LOW RANK GROUP SPARSE REPRESENTATION BASED CLASSIFIER FOR POSE VARIATION

Shivangi Yadav, Maneet Singh, Mayank Vatsa, Richa Singh, Angshul Majumdar

IIIT-Delhi, India

ABSTRACT

Face recognition under uncontrolled environment persists to be an unresolved problem having challenges such as varying pose, illumination, occlusion etc. In this research, we propose an algorithm for identification of faces with pose and illumination variations. An adaptive dictionary learning framework built upon Group Sparse Representation Classifier is presented in order to learn dictionary parameters and pose invariant sparse codes for given images. Low rank regularization is utilized for dictionary learning, to deal with the noise present in training samples which can hinder the discriminative power of learnt dictionary. Experimental results show state-of-the-art performance on CMU Multi-PIE dataset.

***Index Terms*—** Dictionary, Group Sparse Representation based Classifier, Low Rank, Pose and Illumination variation

Fig. 1. Sample images of an individual from CMU Multi-PIE dataset [7] with pose and illumination variation. First row shows illumination variations and second row refers to the pose variations.

1. INTRODUCTION

Face recognition has been the focus of many biometrics researchers since the past few decades [1]. Face has been established as one of the least invasive biometric modalities, thereby, making it one of the most well explored signatures for person identification as well. It provides discriminative textural and structural information, which is often used for identity recognition. Many algorithms have been proposed to automate this task under several covariates such as varying resolution, occlusion, disguise etc. [2, 3, 4]. Though recent algorithms claim high accuracies [5, 6], the performance of the same in real-world conditions is still an unresolved issue.

One of the major challenges associated with automated face recognition in completely unconstrained scenarios is the presence of pose and illumination variations. Algorithms that utilize only frontal, well-illuminated face images for learning a classification model are often ineffective in the presence of such fluctuations. This is primarily because the distribution of data on which the classifier is trained might differ from the distribution of the test samples. For example, Figure 1 shows images from a single individual with varying pose and illumination. It is clear that face recognition with such variations require special attention. One possible solution to address this is to obtain large amount of training samples for all possible variations, which in itself is a challenging task. This generates a need for a less data-intensive algorithm that can handle such variations in the data distribution.

In this research, a Low Rank Group Sparse Representation based algorithm is proposed for face recognition with pose and illumination variations. The algorithm is built upon existing Group Sparse Classifier [8] and utilizes incremental learning [9] with trace norm regularizer [10] for addressing the given problem. In Dictionary Learning approaches, images are represented as a linear combination of atoms of a dictionary. Generally, for a given dictionary, the total number of atoms are large as opposed to the atoms used for the reconstruction of a given image, which results in sparse coefficients for the image. Recently, Group Sparse Classifier has been proposed which assumes that a test sample can be represented as a linear combination of training samples belonging to the same group as that of the given test sample. Since the samples are linearly correlated, the dictionary for a particular group should fall in a low dimensional manifold [11]. To enforce this, a trace norm regularizer on the group-wise dictionaries is introduced in the dictionary learning protocol. As mentioned earlier, since the distribution of the test samples (target domain) might differ from the distribution of the training samples (source domain), the above framework is learnt in an incremental manner.

The paper has been organized as follows: Section 2 briefly describes the existing work on this problem, which is followed by the proposed algorithm in Section 3. Section 4 gives details about the experimental setup and results of the proposed algorithm, as well as comparison with existing approaches. Conclusion and future work are presented in Section 5.

2. RELATED WORK

In literature, major approaches applied to face recognition with pose and illumination variations include methods like 3-D face reconstruction, image mosaicing, deep learning and domain adaptation. Passils *et al.* [12] proposed to learn a 3-D model for face recognition using facial symmetry across different poses. Zhu *et al.* [13] aims to learn face Identity-Preserving features (FIP) using deep learning, however, as is the case with Deep Learning, this approach requires large amount of training data. Singh *et al.* [14] describe a face mosaicing scheme to generate a composite face from frontal and semi-profile faces. Qiu *et al.* [9] proposed a dictionary learning framework, Domain Adaptive Dictionary Learning (DADL), to transfer information from source domain to target domain for re-identification of faces. This algorithm addresses the problem of variation in data distribution between source and target domain but it is unable to handle any variation within the target domain itself. The next section presents the proposed algorithm, which aims to overcome these shortcomings.

3. LOW RANK GROUP SPARSE REPRESENTATION BASED CLASSIFIER (LR-GSRC)

This section describes the proposed algorithm, prior to which some background knowledge about sparse coding, dictionary learning and GSRC is discussed.

3.1. Sparse Representation and Dictionary Learning

Sparse modeling of data has attracted a lot of attention in the past few years. In dictionary learning algorithms, images are represented as linear combination of few atoms of a dictionary. Given a signal y and dictionary D, sparse representation of y can be learned through following optimization problem:

$$\hat{x} = \arg\min_{x} \|x\|_o \textit{ subject to } y = Dx \tag{1}$$

where, $\|x\|_o$ refers to l_o norm that gives the number of nonzero entries in vector x.

Recently, many new approaches have been discussed to learn an efficient dictionary [15, 16] from the given data. It has mainly been influenced by recent advances in sparse algorithms and representation theory. One of the established methods of learning a dictionary from training samples is the K-SVD algorithm [17]. Given a sample y, K-SVD aims to learn a dictionary D and its sparse code x such that the reconstruction error is minimized:

$$\arg\min_{D,X} \|Y - DX\|_F^2 \quad \textit{s.t.} \vee_i, \ \|x_i\|_o \leq T \tag{2}$$

where, $X = [x_1, ..., x_N], \ x_i \,\epsilon \in R^k$ are sparse codes of N input signals Y,$D = [d_1, ..., d_k], \ d_i \,\epsilon \in R^n$ and T restricts the signal to have less than T items in its decomposition. k represents the number of atoms in learned dictionary and n represents the number of samples on which the dictionary has been learnt.

3.2. Group Sparse Representation based Classification

An established sparse representation based classification approach is Sparse Representation based Classification (SRC) [18]. It assumes that a given test sample can be represented as a linear combination of training samples belonging to the same class as the given test sample:

$$v_{test} = \alpha_{k,1}v_{k,1} + \alpha_{k,2}v_{k,2} + ... + \alpha_{k,n}v_{k,n} + \epsilon \tag{3}$$

where, v_{test} belongs to class k, $v_{i,k}$ represents i^{th} training sample from k^{th} class and ϵ is the approximation error.

Since the correct class of v_{test} is not known at the time of classification therefore, SRC represents v_{test} as a linear combination of all the training samples from all classes. For classification, SRC aims to learn the coefficients α in eq.(3) for v_{test} such that α values for the correct class are non-zero while, remaining are zero. This results in a sparse vector for α which is solved by the following minimization problem:

$$\min_{\alpha} \|v_{test} - V\alpha\|_2^2 + \lambda \|\alpha\|_1 \tag{4}$$

Majumdar *et al.* [19] and Elhamifar *et al.* [20] have claimed that l_1-norm does not explicitly impose the sparsity constraint. Instead, it can be better enforced using supervised $l_{2,1}$-norm. Thus, the minimization changes to:

$$\min_{\alpha} \|v_{test} - V\alpha\|_2^2 + \lambda \|\alpha\|_{2,1} \tag{5}$$

Using SRC as basis, Group Sparse Representation based classification [8] aims to handle multiple data sources and features for each data point:

$$v_{test}^i = \alpha_{k,1}^i v_{k,1}^i + \alpha_{k,2}^i v_{k,2}^i + ... + \alpha_{k,n}^i v_{k,n}^i + \epsilon \tag{6}$$

where, v_{test}^i refers to i_{th} modality of test sample v_{test}.

The following subsection builds upon the pre-requisites and presents the proposed algorithm.

3.3. Proposed Algorithm

In this subsection, the proposed algorithm, Low Rank Group Sparse Representation based Classification (LR-GSRC) method for face recognition is presented. Some notations that are needed to facilitate further discussions are given below.

Let Y_s = [$Y_{1,s}$, $Y_{2,s}$,....,$Y_{c,s}$] contain all samples from source domain, with a total of N_s instances from c different classes. Hence, $Y_{i,s}\ \epsilon \in R^{n*ms_i}$, where n is the dimension of the samples and ms_i refers to the i_{th} class size in the Source Domain. Similarly, Y_t = [$Y_{1,t}$, $Y_{2,t}$,....,$Y_{c,t}$] contains samples from the target domain such that $Y_{i,t}\ \epsilon \in R^{n*mt_i}$. From Y_s, a dictionary D_j is learnt for each class. D = [D_1,D_2,..,D_c], where D_j is the dictionary for j^{th} class and $D_j\ \epsilon \in R^{n*p}$. Here p depicts the number of atoms in the dictionary.

For Group Sparse Classifier with i groups in the source domain, Y_t^i represents instances from target data belonging to i^{th} group and D_j^i represents dictionary from i^{th} group and j^{th} class. Our aim here is to incrementally learn Group Sparse coefficients and dictionary such that at k^{th} iteration dictionary $D_{*,k}$ is closer to the target domain as compared to the $k-1^{th}$ iteration dictionary. Here $D_{*,k}$ refers that dictionary D = [D_1,D_2,..,D_c] learned at k^{th} iteration for all classes c in the data.

3.3.1. Training

Given Y_t and Y_s (instances from target and source domain respectively), the algorithm is as follows:

Step 1: Learn the source dictionary $D_{*,o}$ using samples from Y_s. Using this dictionary as initial point, our aim is to incrementally learn Group Sparse coefficients α and target dictionary that gives the best representation for the target domain.

Step 2: Given source dictionary $D_{*,o}$, α for GSRC is learnt using the following formulation:

$$\min_{\alpha} \left\|Y_t^i - D_{*,k}^i \alpha^i\right\|_2^2 + \lambda \left\|\alpha^i\right\|_{2,1} + \sum_j \left\|D_{j,k}^i\right\|_* \tag{7}$$

here, $\|a\|_*$ refers to trace norm that is used as low rank regularization on dictionary. $D_{j,k}^i$ represents dictionary for i^{th} group and j^{th} class at k^{th} iteration.

Step 3: $D_{*,k}$ is updated for the next intermediate domain $k+1$ to incrementally adapt to the target data [9]. $D_{*,k+1}$ is learnt on the basis of its coherence with the dictionary in k^{th} domain and residual of instances in Y_t. The residual,$Z_{*,k}$ is obtained using the following:

$$X_{*,k} = \arg\min_X \|Y_t - D_{*,k}X\|_F^2 \,, s.t. \vee_i, \|p_i\|_o \leq T \tag{8}$$

$$Z_{*,k} = \|Y_t - D_{*,k}X_{*,k}\|_F^2 \tag{9}$$

here, $X_{*,k} = [p_1, ..., p_{N_t}]$ refers to the sparse coefficients of data instances in Y_t, obtained using the dictionary from k^{th} iteration. p_i refers to sparse coefficients of data instances belonging to class i. The updation in $D_{*,k}$ atoms, Δ $D_{*,k}$, to obtain $D_{*,k+1}$ is formulated using following minimization:

$$\min_{\Delta D_{*,k}} \|Z_{*,k} - \Delta D_{*,k}X_{*,k}\|_F^2 + \lambda \|\Delta D_{*,k}\|_F^2 \tag{10}$$

The first term is responsible for adjustments in atoms of dictionary $D_{*,k}$ in order to decrease the residual reconstruction error $Z_{*,k}$. The second term is used to control sudden changes in dictionary atoms between current domain and next domain. Hence, $D_{*,k+1}$ can be formulated using:

$$\Delta D_{*,k} = Z_{*,k}X_{*,k}^T(\lambda I + X_{*,k}X_{*,k}^T)^{-1} \tag{11}$$

$$D_{*,k+1} = D_{*,k} + \Delta D_{*,k} \tag{12}$$

The above two steps are repeated to learn intermediate representations till the best representative dictionary of the target data is obtained. This is enforced by a stopping criteria: $\|\Delta D_{*,k}\|_F < \delta$. This approach has been summarized in Algorithm 1.

Data: Source Dictionary $D_{*,o}$ learnt after Step 1, target data Y_t, sparsity level T,
Result: $D_{*,k}$ and α for intermediate domains initialization;
do
1. Learn Group Sparse coefficients α for Dictionary $D_{*,k}$ using equation (7)
2. Obtain $Z_{*,k}$ from Y_t and $D_{*,k}$ using equation (8) and (9)
3. Update atoms in $D_{*,k}$ to get next intermediate domain $D_{*,k+1}$ using (10), (11) and (12) ;

while $\|\Delta D_{*,k}\|_F < \delta$;

Algorithm 1: Low Rank GSRC

3.3.2. Testing

For a given test sample, following steps are followed:
1. For each class c, reconstruct a sample $v_{recon}(c)$ by the linear combination of training samples from that class:

$$v_{recon}(k) = V_k\alpha_k \tag{13}$$

2. Calculate error between the given test sample and reconstructed sample
3. Assign the test sample to the class having minimum reconstruction error

4. EXPERIMENTS AND RESULTS

Experiments are performed on a subset of CMU MultiPIE dataset consisting of 20 images per subject with varying pose and illumination conditions. Experiments have been performed under two setups and comparison has been drawn with existing approaches.

4.1. Experimental Setup - 1

For Setup - 1, a subset of CMU MultiPIE dataset having pose and Illumination variations was selected. The subset was further divided into source and target domain. The source domain was only used for training while the target domain was

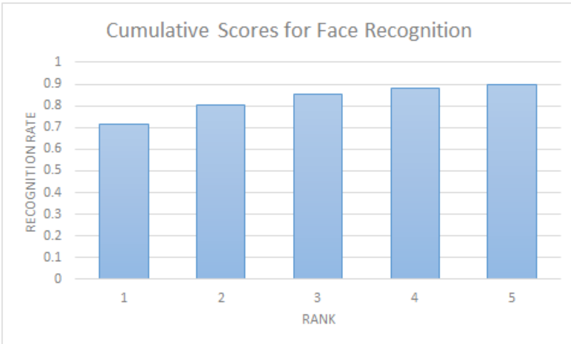

Fig. 2. Cumulative Face Recognition rate of proposed algorithm with multiple poses in target domain (Experimental Setup-1)

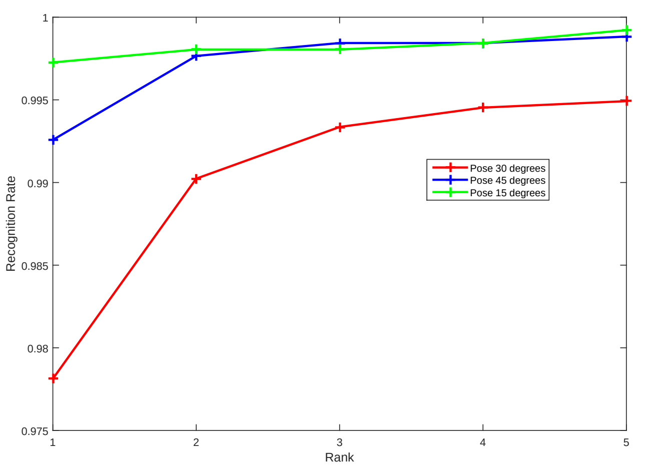

Fig. 3. Cumulative Matchscore Curves of proposed algorithm with each target domain (pose) individually (Experimental Setup-2)

divided into mutually exclusive training and testing sets. The source domain contained instances having pose variations of 0, 45 and -45 degrees. For each user, the source dictionary is learnt on this data. Similarly, target domain contained instances having pose variations of 30, 15 and -30 degrees, this also ensured that the training and testing instances were mutually exclusive. Using the proposed algorithm, the learnt dictionaries are adapted to get the best representation of the target domain. Testing of trained and adapted classifier was done on test data consisting of poses at 45, 30, 15, 0, -30 and -45 degrees.

4.2. Experimental Setup - 2

Some of the existing algorithms that have dealt with pose variations have presented their results on CMU Multi-PIE dataset however, on each target domain individually. Therefore, under Setup - 2 as opposed to Setup -1, separate domains are created for distinct pose variations. As mentioned above, the source domain contained instances from poses 0, 45 and -45 degrees on which the source dictionary is learned. Target domain consists of images having pose variations of 30, 15 and -30 degrees, individually. The source dictionaries are separately adapted for the target domains.

4.3. Results

Results obtained on experimental setup - 1 and setup - 2 have been summarized below:

- Rank-1 accuracy of 71.45% is obtained for Experimental Setup - 1. The cumulative face recognition scores for this experiment are given in Figure 2.
- Table 1 presents the results obtained using Experimental Setup - 2 and it's comparison with some existing state-of-the-art algorithms for pose variations. It is observed that the proposed algorithm outperforms the existing state-of-the-art algorithms for poses at 15 and 45 degrees by obtaining an accuracy of 99.72% and 99.25% respectively. Also, it performs well for pose variations of 30 degrees by reporting an accuracy of 97.81%. The cumulative match curves obtained for the given setup are given in Figure 3.
- Results obtained from Setup-1 motivate the use of the proposed algorithm even when the target domain has a mix of varying poses, i.e. it eliminates the need to address each pose variation individually.

Method	15^o	30^o	45^o
GMLDA [21]	99.7	99.2	98.6
FDDL [22]	96.8	90.6	94.4
SDDL [23]	98.4	98.2	98.9
LR-GSRC	**99.72**	97.81	**99.25**

Table 1. Comparison of proposed algorithm with other Domain adaptation algorithms for pose

These results motivate the use of trace-norm with group sparse classifier for incremental dictionary learning. The performance of the proposed algorithm, under different experimental setups for face recognition with varying pose and illumination conditions, also suggests the usage of LR-GSRC for domain adaptation.

5. CONCLUSION

In this research, a novel framework for addressing the problem of face recognition with pose and illumination variation has been proposed. The algorithm, low rank group sparse representation based classifier, learns group-sparse coefficients on low-rank dictionaries with incremental learning. Results under multiple experimental setups on the CMU Multi-PIE dataset support the effectiveness of the algorithm.

6. REFERENCES

[1] Anil K Jain and Stan Z Li, *Handbook of face recognition*, vol. 1, Springer, 2005.

[2] Himanshu S Bhatt, Richa Singh, Mayank Vatsa, and Nalini K Ratha, "Improving cross-resolution face matching using ensemble-based co-transfer learning," *IEEE Transactions on Image Processing*, vol. 23, no. 12, pp. 5654–5669, 2014.

[3] Jongsun Kim, Jongmoo Choi, Juneho Yi, and Matthew Turk, "Effective representation using ica for face recognition robust to local distortion and partial occlusion," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 27, no. 12, pp. 1977–1981, 2005.

[4] Meng Yang, Lei Zhang, Jian Yang, and David Zhang, "Robust sparse coding for face recognition," in *IEEE Conference on Computer Vision and Pattern Recognition*. IEEE, 2011, pp. 625–632.

[5] Florian Schroff, Dmitry Kalenichenko, and James Philbin, "Facenet: A unified embedding for face recognition and clustering," in *IEEE Conference on Computer Vision and Pattern Recognition*, June 2015.

[6] Y. Taigman, Ming Yang, M. Ranzato, and L. Wolf, "Deepface: Closing the gap to human-level performance in face verification," in *IEEE Conference on Computer Vision and Pattern Recognition*, June 2014, pp. 1701–1708.

[7] Ralph Gross, Iain Matthews, Jeffrey Cohn, Takeo Kanade, and Simon Baker, "Multi-pie," *Image and Vision Computing*, vol. 28, no. 5, pp. 807–813, 2010.

[8] Gaurav Goswami, Paritosh Mittal, Angshul Majumdar, Mayank Vatsa, and Richa Singh, "Group sparse representation based classification for multi-feature multimodal biometrics," *Information Fusion*, 2015.

[9] Qiang Qiu, Jie Ni, and Rama Chellappa, "Dictionary-based domain adaptation methods for the re-identification of faces," in *Person Re-Identification*, pp. 269–285. Springer, 2014.

[10] Zaid Harchaoui, Matthijs Douze, Mattis Paulin, Miroslav Dudik, and Jérôme Malick, "Large-scale image classification with trace-norm regularization," in *IEEE Conference on Computer Vision and Pattern Recognition*. IEEE, 2012, pp. 3386–3393.

[11] Raghunandan Keshavan, Andrea Montanari, and Sewoong Oh, "Matrix completion from noisy entries," in *Advances in Neural Information Processing Systems*, 2009, pp. 952–960.

[12] Georgios Passalis, Panagiotis Perakis, Theoharis Theoharis, and Ioannis A Kakadiaris, "Using facial symmetry to handle pose variations in real-world 3d face recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 33, no. 10, pp. 1938–1951, 2011.

[13] Zhenyao Zhu, Ping Luo, Xiaogang Wang, and Xiaoou Tang, "Deep learning identity-preserving face space," in *IEEE International Conference on Computer Vision*, 2013, pp. 113–120.

[14] Richa Singh, Mayank Vatsa, Arun Ross, and Afzel Noore, "Performance enhancement of 2d face recognition via mosaicing," in *Fourth IEEE Workshop on Automatic Identification Advanced Technologies*. IEEE, 2005, pp. 63–68.

[15] Qiang Zhang and Baoxin Li, "Discriminative k-svd for dictionary learning in face recognition," in *IEEE Conference on Computer Vision and Pattern Recognition*. IEEE, 2010, pp. 2691–2698.

[16] Long Ma, Chunheng Wang, Baihua Xiao, and Wen Zhou, "Sparse representation for face recognition based on discriminative low-rank dictionary learning," in *IEEE Conference on Computer Vision and Pattern Recognition*. IEEE, 2012, pp. 2586–2593.

[17] Michal Aharon, Michael Elad, and Alfred Bruckstein, "K-svd: An algorithm for designing overcomplete dictionaries for sparse representation," *IEEE Transactions on Signal Processing*, vol. 54, no. 11, pp. 4311–4322, 2006.

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

[19] Angshul Majumdar and Rabab K Ward, "Robust classifiers for data reduced via random projections," *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics*, vol. 40, no. 5, pp. 1359–1371, 2010.

[20] Ehsan Elhamifar and René Vidal, "Robust classification using structured sparse representation," in *IEEE Conference on Computer Vision and Pattern Recognition*. IEEE, 2011, pp. 1873–1879.

[21] Abhishek Sharma, Abhishek Kumar, Hal Daume III, and David W Jacobs, "Generalized multiview analysis: A discriminative latent space," in *IEEE Conference on Computer Vision and Pattern Recognition*. IEEE, 2012, pp. 2160–2167.

[22] Meng Yang, Lei Zhang, Xiangchu Feng, and David Zhang, "Fisher discrimination dictionary learning for sparse representation," in *IEEE International Conference on Computer Vision,2011*. IEEE, 2011, pp. 543–550.

[23] Sumit Shekhar, Vishal Patel, Hien Nguyen, and Rama Chellappa, "Generalized domain-adaptive dictionaries," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2013, pp. 361–368.