitions the eigenspace into various subspaces. However, their variances are extracted from within-class scatter matrix and does not consider partitions within each class. Hence, this method would fail in capturing the crucial within-subclass discriminant information.

Another class of emerging algorithms is the deep learning which uses convolutional neural network and millions of (external) face images for training and obtain very high accuracy rates [12, 13]. However, our proposed method is still attractive because it uses small number of training samples and does not use any external training data but can achieve comparable performances.

2. DISCRIMINANT ANALYSIS

2.1. Scatter matrices and discriminant analysis

The problem of discriminant analysis is generally solved by maximization of the Fisher criterion [1]. This involves between-class (S_b) and within-class ($S_w = \frac{1}{n} \sum_{i=1}^{C} \sum_{j=1}^{n_i} (x_{ij} - \mu_i)(x_{ij} - \mu_i)^T$) scatter matrices, where C is the number of classes or persons, μ_i is the sample mean of class i, μ is the global mean, $x_{ij} \in \mathbb{R}^l$, by lexicographic ordering the pixel elements of image of size $l = width \times height$, is the jth sample of class i, n_i is the number of samples in i^{th} class and $n = \sum_{i=1}^{C} n_i$ is the total number of samples.

LDA assumes that the class distributions are homoscedastic, which is rarely true in practice for FR. We assume that there exist subclass homoscedastic partitions of the data and model each class as mixtures of Gaussians subclasses, whose objective function is defined as $J(\Psi) = \frac{tr(\Psi^T S_{bs} \Psi)}{tr(\Psi^T S_{ws} \Psi)}$, where

tr represents trace of a matrix, Ψ denotes a transformation matrix, S_{bs} is the between-subclass scatter matrix and S_{ws} is the within-subclass scatter matrix defined as

$$S_{ws} = \sum_{i=1}^{C} p_i \sum_{j=1}^{H_i} \frac{q_{H_i}}{G_{ij}} \sum_{k=1}^{G_{ij}} (x_{ijk} - \mu_{ij})(x_{ijk} - \mu_{ij})^T. \quad (1)$$

H_i denotes the number of subclasses of the ith class and G_{ij} denotes the number of samples in j^{th} subclass of ith class. $x_{ijk} \in \mathbb{R}^l$ is the kth image vector in j^{th} subclass of ith class. $\mu_{ij} = \frac{1}{G_{ij}} \sum_{k=1}^{G_{ij}} x_{ijk}$ is the sample mean of jth subclass of the ith class. p_i and q_{H_i} are the estimated prior probabilities. If we assume that each class and subclasses have equal prior probabilities then $p_i = \frac{1}{C}$ and $q_{H_i} = \frac{1}{H_i}$. If S_{ws} is nonsingular, the optimal projection vectors Ψ is chosen as the matrix with orthonormal columns which maximizes the ratio of the determinant of the between-subclass matrix of the projected samples to the determinant of the within-subclass scatter of the projected samples.

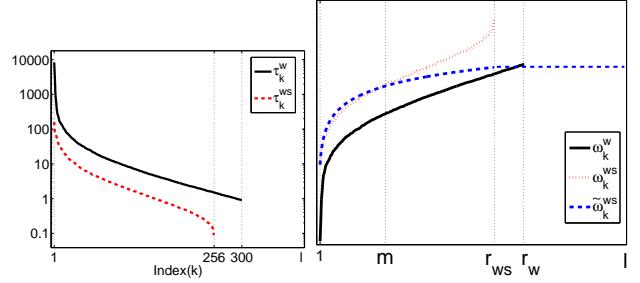## 2.2. Partitioning of a face class into subclasses

We investigate various popular spatial partition trees to partition each face class into subclasses for within-subclass discriminant analysis, [14, 15]: (i) k-d tree, (ii) RP tree, (iii) PCA tree, and (iv) k-means tree. k-d trees and RP trees are built by recursive binary splits. They differ only in the nature of the split. Unlike k-d tree, RP tree adapts to intrinsic low dimensional structure without having to explicitly learn face structure. In PCA tree, the partition axis is obtained by computing the principal eigenvector of the covariance matrix of the face image data. Since face image appear very differently under various contexts this kind of partitioning would be advantageous for data that are heterogeneously distributed in all dimensions. k-means tree is built based on nearest neighbor (NN) clustering of face appearances. In our experiments, we use the implementations of spatial partitioning trees by Freund *et al.* [14], with their default parameters of maximum depth up to eight layers and no overlap in samples splitting.

2.3. Regularization of within-subclass eigenspace

2.3.1. Divisions in within-subclass eigenspace

Formation of subclasses using spatial partition tree helps in capturing the variances more closely in appearances of the same individual. We compute the eigenvectors $\Psi^{ws} = \{\psi_1^{ws}, \ldots, \psi_l^{ws}\}$ corresponding to the eigenvalues $\Lambda^{ws} = \{\lambda_1^{ws}, \ldots, \lambda_l^{ws}\}$ of S_{ws} described by (1), where the eigenvalues are sorted in descending order.

A typical plot of $\tau_k^w = \sqrt{\lambda_k^w}$ and $\tau_k^{ws} = \sqrt{\lambda_k^{ws}}$ computed from within-class (Λ^w) and within-subclass (Λ^{ws}) scatter matrices on real face training images are shown in Fig. 1 left. The decay of the eigenvalues of S_{ws} is much faster than S_w.

Fig. 1. Left, A typical real eigenspectrum computed from within-class and within-subclass sorted in descending order. Right, Inverse of a typical real eigenspectrum computed from within-class and within-subclass and weighting function $\tilde{\omega}_k^{ws}$.

This is because within-class matrices capture variances that arise in the face appearances of the same individual. Forming subclasses, further lower their values. Their differences are of the order 10^2 for larger and smaller ones. In both curves, as pointed out in [16, 17], the characteristic bias is most pronounced when the population eigenvalues tend toward equality, and it is correspondingly less severe when their values are highly disparate. Therefore, the smallest eigenvalues are biased much more than the largest ones [16, 10, 17].

The whitened eigenvector matrix $\overline{\Psi}^{ws} = \{\psi_1^{ws}/\tau_1^{ws}, \ldots, \psi_l^{ws}/\tau_l^{ws}\}$, $\tau_k^{ws} = \sqrt{\lambda_k^{ws}}$ in Fig. 1 left, is used to project the image vector x_{ij} before constructing the between-subclass scatter matrix. This is equivalent to image vector x_{ij} first transformed by the eigenvector $y_{ij} = \Psi^{ws T} x_{ij}$, and then multiplied by a scaling function $\omega_k^{ws} = 1/\tau_k^{ws}$ (whitening process). Truncating dimensions is equivalent to set $\omega_k^{ws} = 0$ for these dimensions as done in Fisherface and many other variants of LDA [18, 2, 3]. The scaling function is thus

$$\omega_k^{ws} = \begin{cases} 1/\sqrt{\lambda_k^{ws}}, & k \leq r_{ws} \\ 0, & r_{ws} < k \leq l \end{cases}, \quad (2)$$

where $r_{ws} \leq min(l, \sum_{i=1}^{C} \sum_{j=1}^{H_i} (G_{ij} - 1))$, is the rank of S_{ws}. Similar scaling function (ω_k^w) with $r_w \leq min(l, n - C)$ are also applicable for S_w scatter matrix.

The inverses of τ_k^{ws} and τ_k^w pose two problems as shown in Fig. 1 right. Firstly, the eigenvectors corresponding to the zero eigenvalues are discarded or truncated as the features in the null space are weighted by a constant zero. This leads to the loses of important discriminative information that lies in the null space [2, 6, 19, 3]. Secondly, when the inverse of the square of the eigenvalues are used to scale the respective eigenvectors, features get undue weighage, noises get amplified and tend to over-fit the training samples.

2.3.2. Within-subclass eigenspectrum modeling

We use a median operator and its parameter similar to that used in [10, 20] to find the pivotal point m for decreasing the

decay of the eigenspectrum. A typical such m value of a real eigenspectrum is shown in Fig. 1 right. We use the function form $1/f$, similar to [21, 10], to estimate the eigenspectrum as $\tilde{\lambda}_k^{ws} = \frac{\alpha}{k+\beta}$, $1 \leq k \leq r_{ws}$, where α and β are two constants, used to model the real eigenspectrum in the initial portion. We determine α and β by letting $\tilde{\lambda}_1^{ws} = \lambda_1^{ws}$ and $\tilde{\lambda}_m^{ws} = \lambda_m^{ws}$, which yields $\alpha = \frac{\lambda_1^{ws}\lambda_m^{ws}(m-1)}{\lambda_1^{ws}-\lambda_m^{ws}}$, $\beta = \frac{m\lambda_m^{ws}-\lambda_1^{ws}}{\lambda_1^{ws}-\lambda_m^{ws}}$. Fig. 1 right, shows inverse of the square roots of a real eigenspectrum, $\omega_k^{ws} = 1/\tau_k^{ws}$, where $\tau_k^{ws} = \sqrt{\lambda_k^{ws}}$, and the stable portion of its model, $\tilde{\omega}_k^{ws} = 1/\tilde{\tau}_k^{ws}$, such that, $\tilde{\tau}_k^{ws} = \sqrt{\tilde{\lambda}_k^{ws}}$. We see that the model $\tilde{\omega}_k^{ws}$ fits closely to the real ω_k^{ws} in the reliable space but has slower decay in the unstable space.

2.3.3. Feature extraction

From Fig. 1 right, it is evident that noise component is small as compared to face components in range space but it is dominating in unstable region. Thus, the estimated eigenspectrum $\tilde{\lambda}_k^{ws}$ is given by

$$\tilde{\lambda}_k^{ws} = \begin{cases} \lambda_k^{ws}, & k < m \\ \frac{\alpha}{k+\beta}, & m \leq k \leq r_{ws} \\ \frac{\alpha}{r_{ws}+1+\beta}, & r_{ws} < k \leq l \end{cases} \quad (3)$$

The feature scaling function is then $\tilde{\omega}_k^{ws} = \frac{1}{\sqrt{\tilde{\lambda}_k^{ws}}}$, $k = 1, 2, ..., l$. Fig. 1 right, shows the proposed feature scaling function $\tilde{\omega}_k^{ws}$. Using this scaling function and the eigenvectors ψ_k^{ws}, training data are transformed to $\tilde{y}_{ij} = \tilde{\mathbf{\Psi}}_l^{wsT} x_{ij}$, where $\tilde{\mathbf{\Psi}}_l^{ws} = [\tilde{\omega}_k^{ws}\psi_k^{ws}]_{k=1}^l$. New between-subclass and total subclass scatter matrices are formed by vectors $\tilde{y}_{ij}$ of the transformed training data as

$$\begin{cases} \tilde{S}_{bs} = \sum_{i=1}^C \frac{p_i}{H_i} \sum_{j=1}^{H_i} (\tilde{\mu}_{ij} - \tilde{\mu})(\tilde{\mu}_{ij} - \tilde{\mu})^T, \\ \tilde{S}_{ts} = \sum_{i=1}^C \frac{p_i}{n_i} \sum_{j=1}^{n_i} (\tilde{y}_{ij} - \tilde{\mu})(\tilde{y}_{ij} - \tilde{\mu})^T, \end{cases} \quad (4)$$

where $\tilde{\mu}_{ij} = \frac{1}{G_{ij}} \sum_{k=1}^{G_{ij}} \tilde{y}_{ijk}$ and $\tilde{\mu} = \frac{1}{C} \sum_{i=1}^C \tilde{\mu}_i$, such that $\tilde{\mu}_i = \frac{1}{n_i} \sum_{j=1}^{n_i} \tilde{y}_{ij}$. In this work, we employ the total scatter matrix $\tilde{S}_{ts}$ of the regularized training data to extract the discriminative features because of its greater noise tolerance as compared to $\tilde{S}_{bs}$. The transformed features $\tilde{y}_{ij}$ will be decorrelated for $\tilde{S}_{ts}$ by solving the eigenvalue problem. Selecting the eigenvectors with the d largest eigenvalues, $\tilde{\mathbf{\Psi}}_d^{ts} = [\tilde{\psi}_k^{ts}]_{k=1}^d$, the proposed feature scaling and extraction matrix is given by $\mathbf{U} = \tilde{\mathbf{\Psi}}_l^{ws}\tilde{\mathbf{\Psi}}_d^{ts}$, which transforms a face image vector x, $x \in \mathbb{R}^l$, into a feature vector z, $z \in \mathbb{R}^d$, by $z = \mathbf{U}^T x$.
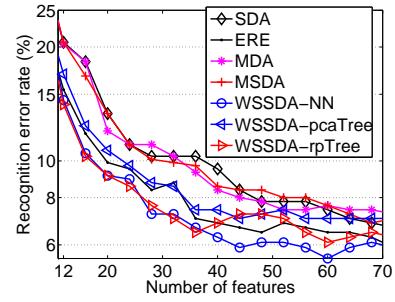
3. EXPERIMENTAL RESULTS

Three popular benchmark databases are used to evaluate our proposed approach of whole space subclass discriminant analysis (WSSDA) for FR. All images are normalized following the CSU Face Identification Evaluation System [22]. We

follow the rule described in [5] that every class is partitioned by the same number of subclasses h (equally balanced), such that $H_i = h, \forall i$. k-means tree is built based on nearest neighbor (NN) clustering of face appearances, abbreviated as WSSDA-NN. We divide each class into two subclasses for AR and FERET databases and each class into four subclasses for YouTube database. We test our approach using PCA and RP (random projection) decision trees [15], abbreviated as WSSDA-pcaTree and WSSDA-rpTree respectively. Cosine distance measure and the first nearest neighborhood classifier (1-NN) are applied to test the proposed WSSDA approach.

3.1. Results on AR database

In AR database [23], color images are converted to gray-scale and cropped into the size of 120×170 same as that in [23, 10]. Seventy-five subjects with 14 non-occluded images per subject are selected from the AR database. The first 7 images of 75 subjects are used in the training and also serve as gallery images. The second 7 images of the 75 subjects serve as probe images. Fig. 2 shows the recognition error rate on the test set against the number of features d used in the matching. Methods that discard the null space of the within-class or
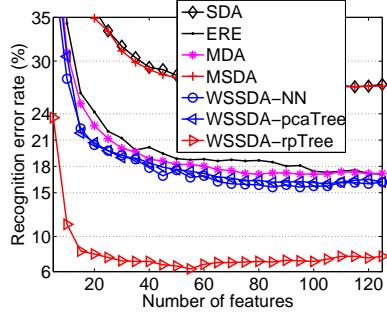

Fig. 2. Recognition error rate against the number of features used in the matching on the AR database of 525 training images (75 people) and 525 testing images (75 people).

within-subclass scatter matrices perform poorly as compared to the methods that utilize this valuable space for discriminant analysis. ERE approach evaluates the discriminant value in the whole space of S_w, so it performs better than all other approaches except our method. Partitioning each class into subclasses has helped in better capturing the variances arising from within-subclass scatter matrix. The proposed WSSDA approaches evaluate the discriminant value in the whole space of S_{ws}, hence they alleviate the over-fitting problem. Among them, WSSDA-NN outperform all other approaches across all different number of features.

3.2. Results on FERET database

Using this database, we test the performance of the proposed algorithm on how well it generalizes to new test datasets (with

new subjects), when trained with different set of subjects. It is constructed with normalized image size of 130×150, similar to one data set used in [24, 10, 25], by choosing 256 subjects with at least four images per subject. 512 images of the first 128 subjects are used for training and the remaining 512 images of another 128 subjects serve as testing images. There is no overlap in subjects between the training and testing sets. For each subject, the ith image is chosen to form the gallery set and the remaining 3 images serve as the probe images to be identified. Fig. 3 shows the average recognition error rates over the 4 probe sets, each of which has a different gallery set.
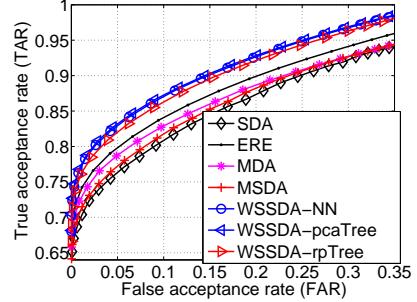

Fig. 3. Recognition error rate against the number of features used in the matching on the FERET database of 512 training images (128 people) and 512 testing images (128 people).

As described previously, methods like MSDA and SDA perform very badly because they discard the null space of the within-class scatter matrix and their eigenfeatures are inversely weighted by very small or near zero eigenvalues causing severe over-fitting problem. MDA outperforms ERE, SDA and MSDA approaches because it performs discriminant analysis using the scatter matrix arising from within-subclass variances. However, it discards the crucial null space of the within-subclass scatter matrix. The proposed WSSDA approaches (especially with rp-Tree partition) achieve consistently lowest recognition error rates for all number of features.

3.3. Results on YouTube database

YouTube faces (YTF) database [26] contains 3,425 videos of 1,595 different people under real-world scenarios and hence very challenging. The longest and shortest videos contain 48 and 6070 frames, respectively, with an average of 181.3 frames per video. We directly crop the image centered on the face according to the provided data [26] and then resize them into 40×24 which is similar to [27]. We use the same 5,000 video pairs from the YTF database described in [26] and follow their splits/partitioning schemes for restricted video face verification. Experimental results are presented in this work for each approach where the minimum equal error rate is obtained. Fig. 4 shows the average receiver operating characteristics (ROC) curves that plots the true acceptance rate (TAR)

Fig. 4. TAR against FAR on YouTube database [26].

against the false acceptance rate (FAR) following the 10-fold cross-validation pairwise tests protocol suggested for the YTF database [26]. It shows again that the proposed WSSDA approaches achieve higher TAR for their corresponding FAR among other tested approaches consistently for all different operating points.

For an accurate record, verification performance in terms of equal error rate (EER) obtained from the above experiments are shown in Table 1. From Fig. 4 and Table 1, we observe that the proposed WSSDA approaches, *i.e.* WSSDA-NN, WSSDA-pcaTree and WSSDA-rpTree approaches outperform recent results on YTF database and achieves lowest EER among all the compared approaches except the Deep-Face which uses an external data of 4.4 million face images for training [12].

Table 1. Equal Error Rate (EER %) of various approaches on YouTube face image database.

Method	ERR
MBGS [26]	25.3
APEM Fusion [28]	21.4
STFRD+PMML [27]	19.9
MMMRF Fusion [29]	12.6
DeepFace [12]	8.6

Method	ERR
SDA	14.0
ERE	12.8
MDA	13.1
MSDA	13.6
WSSDA-NN	**11.8**
WSSDA-pcaTree	**11.6**
WSSDA-rpTree	**12.2**

4. CONCLUSIONS

This paper addresses the problems of discriminant analysis using within-class and within-subclass scatter matrices for FR. Each class is divided into subclasses using spatial partition trees so as to approximate the underlying distribution with mixture of Gaussians and perform whole space subclass discriminant analysis among these subclasses. This work proposes a regularization methodology that enables discriminant analysis in the whole eigenspace of the within-subclass scatter matrix. Low dimensional face discriminative features are extracted after performing discriminant evaluation in the entire eigenspace of within-subclass scatter matrix. Experimental results on popular databases, AR, FERET and challenging unconstrained YouTube face database show the superiority of our proposed approach on all three databases.

5. REFERENCES

[1] R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern classification*, John Wiley and Sons, New York, 2001.

[2] H. Cevikalp, M. Neamtu, M. Wilkes, and A. Barkana, "Discriminative common vectors for face recognition," *IEEE PAMI*, vol. 27, no. 1, pp. 4–13, January 2005.

[3] L. F. Chen, H. Y. M. Liao, M. T. Ko, J. C. Lin, and G. J. Yu, "A new lda-based face recognition system which can solve the small sample size problem," *Pattern Recognition*, vol. 33, no. 10, pp. 1713–1726, 2000.

[4] T. Hastie and R. Tibshirani, "Discriminant analysis by gaussian mixtures," *Journal of the Royal Statistical Society, Series B*, vol. 58, 1996.

[5] M. Zhu and A. Martinez, "Subclass discriminant analysis," *IEEE PAMI*, vol. 28, no. 8, pp. 1274–1286, 2006.

[6] W. Liu, Y. Wang, S. Z. Li, and T. N. Tan, "Null space approach of fisher discriminant analysis for face recognition," in *ECCV*, 2004, pp. 32–44.

[7] N. Gkalelis *et al.*, "Mixture subclass discriminant analysis," *IEEE SPL.*, vol. 18, no. 5, pp. 319–322, 2011.

[8] N. Gkalelis, V. Mezaris, I. Kompatsiaris, and T. Stathaki, "Mixture subclass discriminant analysis link to restricted gaussian model and other generalizations," *IEEE TNNLS.*, vol. 24, no. 1, pp. 8–21, 2013.

[9] B. Mandal, X. Jiang, HL. Eng, and A. Kot, "Prediction of eigenvalues and regularization of eigenfeatures for human face verification," *Pattern Recognition Letters*, vol. 31, no. 8, pp. 717–724, 2010.

[10] X. D. Jiang, B. Mandal, and A. Kot, "Eigenfeature regularization and extraction in face recognition," *IEEE PAMI*, vol. 30, no. 3, pp. 383–394, March 2008.

[11] X. D. Jiang, B. Mandal, and A. Kot, "Face recognition based on discriminant evaluation in the whole space," in *IEEE 32nd ICASSP*, Hawaii, April 2007, pp. 245–248.

[12] Taigman *et al.*, "Deepface: Closing the gap to human-level performance in face verification," in *CVPR*, Columbus, OH, Jun 2014, pp. 1701–1708.

[13] Y. Sun, X. Wang, and X. Tang, "Deep learning face representation from predicting 10,000 classes," in *CVPR*, 2014, pp. 1891–1898.

[14] Freund *et al.*, "Learning the structure of manifolds using random projections.," in *NIPS*, 2007, vol. 7, p. 59.

[15] J. Wang *et al.*, "Trinary-projection trees for approximate nearest neighbor search," *IEEE PAMI*, vol. 36, no. 2, pp. 388–403, 2014.

[16] J. H. Friedman, "Regularized discriminant analysis," *Journal of the American Statistical Association*, vol. 84, no. 405, pp. 165–175, March 1989.

[17] B. Mandal and H-L. Eng, "Regularized discriminant analysis for holistic human activity recognition," *IEEE Intelligent Systems*, vol. 27, no. 1, pp. 21–31, 2012.

[18] X. Zhuang and D. Dai, "Improved discriminant analysis for high-dimensional data and its application to face recognition," *PR*, vol. 40, pp. 1570–1578, 2007.

[19] B. Mandal, X. D. Jiang, and A. Kot, "Dimensionality reduction in subspace face recognition," in *IEEE ICICS*, Dec 2007.

[20] B. Mandal, X. D. Jiang, and A. Kot, "Verification of human faces using predicted eigenvalues," in *19th ICPR*, Tempa, Florida, USA, Dec 2008.

[21] B. Moghaddam, "Principal manifolds and probabilistic subspace for visual recognition," *IEEE PAMI*, vol. 24, no. 6, pp. 780–788, June 2002.

[22] R. Beveridge, D. Bolme, M. Teixeira, and B. Draper, "The csu face identification evaluation system users guide 2013: Version 5.0," *Technical Report: http://www.cs.colostate.edu/evalfacerec/data/normalization.html*.

[23] A. M. Martinez, "Recognizing imprecisely localized, partially occluded, and expression variant faces from a single sample per class," *IEEE PAMI*, vol. 24, no. 6, pp. 748–763, June 2002.

[24] J. Lu *et al.*, "Ensemble-based discriminant learning with boosting for face recognition," *IEEE TNN*, vol. 17, no. 1, pp. 166–178, 2006.

[25] X. D. Jiang, B. Mandal, and A. Kot, "Complete discriminant evaluation and feature extraction in kernel space for face recognition," *Machine Vision and Applications, Springer*, vol. 20, no. 1, pp. 35–46, Jan 2009.

[26] L. Wolf, T. Hassner, and I. Maoz, "Face recognition in unconstrained video with matched background similarity," in *IEEE CVPR*, Jun 2011, pp. 529–534.

[27] Z. Cui *et al.*, "Fusing robust face region descriptors via multiple metric learning for face recognition in the wild," in *IEEE CVPR*, Jun 2013.

[28] H. Li, G. Hua, Z. Lin, J. Brandt, and J. Yang, "Probabilistic elastic matching for pose variant face verification," in *IEEE CVPR*, Jun 2013.

[29] L. Rowden, B. Klare, J. Klontz, and A. K. Jain, "Video-to-video face matching: Establishing a baseline for unconstrained face recognition," in *IEEE BTAS*, Oct 2013.