A New Face Recognition Algorithm based on Dictionary Learning for a Single Training Sample per Person

Yang Liu
yl504@cam.ac.uk
Ian Wassell
ijw24@cam.ac.uk
Computer Laboratory,
University of Cambridge

Abstract

The number of the training samples per person has a significant impact on face recognition (FR) performance. For the single training sample per person (STSPP) problem, most traditional FR algorithms exhibit performance degradation owing to the limited information available to predict the variance of the query sample. This paper proposes a new method for the STSPP problem in FR, namely the Learn-Generate-Classify (LGC) method. The LGC method first learns the relationship between the multiple images of a subject based on dictionary learning from a generic training set. Then it predicts the intra-class variance of the gallery set using the learned relationship. Based on the predicted information, synthetic samples can be generated, thus extending the single sample gallery set to one having multiple samples. Finally, we can classify the query samples using the traditional sparse representation classification (SRC) framework on the multi-sample gallery set. We verified the effectiveness of the new LGC method on the CMU Multi-pie database, with different illumination, expression and pose variation factors. It shows that the LGC method demonstrates a promising FR performance with only a STSPP.

1 Introduction

In the field of computer vision and bioinformatics, FR has captured considerable attention from both academia and industry [22]. In many practical scenarios (e.g., e-passports, driving licences), there may be as few as a single sample image per person, which degrades the recognition performance dramatically owing to the limited information available to predict the variance of the query sample. It has been one of the biggest challenges in FR to achive a robust recognition performance in the single training sample per person (STSPP) scenario.

In the STSPP problem, due to the curse of dimensionality, many traditional recognition methods cannot be applied directly. For the STSPP problem, the sparse representation classification (SRC) method has been shown to be ineffective, because it is excessively sensitive to the quality and quantity of training samples [20]. In the literature, various methods have been proposed to deal with the STSPP problem [15], which can be divided into two categories, i.e., those with and those without learning from a generic training set.

Pages 69.1-69.11
DOI: https://dx.doi.org/10.5244/C.29.69

The methods without generic learning include robust feature extraction, synthetic sample generation and local block methods. For discriminant feature extraction, the common features employed are the histogram of gradients (HOG) [8], the scale-invariant feature transform (SIFT) [11] and the local binary patterns (LBP) [2]. To improve the performance of the STSPP, the generation of synthetic images has been propsed [6]. Common methods for synthetic sample generation are based on singular value decomposition (SVD) [21] or the geometry transform[6]. There are also some local block methods, such as local patch based LDA [3] and multi manifold learning from local patches [12]. Although these methods improve recognition rate for the STSPP problem, they are still quite sensitive to extreme illumination, expression and pose variations.

In generic learning methods, a generic training set is introduced, which includes a number of sample images per subject but is totally separate from the single sample per subject training gallery set. Generic learning methods outperform the other approaches since the discriminant features are learned from the generic training set. For example, Su et al propose the Adaptive generic learning (AGL) [14] method which applies the within and between class scatter matrices computed from a generic training set to the gallery set. Deng et al propose the extended SRC method [7], which computes an intra-class variance matrix from the generic training set and then applies this intra-class variant dictionary to represent the possible variation between the query and training samples in the gallery set.

Although some improvements have been made, these methods rarely work effectively for extreme variations, especially, for pose. In this chapter, we propose a new framework to deal with the STSPP problem, known as the learn-generate-classify (LGC) method. The overall framework of the LGC method is presented in Fig.1. In this framework, three sets in a STSPP classification task are defined in advance, namely the generic training set, the gallery training set and the test set. The generic training set is composed of multiple sample images per subject, that are distinct from the subjects in the gallery training and test sets. The identities in the gallery training set and the test set are the same. However, note that in a STSPP problem, the gallery training set consists of only one sample per subject. The test set is normally chosen to have multiple sample images for each subject contained in the gallery training set, i.e., there is variability between the sample images for each subject. The LGC method first learns the intra-class variation between the different images of a subject based on dictionary learning from a generic training set. Subsequently these intra-class variations are directly used in the gallery set by using knowledge transfer. Based on these variations, we can generate some synthetic samples, thus extending a single sample gallery set to one having multiple samples per subject. Finally, the traditional SRC framework is used on the multi-sample gallery set to classify the query samples in the test set. In this paper, we verified the effectiveness of this method on the CMU Multi-pie database [9], including illumination, expression and pose variation. These results show that the LGC method has a promising FR performance for STSPP.

The structure of the remainder of the paper is as follows. We propose the LGC method in Section 2. Both reconstructive and discriminant dictionary learning schemes are proposed. Furthermore, we show how to learn a dictionary and how to construct the synthetic images and the recognition framework. Section 3 presents the experimental results and analysis. Finally, we present our conclusions in Section 4.

2 Learn-Generate-Classify Method

Considering the fact that face variations for different subjects share many similarities, an additional generic training set with multiple samples per person might bring new and useful information. Looking into the relationship between the different images of one subject, we can capture the intra-class variance in the generic training set. These intra-class variations are directly used to generate some synthetic samples, thus extending the single training sample gallery set to one having multi samples per subject. Finally, we can use the traditional SRC framework on the multi-sample gallery set to classify the query samples.

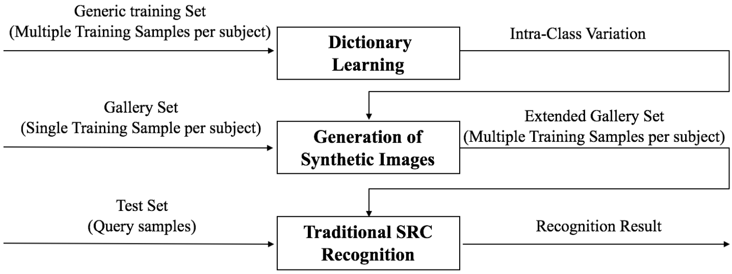

Figure 1: The overall framework of LGC method

2.1 Representation of face images having variation

In general, two important factors are used to characterize a training database of face images, specifically the number of subjects and the types of image variations of each subject (e.g., illumination, expression and pose). Ideally, in the generic training set, we can assume that the number of subjects present is large enough to retain the generalization and so the intra-class variations are similar among different subjects. Actually, since different subjects share some similarities, a single-sampled gallery image, denoted as x, can be represented by a linear combination of other images (excluding the gallery subject) in the generic training set.

If we define D as a subset of the generic training set and α as the coefficients of x over D, they can be described as follows,

$$x = D\alpha. \tag{1}$$

Each column of D is a vectorised training sample of a generic subject, and we assume the variation of each subject in D can be shared and regarded as the one of the gallery subjects.

Based on similar variations among different subjects, we further define $x_{(v)}$ as a sample with the same identity as x but with a specific variation type v and that $x_{(v)}$ can be represented as follows,

$$x_{(v)} = D_{(v)}\alpha_{(v)}, \tag{2}$$

where $D_{(v)}$ is the counterpart of D where the subscript denotes the type of variation. Subjects with similar frontal images should be similar in the other views, and indeed this observation has been successfully used in [2][13][18]. Thus given that $x_{(v)}$ and $D_{(v)}$ show similar variations to x and D respectively, the representation coefficients α and $\alpha_{(v)}$ should be similar as well, i.e.,

$$\alpha \approx \alpha_{(v)}. \tag{3}$$

2.2 Dictionary Learning to capture the intra-class variance from the generic set

We assume that there is a discriminant space, in which the representation of the face image is invariant to illumination, expression and pose. Specifically, it is assumed that the faces of the same subject in different views can be represented as the same one in the discriminant space. For a specific subject, the representation of its samples from different views with respect to the corresponding view-dictionary will be the same. Consequently, for the i^{th} subject, we have the following set of equations

$$\begin{cases} x_{i1} = D_1\alpha_i + e_1 \\ \quad\quad \cdots \\ x_{ip} = D_p\alpha_i + e_p \\ \quad\quad \cdots \\ x_{iN} = D_N\alpha_i + e_N \end{cases}, \tag{4}$$

where the sparse representation α_i is shared among the different view conditions of the subject i. The dictionary D_p is the corresponding dictionary in the p^{th} view condition. There are a total of N different view conditions and e_p is the residual for the recovered image based on the dictionary D_p and sparse representation α_i .

This model is similar to the joint sparse model employed in [12], since the sparse representation vector in the discriminant space is shared among all the view conditions of the same subject. However, the difference here is that the sparse representation α_i is based on different dictionary bases, changing according to the view condition. Thus in our objective function Eq.(5), we want to find simultaneously the dictionary D and sparse representation to minimize the reconstruction error, where the regularization term $\|\alpha\|_1$ is used to guarantee the representation signal α satisfies the sparsity constraint, and λ_2 is a scale parameter to balance the two terms;

$$arg \min_{D,\alpha} \|X - D\alpha\|_2 + \lambda_2 \|\alpha\|_1 . \tag{5}$$

Although the dictionary we learn via Eq.(5) is reconstructive, it is not discriminant. It is so because we do not use the label information of the training samples X in the dictionary learning process. In order to make the learned dictionary both reconstructive and discriminant, the dictionary learning process should add another constraint to encourage the images from the same subject to have similar sparse coefficients and those from different subject to have dissimilar sparse coefficient representations. More specifically, we need to consider a new label consistency constraint [10], called the 'discriminant sparse-code error' and combine it with the reconstruction error to form a unified objective function, i.e.,

$$arg \min_{D,W,\alpha} \|X - D\alpha\|_2 + \lambda_1 \|Q - W\alpha\|_2 + \lambda_2 \|\alpha\|_1 , \tag{6}$$

where λ_1 controls the relative contribution of the reconstructive term and the discriminant term. The consistently label Q is the ground truth for which dictionary columns should contribute to each of the training images. Q is a $K*N$ matrix where K is the number of dictionary columns and N is the number of training samples in matrix X. Q is deemed as the ground truth discriminative sparse coefficients of the input signal X in the classification. The i^{th} column of Q, namely q_i, represents a discriminative sparse coefficient corresponding to a training sample x_i. The nonzero values occur at the places where the input signal x_i and dictionary atom d_k share the same label. Thus we define Q as

$$Q = [q_1, q_2, ..., q_N] \in R^{K*N} \quad q_i = \left[q_i^1, q_i^2, ..., q_i^k\right] = [0...1, 1...0]^T \in R^K. \tag{7}$$

The matrix W is a linear transformation matrix and the linear transformation $g(W, \alpha) = W\alpha$ transforms the original sparse codes α to the most discriminate sparse feature domain. The term $\|Q - W\alpha\|_2$ represents the discriminative sparse coefficient error and our objective is to minimize this error so that the sparse representation is more discriminative. In other words, it forces the transformed sparse representation to approximate the ground truth discriminative code Q, which encourages the training images of the same subject to have similar sparse codes, that should encourage good recognition performance.

The objective function Eq.(6) is not jointly convex to D, W, α. Therefore, we solve this problem by breaking it into two sub problems, and alternately update the unknown variables. It involves a sparse coding state using a pursuit algorithm, such as Orthogonal Matching Pursuit (OMP) [17] or the FOCal Underdetermined System Solver (FOCUSS) [4], followed by an update of the dictionary. We use the classical K-SVD method to update the dictionary atoms gradually. The pseudo code of the K-SVD optimization process can found in [1].

2.3 Generation of Synthetic samples

To generate multiple synthetic samples from the single training sample, two steps are required. The first step is to recover the discriminant representation α_i over the generic training set, which aims to represent the query sample as a linear combination of the images from the generic set. This process represents a transformation from the observation space to the new discriminant space. The second step is to synthesize a series of virtual images using α_i and the corresponding view-basis dictionary, which represents an information flow from the discriminant representation back to the observation space. More specifically, these two steps will now be illustrated.

The first step is to calculate the sparse coefficient representation of the single training sample. For the single sample of the j^{th} person in the p^{th} view condition $y_{j,p}$, we calculate the sparse coefficient α_j, using the p^{th} view of the updated dictionary D_p, as shown in Eq.(8)

$$arg \min_{\alpha_j} \left\|y_{j,p} - D_p \alpha_j\right\|_2 + \lambda \left\|\alpha_j\right\|_1. \tag{8}$$

Since we assume that different images of the same subject should share the same sparse representation, we can use the sparse coefficient α_j to construct a series of synthetic images by using different view-basis dictionaries as shown in Eq.(9). Here, we assume the reconstructive errors that correspond to the different view dictionaries are the same. Thus, we use the $\alpha_{j,v_1}, \alpha_{j,v_2}, ..., \alpha_{j,v_{(p-1)}}$, as the new atoms to extend the gallery set from a single training sample per person to one having multiple samples per person. Fig.2 shows an example of the ground truth images of a subject and the synthetic ones both with and without the sum of residuals. We can see that the synthetic images with the sum of residuals look more similar to the original ones than those without.

$$\begin{pmatrix} \alpha_{j,v_1} \\ \alpha_{j,v_2} \\ \cdots \\ \alpha_{j,v_{p-1}} \end{pmatrix} = \begin{pmatrix} D_1 \\ D_2 \\ \cdots \\ D_{p-1} \end{pmatrix} \alpha_j + \begin{pmatrix} e_1 \\ e_2 \\ \cdots \\ e_{p-1} \end{pmatrix}. \tag{9}$$

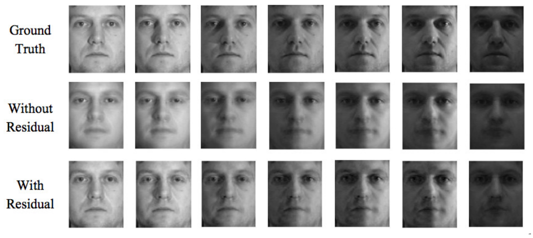

Figure 2: Comparision of the original and synthetic images

2.4 SRC recognition algorithm based on the extended Gallery set

We use the traditional SRC framework for classification. The key idea of the SRC method is to represent the test images as the linear combination of training data from the class to which it belongs. More specifically, we can only represent the test samples in an over-complete dictionary in which atoms are all the training images. Since there are limited numbers of images of each class, which accounts for only a small portion of the whole training set, most of the weight coefficients for the different training subjects should be zero. Thus, the resulting weight satisfies the sparsity requirement of using CS, which means the sparsest coding vector can be found by using the efficient l_1-minimization technique. Finally, the test sample can be classified by calculating the representation residuals among the classes, of which the minimum will correspond to the one to which the test sample belongs.The pseudo code of SRC algorithm can be found in [19].

2.5 Overall algorithm

The overall LGC algorithm is summarized in Algorithm 1. In each iteration of the LGC, the objective function of the representation error will decrease and so the proposed dictionary learning algorithm will converge.

3 Experimental Results

In this section, we perform FR with STSPP on the large-scale CMU Multiple PIE database[9]. We first discuss the experiment parameters in section 4.3.1. In section 4.3.2, we test the robustness of LGC in response to variations, specifically illumination, pose and expression. We compare the proposed LGC method with some other state of the art methods, including Nearest Neighbor (NN) [8], Support Vector Machine (SVM) [16], Sparse Representation Classification (SRC) [19], Adaptive Generic Learning (AGL) [14] and Extended SRC

Algorithm 1 Learn-Generate-Classify (LGC) Method

Initialization: D and Q by Eq.(4) and Eq.(7)

Variation Dictionary Learning
while not converge **do**
    Update α in Eq.(6) with fixed Dictionary D and Q
    Update dictionary D atom by atom via K-SVD Algorithm in paper [1]
end while
Generate Synthetic images via Eq.(9)
Classify the test sample via SRC Algorithm in paper [19]

Output: Classification result

(ESRC)[7]. Among these, NN, SVM and SRC do not use a generic training set, while the others do.

3.1 Parameter Setting

There are two regularization parameters, λ_1 and λ_2, where λ_1 controls the relative contribution of the reconstructive term and the discriminant term. Parameter λ_1 should be set to a relatively small value to improve tolerance to global variation. Parameter λ_2 guarantees that the coefficient should be sparse and so should be set to a relatively large value to enforce sparse representation over the learned dictionary. Parameter λ_2 should be fixed in two phases, first in the representation of face images with variation and second in the generation of synthetic samples. In all the experiments we fix λ_1=0.001 and λ_2=0.02. In addition, the number of iterations n in KSVD should be set beforehand. In the following experiments, we fixed n=3 to balance the learned accuracy against the speed of computation.

3.2 Robustness to various variations

We evaluate the robustness of all the selected methods on the large scale CMU Multi-Pie database. The images in this database are captured in four sessions with simultaneous variation of illumination, pose and expression. For each subject in each session, there are 20 illumination scenarios, indexed from 0-19 per expression per pose. In all experiments, the images of the first 100 subjects in session 1 are deemed as the gallery images. The remaining 149 subjects are used for generic training. For the gallery images, we select the single frontal image with illumination 7 and neutral expression as the single training sample. All the other images of the first 100 subjects in sesssion 1 are deemed as the test images. In the following tests, we choose the corresponding illumination, expression or pose in the generic training set to learn the variation dictionary. In addition we use the learned dictionary to generate synthetized images. The image is cropped to 176*137 pixels. Except for AGL which learns its own features, all the other methods use a 90-dimension Eigen-face for dimension reduction.

3.2.1 Illumination variation

We use all the frontal images with a neutral expression in sessions 2, 3 and 4 as the test images. Fig.3(a) shows an example of the original images from the same subject having different illumination in sessions 2, 3 and 4 (S2, S3 and S4) and the single training sample in the gallery set. Table 1 shows the recognition rates achieved by the different methods.

Figure 3: Example images with illumination or pose variations

Table 1: Recognition Rates for illumination variation

Session	S2	S3	S4
NN	43.64%	40.23%	38.99%
SVM	43.64%	40.53%	42.50%
SRC	44.66%	38.80%	43.21%
AGL	84.37%	79.51%	79.33%
ESRC	89.25%	84.06%	87.40%
LGC	**90.72%**	**88.42%**	**89.68%**

From Table 1, it can be shown that the LGC achieves the best recognition performance in all cases, which means that LGC can effectively capture the illumination variation from the generic training set. ESRC performs the second best, followed by AGL. We can see that those learning methods employing a generic set are much better than the methods that do not. The recognition rates of NN, SVM, SRC are similar, at between 38.80-44.66%. This is because the illumination variation cannot be learned effectively from the gallery set itself.

3.2.2 Pose Variation

In this experiment, the test sample includes face images with pose 05-0 in Session 2 (P05-0-S2), pose 04-1 in Session 3 (P04-1-S3), and pose 04-1 in Session 4 (P04-1-S4). Fig.3(b) shows examples of images from the same subject with different poses in sessions 2, 3, 4 and the single training sample in the gallery set in the first session. The recognition rates of all the methods evaluated are shown in Table 2.

Table 2: Recognition Rates for pose variation

Session	P05-0-S2	P04-1-S3	P04-1-S4
NN	19.01%	8.99%	6.72%
SVM	18.54%	8.73%	6.72%
SRC	18.64%	9.21%	6.99%
AGL	50.71%	23.58%	19.88%
ESRC	53.95%	29.74%	22.58%
LGC	**56.77%**	**33.74%**	**29.25%**

From Table 2, we can see that the LGC method performs quite effectively in the presence of pose variation. Indeed, it performs the best in all cases tested, and the recognition rate is much higher than the second best namely ESRC by at least 3, 4 and 7 percentage points respectively. We can see that when the pose variation becomes more severe, i.e., Pose 4-1 compared to Pose 5-0, the LGC method outperforms other methods by significant margins.

In addition, the LGC method can achieve a better performance over a wider range of pose variation than can ESRC. Actually, for the pose variation scenario, it is always much more accurate to capture the inta-class variation based on dictionary learning than using pixel differences. Thus, the LGC method is a better candidate for solving the pose variation problem than ESRC.

3.2.3 Expression variation

In this experiment, the test samples include the frontal face images having the following expressions, smile in session 1 (Smi-S1), surprise in Session 2 (Sur-S2), squint in Session 2 (Squ-S2), smile in Session 3 (Smi-S3), disgust in S3 (Dis-S3) and scream in Session 4 (Scr-S4). Examples of all these expressions are shown in Fig.4. The recognition rates of all the methods evaluated are presented in Table 3.

Figure 4: Example images with Expression variations

Table 3: Recognition Rates for expression variation

Expression	Smi-S1	Sur-S2	Squ-S2	Smi-S3	Dis-S3	Scr-S4
NN	43.91%	17.18%	32.02%	27.22%	20.23%	7.55%
SVM	44.03%	17.18%	29.70%	28.35%	20.55%	8.05%
SRC	44.13%	19.74%	31.80%	23.16%	18.80%	8.31%
AGL	83.78%	31.34%	35.79%	40.32%	25.88%	10.59%
ESRC	82.67%	**43.93%**	42.20%	50.60%	**35.64%**	13.99%
LGC	**91.39%**	41.67%	**43.73%**	**56.02%**	32.78%	**14.01%**

From Table 3, we can see that LGC achieves the best recognition performance in half of the cases investigated. It performs best when the training expression variation type is same as the test expression variation type. Since we use the smile expression as the generic training set, we can see the LGC is much better than the other methods when the test image is also in the smile expression. For the Smi-S1 and Smi-S3, LGC is better than ESRC by 9 and 6 percentage points respectively. However, when the generic training variation type is different from the test variation type, LGC is better than ESRC by only about 0-2 percentage points, or can even be worse than ESRC when the test image is either the surprise or disgust expressions. In conclusion, we can see that the LGC is effective for the single training sample problem and can tolerate some variation of expression, especially when the generic training is pertinent, otherwise, the recognition performance may degrade dramatically.

4 Conclusion

In this paper, we propose a Learn-Generate-Classify (LGC) method for the challenging task of FR with only a STSPP. The LGC method utilizes the advantage of both generic learning and synthesized sample generation for the STSPP problem in FR. First, it learns a series of dictionaries with different views from a generic training set in order to predict the intra-class variation in the Gallery set. Second, it synthesizes multiple images per subject to extend the single-sample gallery set. Finally based on the synthesized set, the Sparse Representation Classification (SRC) framework is used for classification. According to the experimental results, the LGC method is shown to outperform the other methods evaluated when the variations exist globally, for example illumination and pose variation. However, it has little performance advantage when the variations are centred at local image patches, i.e., expression variation. The extensive experiments on the CMU Multi-Pie database with various face variations have demonstrated the effectiveness of LGC method for the STSPP problem.

References

[1] Michal Aharon, Michael Elad, and Alfred Bruckstein. K-SVD: An algorithm for designing overcomplete dictionaries for sparse representation. *Signal Processing, IEEE Transactions on*, 54(11):4311–4322, 2006.

[2] Timo Ahonen, Abdenour Hadid, and Matti Pietikäinen. Face recognition with local binary patterns. In *Computer vision-eccv 2004*, pages 469–481. Springer, 2004.

[3] Songcan Chen, Jun Liu, and Zhi-Hua Zhou. Making FLDA applicable to face recognition with one sample per person. *Pattern recognition*, 37(7):1553–1555, 2004.

[4] Shane F Cotter, Bhaskar D Rao, Kjersti Engan, and Kenneth Kreutz-Delgado. Sparse solutions to linear inverse problems with multiple measurement vectors. *Signal Processing, IEEE Transactions on*, 53(7):2477–2488, 2005.

[5] Thomas Cover and Peter Hart. Nearest neighbor pattern classification. *Information Theory, IEEE Transactions on*, 13(1):21–27, 1967.

[6] Fernando De la Torre, Ralph Gross, Simon Baker, and BVKV Kumar. Representational oriented component analysis (ROCA) for face recognition with one sample image per training class. In *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*, volume 2, pages 266–273. IEEE, 2005.

[7] Weihong Deng, Jiani Hu, and Jun Guo. Extended SRC: Undersampled face recognition via intraclass variant dictionary. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 34(9):1864–1870, 2012.

[8] Pedro Felzenszwalb, David McAllester, and Deva Ramanan. A discriminatively trained, multiscale, deformable part model. In *Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on*, pages 1–8. IEEE, 2008.

[9] Ralph Gross, Iain Matthews, Jeffrey Cohn, Takeo Kanade, and Simon Baker. Multi-pie. *Image and Vision Computing*, 28(5):807–813, 2010.

[10] Zhuolin Jiang, Zhe Lin, and L.S. Davis. Label consistent K-SVD: Learning a discriminative dictionary for recognition. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 35(11):2651–2664, Nov 2013. ISSN 0162-8828. doi: 10.1109/TPAMI.2013.88.

[11] Yan Ke and Rahul Sukthankar. PCA-SIFT: A more distinctive representation for local image descriptors. In *Computer Vision and Pattern Recognition, 2004. CVPR 2004. Proceedings of the 2004 IEEE Computer Society Conference on*, volume 2, pages II–506. IEEE, 2004.

[12] Jiwen Lu, Yap-Peng Tan, and Gang Wang. Discriminative multimanifold analysis for face recognition from a single training sample per person. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 35(1):39–51, 2013.

[13] H. Mohammadzade and D. Hatzinakos. Projection into expression subspaces for face recognition from single sample per person. *Affective Computing, IEEE Transactions on*, 4(1):69–82, Jan 2013.

[14] Yu Su, Shiguang Shan, Xilin Chen, and Wen Gao. Adaptive generic learning for face recognition from a single sample per person. In *Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on*, pages 2699–2706. IEEE, 2010.

[15] Xiaoyang Tan, Songcan Chen, Zhi-Hua Zhou, and Fuyan Zhang. Face recognition from a single image per person: A survey. *Pattern recognition*, 39(9):1725–1745, 2006.

[16] Simon Tong and Daphne Koller. Support vector machine active learning with applications to text classification. *The Journal of Machine Learning Research*, 2:45–66, 2002.

[17] Joel A Tropp and Anna C Gilbert. Signal recovery from random measurements via orthogonal matching pursuit. *Information Theory, IEEE Transactions on*, 53(12):4655–4666, 2007.

[18] Lior Wolf, Tal Hassner, and Yaniv Taigman. Effective unconstrained face recognition by combining multiple descriptors and learned background statistics. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 33(10):1978–1990, 2011.

[19] John Wright, Allen Y Yang, Arvind Ganesh, Shankar S Sastry, and Yi Ma. Robust face recognition via sparse representation. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 31(2):210–227, 2009.

[20] John Wright, Yi Ma, Julien Mairal, Guillermo Sapiro, Thomas S Huang, and Shuicheng Yan. Sparse representation for computer vision and pattern recognition. *Proceedings of the IEEE*, 98(6):1031–1044, 2010.

[21] Daoqiang Zhang, Songcan Chen, and Zhi-Hua Zhou. A new face recognition method based on SVD perturbation for single example image per person. *Applied Mathematics and computation*, 163(2):895–907, 2005.

[22] Wenyi Zhao, Rama Chellappa, P Jonathon Phillips, and Azriel Rosenfeld. Face recognition: A literature survey. *Acm Computing Surveys (CSUR)*, 35(4):399–458, 2003.