LOCAL STRUCTURE BASED SPARSE REPRESENTATION FOR FACE RECOGNITION WITH SINGLE SAMPLE PER PERSON

Fan Liu, Jinhui Tang, Yan Song, Xinguang Xiang, Zhenmin Tang

School of Computer Science and Engineering, Nanjing University of Science and Technology, China

ABSTRACT

In this paper, we propose local structure based sparse representation classification (LS_SRC) to solve single sample per person (SSPP) problem. By adopting the "divide-conquer-aggregate" strategy, we successfully alleviate the dilemma of high data dimensionality and small samples, where we first divide the face into local blocks, and classify each local block, and then integrate all the classification results by voting. For each block, we further divide it into overlapped patches and assume that these patches lie in a linear subspace. This subspace assumption reflects local structure relationship of the overlapped patches and makes SRC feasible for SSPP problem. To lighten the computing burden, we further propose local structure based collaborative representation classification (LS_CRC). Experimental results on three public face databases show that our methods not only generalize well to SSPP problem but also have strong robustness to expression, illumination, little pose variation, occlusion and time variation.

***Index Terms*—** Face recognition, single sample per person problem, sparse representation, collaborative representation, local structure

1. INTRODUCTION

Face recognition is an active research topic that has attracted significant attention in the domain of computer vision and pattern recognition for many years [1], due to its scientific challenges and potential applications. In the prior literature, plenty of research efforts have been dedicated into face recognition problem and tremendous progress has been achieved [2, 3, 4, 5, 6].

In spite of the tremendous achievements, there are still many challenges caused by the large face appearance variations of illumination, expression, pose, noise, occlusion, etc. [7]. The typical approach in handling these variations is to use large and representative training sample sets. However, in many real-world applications such as identity card verification, passport verification in customs, law enforcement, surveillance or access control, only one training sample per person is available. This problem is called *single sample per person* (SSPP) problem [8] which has become one of the greatest challenges in real world applications.

In this paper, we propose a simple yet effective face classification scheme, called as local structure based sparse representation classification (LS_SRC), to solve SSPP problem. Motivated by the "divide-and-conquer" [9] strategy, we partition each face into a set of overlapped blocks and classify each blocks, then aggregate the classification results by voting to make the final decision. To generate classification result for each block, we further divide block into overlapped patches and assume that these patches lie in a linear subspace due to the fact that the patches are strongly similar. Intuitively, if the training sample and test sample describe the same person, then as to each pair of corresponding local blocks, the subspaces that the local patches lie in should be the same. In other words, the central patch of the test block can be approximately represented by a linear combination of the patches in the corresponding block from the same class. Therefore, the local structure relationship in each block makes SRC feasible even when encountering SSPP problem. To lighten the computing burden, we further propose local structure based collaborative representation classification (LS_CRC). Experimental results on three public face databases show that our methods not only generalize well to single sample per person problem but also have strong robustness to expression, illumination, little pose variation, occlusion and time variation.

The rest of this paper is organized as follows. We start by introducing prior work in Section 2. Then in Section 3, we present the proposed local structure based methods: local structure based sparse representation classification (LS_SRC)and local structure based collaborative representation classification (LS_CRC). Section 4 demonstrates experiments and results. Finally, we conclude in Section 5 by highlighting key points of our work.

2. PRIOR WORK

Recently, sparse representation based classification (SRC) [5] and collaborative representation based classification (CRC)[6] have shown very effective face recognition performance. However, they always require a rich set of training samples of each subject. This greatly limits their performance for SSPP problem. During the last two decades, many methods have been proposed to deal with SSPP problem. Chen et al. [10] proposed BlockFLD method which generates multiple training samples for each person by partitioning each face image into a set of same sized blocks and then applies FLD-based methods with these blocks. There are also some methods using a generic training set to extract the discriminatory information. The works [11, 12, 13] are under this framework. Shan et al. [14] proposed the Adaptive Generic Learning (AGL) method, which does not directly employ the discriminatory information, but adapts it to the persons to be identified. In addition, as local region partition usually can lead to significant improvement in recognition rate and robustness, many patch (or block) based methods have been proposed to solve the SSPP problem. For example, Zhu et al. [15] proposed patch based CRC (PCRC) for small sample size (SSS) problem. Kumar et al. [16] proposed patch based n-nearest classifier to improve the stability and generalization ability. However, they all ignore the local structure relationship of the overlapped patches.

3. LOCAL STRUCTURE BASED SPARSE REPRESENTATION FOR FACE RECOGNITION

As a powerful classifier, the classification ability of SRC relies on computing the correct sparse solution of an underdetermined linear system. But, the correct sparse solution can only be recovered when the number of training samples is sufficiently larger than the dimension of features [5]. Therefore, when encountering SSPP problem, SRC may fail because the sparse solution is not correct any more. To address this issue, we propose a simple yet effective representation method, namely local structure based sparse representation classification (LS_SRC).

3.1. Local Structure

To describe the local structure , we illustrate three kinds of neighborhood in Fig. 1. The P neighbor pixels on a square of radius R form a squarely symmetric neighbor sets. Suppose there are N pixels in an image. For the i-th pixel in the image, its P neighbor pixels can be denoted by $\Omega_P^i = \{i_j | j = 1, \ldots, P\}$.

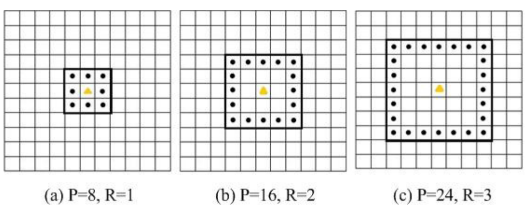

Fig. 1. Squarely symmetric neighbor sets for different R.

For the i-th pixel in the image, we select a $S \times S$ local patch (e.g. S=3, 5) centered at it. All the S^2 pixels within the patch form a m dimensional local patch vector $\boldsymbol{x}_0^i$, where $m = S^2$. Similarly, the neighbor pixel i_j of the i-th also corresponds to a same sized local patch, whose patch vector is denoted by $\boldsymbol{x}_j^i, j = 1, \ldots, P$. Then , the center patch and neighbor patches determine a local block centered at the i-th pixel. Fig. 2 shows an example of a local block containing a central patch and 16 neighbor patches. The size of patch is 3×3 and the size of the block is 7×7. For the pixel on the margin of an image, we use the mirror transform first and then determine its local block.

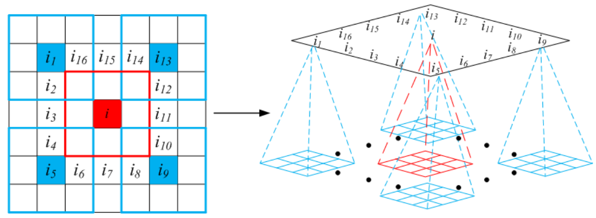

Fig. 2. Illustration of local patches in a local block.

As the patches are overlapped and concentrate in a small block, they are strongly similar. Therefore, we assume that all patches in a local block lie in a linear subspace. This subspace assumption has also been successfully used for feature extraction in [17]. Here, we take advantage of it to improve the robustness of SRC to SSPP problem.

3.2. Local Structure Based Sparse Representation

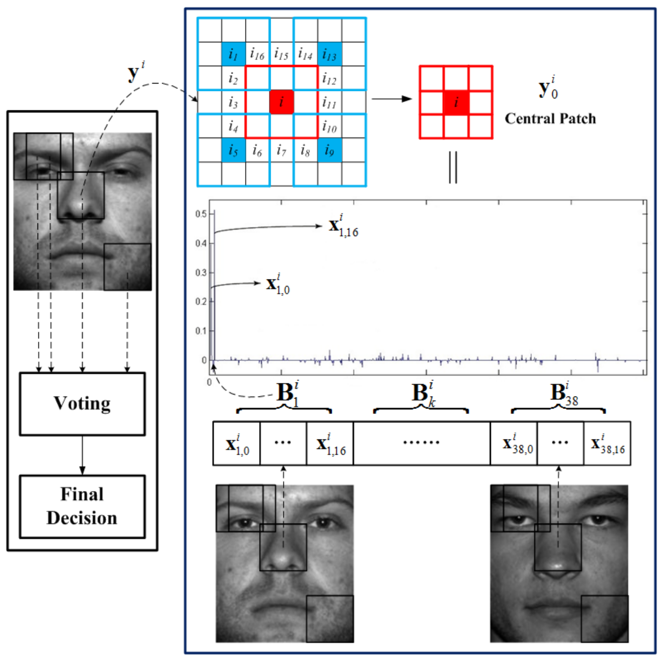

Fig. 3. Diagram of local structure based sparse representation for face recognition.

As shown in Fig. 3, the query image $\boldsymbol{y}$ is first divided into a set of overlapped blocks $\{\boldsymbol{y}^1, \boldsymbol{y}^2, \ldots, \boldsymbol{y}^N\}$. The i-th pixel of the image corresponds to the local block $\boldsymbol{y}^i$, which consists of some overlapped patches, where $\boldsymbol{y}^i = [\boldsymbol{y}_0^i, \boldsymbol{y}_1^i, \ldots, \boldsymbol{y}_P^i] \in R^{m\times(P+1)}$. Similarly, we also decompose the training samples into blocks. Suppose that there is only one training image per person, the i-th block of the k-th person is denoted by $\boldsymbol{B}_k^i = [\boldsymbol{x}_{k,0}^i, \boldsymbol{x}_{k,1}^i, \ldots, \boldsymbol{x}_{k,P}^i] \in R^{m\times(P+1)}$, where $\boldsymbol{x}_{k,j}^i (j = 0, \ldots, P)$ is the local patch vector in the block. All the i-th blocks from K classes form a local dictionary $\boldsymbol{B}^i$, where $\boldsymbol{B}^i = [\boldsymbol{B}_1^i, \boldsymbol{B}_2^i, \ldots, \boldsymbol{B}_K^i]$. According to the description of local structure, $\boldsymbol{B}_k^i$ is supposed to lie in a subspace $\boldsymbol{\Psi}$ and $\boldsymbol{y}^i = [\boldsymbol{y}_0^i, \boldsymbol{y}_1^i, \ldots, \boldsymbol{y}_P^i]$ belongs to a subspace $\boldsymbol{\Phi}$. If the test image $\boldsymbol{y}$ is also from the k-th class and all the training images and test images are well aligned, $\boldsymbol{\Psi}$ and $\boldsymbol{\Phi}$ are theoretically the same subspace or at least they are very close in fact. Therefore, the central patch vector $\boldsymbol{y}_0^i$ will approximately lie in the linear span of all the local patch vectors from $\boldsymbol{B}_k^i = [\boldsymbol{x}_{k,0}^i, \boldsymbol{x}_{k,1}^i, \ldots, \boldsymbol{x}_{k,P}^i]$, which can be written as:

$$\boldsymbol{y}_0^i = \alpha_{k,0}^i \boldsymbol{x}_{k,0}^i + \alpha_{k,1}^i \boldsymbol{x}_{k,1}^i + \ldots + \alpha_{k,P}^i \boldsymbol{x}_{k,P}^i = \boldsymbol{B}_k^i \boldsymbol{\alpha}_k^i \quad (1)$$

Since $\boldsymbol{B}^i = [\boldsymbol{B}_1^i, \boldsymbol{B}_2^i, \ldots, \boldsymbol{B}_K^i]$ is a local dictionary which includes all the i-th blocks from K classes, the linear representation of $\boldsymbol{y}_0^i$ can be rewritten in the form as below:

$$\boldsymbol{y}_0^i = \boldsymbol{B}^i \boldsymbol{\alpha}^i \quad (2)$$

Where $\boldsymbol{\alpha}^i = [0, \ldots, 0, \alpha_{k,0}^i, \alpha_{k,1}^i, \ldots, \alpha_{k,P}^i, 0, \ldots, 0]^T$ is a coefficient vector whose entries are zero except those associated with the k-th class. As $\boldsymbol{y}_0^i$ can be sufficiently represented using only the training block of the same subject, the representation of $\boldsymbol{\alpha}^i$ is naturally sparse if the number of subjects K is reasonably large. Therefore, finding the identity of $\boldsymbol{y}_0^i$ equals finding the sparse solution of (2). This is the same as solving the following optimization problem (l_1-minimization):

$$\hat{\boldsymbol{\alpha}^i} = \arg\min \|\boldsymbol{\alpha}^i\|_1 \quad s.t. \quad \boldsymbol{y}_0^i = \boldsymbol{B}^i \boldsymbol{\alpha}^i \quad (3)$$

According to the classification rule of SRC, $\boldsymbol{y}_0^i$ can be classified by minimizing the following regularized residuals:

$$\min_k r_k(\boldsymbol{y}_0^i) = \|\boldsymbol{y}_0^i - \boldsymbol{B}^i \delta_k(\hat{\boldsymbol{\alpha}^i})\|_2 \quad (4)$$

Where δ_k is a function selecting the coefficients associated with k-th class. The classification outputs of all blocks can then be aggregated. Plurality voting [16] is used for the final decision making, which means that the test sample is finally classified into the class with the largest number of votes. The whole algorithm can be summarized as "divide-conquer-aggregate" procedure.

3.3. Local Structure based Collaborative Representation

Recently, Zhang et al. proposed collaborative representation based classification (CRC) for FR. By using l_2-regularized least square, CRC has much less computational cost than SRC. Inspired by this idea, we propose local structure based collaborative representation classification (LS_CRC) to lessen the computing burden of LS_SRC.

For the test block $\boldsymbol{y}^i$, the collaborative representation of its central patch $\boldsymbol{y}_0^i$ to the local dictionary $\boldsymbol{B}^i$ is computed by the following l_2-regularized least square method:

$$\hat{\boldsymbol{\alpha}^i} = \arg\min\{\|\boldsymbol{y}_0^i - \boldsymbol{B}^i \boldsymbol{\alpha}^i\|_2^2 + \lambda \|\boldsymbol{\alpha}^i\|_2^2\} \quad (5)$$

According to the classification rule of CRC, $\boldsymbol{y}_0^i$ can be classified by minimizing the following regularized residuals:

$$\min_k r_k(\boldsymbol{y}_0^i) = \|\boldsymbol{y}_0^i - \boldsymbol{B}^i \delta_k(\hat{\boldsymbol{\alpha}^i})\|_2 / \|\delta_k(\hat{\boldsymbol{\alpha}^i})\|_2 \quad (6)$$

4. EXPERIMENTAL RESULTS

In this section, we use the Extended Yale B [18], PIE [19], and AR [20] databases to evaluate the proposed methods and compare them with SRC, CRC, patch based SRC (PSRC) and patch based CRC (PCRC). To demonstrate the robustness of our proposed methods to SSPP problem, we also compare them with several popular methods dealing with SSPP problem including BlockFLD [10], patch based nearest neighbor (PNN) classifier [16] and FLDA_single [21]. For the methods we compared, the best result is reported. For our proposed methods, the neighbor set is $P = 8, R = 1$ and the patch size is 11×11. As each block can be classified in the same time, 12 Matlab workers are used for parallel computation.

4.1. Extended YaleB database

The Extended Yale B face database [18] contains 38 human subjects under 9 poses and 64 illumination conditions. All frontal-face images marked with P00 in Extended Yale B are used in our experiment. For each subject, we use the image under the best illumination condition for training, whose azimuth and elevation are both 0 degree. The remaining 63 images under various lighting conditions are used for testing and the average results are reported. The experimental results are shown in Fig. 4.

From the figure, we can see that our proposed methods achieve the best results. Compared with SRC and CRC, they lead to almost 50% improvements. Meanwhile, they also lead to much better results than FLDA_single, BlockFLD and PNN. The comparison with PSRC and PCRC further demonstrates the effectiveness of local structure relationship.

In terms of CPU time, LS_SRC consumes much time because they need to solve l_1-minimization for each block. However, we can significantly improve its computational efficiency by decreasing the number of blocks for classification. Although there are N blocks corresponding to N pixels of the image, we do not need to use all the N blocks for classification.

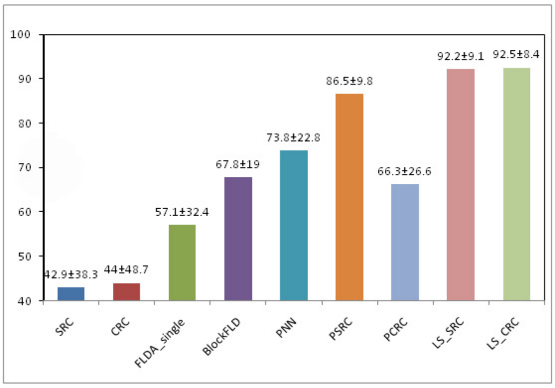

Fig. 4. Recognition accuracy on Extended Yale B database.

4.2. PIE database

In this experiment, we use the five near frontal poses (C05, C07, C09, C27 and C27) of PIE database and select 6 images with different illumination under each pose, thus we get 30 images for each individual. These images will be used for testing. We select a single image with good illumination from C27 for training. The experimental result is shown in Fig. 5.

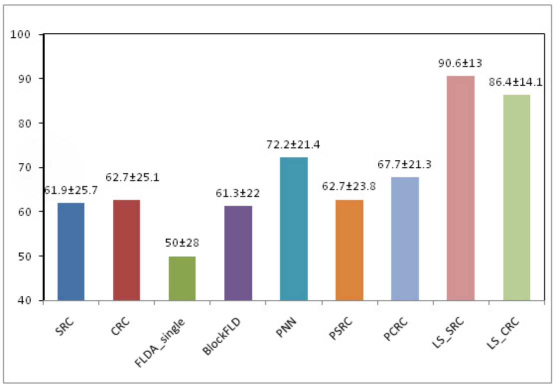

Fig. 5. Recognition accuracy on PIE database.

LS_SRC and LS_CRC still achieve the best results. Compared with SRC, CRC, PSRC and PCRC, they improves at least 18 percentage points. The experimental results also show that our methods are robust to little pose variation since the test images are under five near frontal poses.

4.3. AR database

The AR face database [20] contains over 4,000 color face images of 126 people, including frontal views of faces with different facial expressions, lighting conditions and occlusions. The pictures of 120 individuals taken in two sessions are selected. The single image under natural expression and illumination from session 1 is used for training and the other images from two sessions are used for testing. The classification results on the two sessions are shown in Table 1 and Table 2 respectively.

Table 1. Recognition accuracy on session 1 of AR database.

Method	expression	illumination	sunglasses	scarves	average
SRC	95.3	94.7	88.1	50.6	82.2
CRC	95.3	93.9	86.7	49.2	81.3
PNN	85.8	91.4	70.8	36.9	71.3
BlockFLD	72.5	84.4	72.5	41.7	67.8
FLDA_single	93.1	96.1	89.4	43.1	80.4
PSRC	83.9	93.3	87.2	63.9	82.8
PCRC	84.4	91.4	83.6	63.6	80.8
LS_SRC	**96.1**	**98.3**	**96.7**	**84.5**	**93.9**
LS_CRC	89.2	92.8	83.6	69.5	83.8

Table 2. Recognition accuracy on session 2 of AR database.

Method	expression	illumination	sunglasses	scarves	average
SRC	63.5	62.2	46.9	25.8	49.7
CRC	62.5	60.8	41.7	22.2	47.1
PNN	52.5	51.9	34.2	17.8	39.1
BlockFLD	33.1	41.1	30.7	22.5	31.7
FLDA_single	57.2	53.3	46.1	20.0	44.2
PSRC	47.2	53.1	47.5	38.1	46.5
PCRC	43.9	48.3	38.3	32.5	40.8
LS_SRC	**74.4**	**76.4**	**70**	**61.9**	**70.7**
LS_CRC	57.2	63.9	38.9	34.4	48.6

The experimental results demonstrate that LS_SRC is not only superior to other methods but also robust to expression and illumination variations and occlusion. In the experiments of session 2, we find that LS_SRC is also robust to time variation.

5. CONCLUSION

In this paper, we propose local structure based sparse representation classification (LS_SRC) to solve SSPP problem. To lighten the computing burden of LS_SRC, we also present local structure based collaborative representation classification (LS_CRC). Experimental results show that LS_SRC and LS_CRC generalize well to SSPP problem and have strong robustness to the large variation of expression, illumination, little poses variation, occlusion and time variation. However, the proposed methods rely on the basis that all the training and testing images are well aligned. This constraint is planned to be solved in our future work.

6. ACKNOWLEDGEMENT

This work was partially supported by the Natural Science Foundation of China (NSFC) under Grant 61103059 and 61301106, the Natural Science Foundation of Jiangsu Province under Grant BK2012033 and BK2012397.

7. REFERENCES

[1] Wenyi Zhao, Rama Chellappa, P Jonathon Phillips, and Azriel Rosenfeld, "Face recognition: A literature survey," *Acm Computing Surveys (CSUR)*, vol. 35, no. 4, pp. 399–458, 2003.

[2] Jun Zhang, Yong Yan, and Martin Lades, "Face recognition: eigenface, elastic matching, and neural nets," *Proceedings of the IEEE*, vol. 85, no. 9, pp. 1423–1435, 1997.

[3] Joshua B Tenenbaum, Vin De Silva, and John C Langford, "A global geometric framework for nonlinear dimensionality reduction," *Science*, vol. 290, no. 5500, pp. 2319–2323, 2000.

[4] Mikhail Belkin and Partha Niyogi, "Laplacian eigenmaps for dimensionality reduction and data representation," *Neural computation*, vol. 15, no. 6, pp. 1373–1396, 2003.

[5] John Wright, Allen Y Yang, Arvind Ganesh, S Shankar Sastry, and Yi Ma, "Robust face recognition via sparse representation," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 31, no. 2, pp. 210–227, 2009.

[6] Lei Zhang, Meng Yang, and Xiangchu Feng, "Sparse representation or collaborative representation: Which helps face recognition?," in *Computer Vision (ICCV), 2011 IEEE International Conference on*. IEEE, 2011, pp. 471–478.

[7] P Jonathon Phillips, Patrick J Flynn, Todd Scruggs, Kevin W Bowyer, Jin Chang, Kevin Hoffman, Joe Marques, Jaesik Min, and William Worek, "Overview of the face recognition grand challenge," in *Computer vision and pattern recognition, 2005. CVPR 2005. IEEE computer society conference on*. IEEE, 2005, vol. 1, pp. 947–954.

[8] Xiaoyang Tan, Songcan Chen, Zhi-Hua Zhou, and Fuyan Zhang, "Face recognition from a single image per person: A survey," *Pattern Recognition*, vol. 39, no. 9, pp. 1725–1745, 2006.

[9] Quentin F Stout, "Supporting divide-and-conquer algorithms for image processing," *Journal of Parallel and Distributed Computing*, vol. 4, no. 1, pp. 95–115, 1987.

[10] Songcan Chen, Jun Liu, and Zhi-Hua Zhou, "Making flda applicable to face recognition with one sample per person," *Pattern recognition*, vol. 37, no. 7, pp. 1553–1555, 2004.

[11] Jie Wang, KN Plataniotis, Juwei Lu, and AN Venetsanopoulos, "On solving the face recognition problem with one training sample per subject," *Pattern recognition*, vol. 39, no. 9, pp. 1746–1762, 2006.

[12] Tae-Kyun Kim and Josef Kittler, "Locally linear discriminant analysis for multimodally distributed classes for face recognition with a single model image," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 27, no. 3, pp. 318–327, 2005.

[13] A Majumdar and RK Ward, "Pseudo-fisherface method for single image per person face recognition," in *Acoustics, Speech and Signal Processing, 2008. ICASSP 2008. IEEE International Conference on*. IEEE, 2008, pp. 989–992.

[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*. IEEE, 2010, pp. 2699–2706.

[15] Pengfei Zhu, Lei Zhang, Qinghua Hu, and Simon CK Shiu, "Multi-scale patch based collaborative representation for face recognition with margin distribution optimization," in *Computer Vision–ECCV 2012*, pp. 822–835. Springer, 2012.

[16] Ritwik Kumar, Arunava Banerjee, Baba C Vemuri, and Hanspeter Pfister, "Maximizing all margins: Pushing face recognition with kernel plurality," in *Computer Vision (ICCV), 2011 IEEE International Conference on*. IEEE, 2011, pp. 2375–2382.

[17] Jianjun Qian, Jian Yang, and Yong Xu, "Local structure-based image decomposition for feature extraction with applications to face recognition," *Image Processing, IEEE Transactions on*, vol. 22, no. 9, pp. 3591–3603, 2013.

[18] Athinodoros S. Georghiades, Peter N. Belhumeur, and David 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, no. 6, pp. 643–660, 2001.

[19] Terence Sim, Simon Baker, and Maan Bsat, "The cmu pose, illumination, and expression database," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 25, no. 12, pp. 1615–1618, 2003.

[20] Aleix M Martinez, "The ar face database," *CVC Technical Report*, vol. 24, 1998.

[21] Quan-xue Gao, Lei Zhang, and David Zhang, "Face recognition using flda with single training image per person," *Applied Mathematics and Computation*, vol. 205, no. 2, pp. 726–734, 2008.