Significance of Dictionary for Sparse Coding Based Pose Invariant Face Recognition

Shejin Thavalengal, Srimanta Mandal and Anil Kumar Sao
School of Computing and Electrical Engineering
Indian Institute of Technology Mandi, India
Email: in.shejin.t@ieee.org, srimanta_mandal@students.iitmandi.ac.in, anil@iitmandi.ac.in

***Abstract*—This paper deals with the dictionary for sparse coding based pose invariant face recognition. A two stage face recognition system is proposed in sparse coding frame work. The first stage is designed to recognize the pose of an incoming test image. It is proposed to perform pose classification as a three class classification problem in sparse coding frame work, each class representing images for frontal, left side or right side pose. This is done to choose the appropriate Weighted Decomposition Face (WD Face) dictionary, which is found to be suitable for sparse coding based face recognition. The second stage obtains the identity of the person with the help of pose specific WD Face dictionary chosen in the first stage. Experimental results have shown that the proposed approach is a simple yet robust sparse coding based face recognition system which is invariant of pose and illumination.**

I. Introduction

Even though face recognition has been widely studied for the last two decades [1], it still poses as an interesting as well as challenging field in the literature of computer vision due to the never ending demand for efficient and faster face recognition systems. Recently sparse coding based approaches have been introduced and studied in the context of face recognition [2]–[4]. These approaches denote a test face image of a person ($\mathbf{y}_i$) as a linear combination of the training images of the same person, i.e,

$$\mathbf{y}_i = \sum_{j=1}^{n_i} a_{i,j}\mathbf{v}_{i,j}, \tag{1}$$

where n_i is the number of training images available for the i^{th} person, $\mathbf{v}_{i,j} \in \mathbb{R}^m$ is the j^{th} training face image of the same person and corresponding coefficient is denoted by $a_{i,j}$. In a more general way, equation(1) can be rewritten as

$$\mathbf{y} = \mathbf{D}\mathbf{a}, \tag{2}$$

where $\mathbf{D} \in \mathbb{R}^{m\times n}$ represents the dictionary containing all the available n training face images of all the persons, and vector $\mathbf{a} \in \mathbb{R}^n$ contains zero corresponding to those face images in dictionary which do not belong to the same class as the test face image.

Information related to sparse coding based face recognition and reconstruction of coefficient vector $\mathbf{a}$ is available in references [2]–[10]. Wright *et al.* [2] introduced sparse coding based face recognition across varying illumination and partial occlusion. It has been shown that dictionary $\mathbf{D}$ is no longer important if a large number of training images are available. Deng *et al.* [3] approached sparse coding based face recognition by generating an intra-class dictionary, which contains all possible class variations. This work was focused on solving the partial occlusion problem in face recognition. An illumination and alignment invariant face recognition system is proposed by Wagner *et al.* [4]. These methods require all possible misalignment, illumination or occlusion, which might occur during testing of the image. This issue was addressed by deriving training samples from a generic model of face image [3]. This may not be appropriate in real world face recognition. The derived samples might smear person specific information of face image or may add some artifacts. Moreover, if the modeling of face is not proper or fails to address all possible variations, the system may not provide a reliable result. Hence, we emphasize on extraction of subject specific features which could be the reason for discriminating between two people. These features should not be affected by illumination or pose variations.

Our previous work [5] highlighted the significance of dictionary when a very small number of training images is available. We have proposed a dictionary named WD Face (Weighted Decomposition of Face) dictionary to overcome the sufficiency in training images [5]. This dictionary accomplishes the goal by emphasizing subject specific unique information of a face image, which is crucial for discriminating between two people's face images. The underlying physical and mathematical nature of WD Face dictionary was brought to light in another previous work [8]. WD Face dictionary is proved to overcome the issue of sufficiency of the training data as well as illumination variations. But, this approach cannot be used directly to address the issue of pose variation in face recognition. This is because, WD Face dictionary requires computation of eigenvectors of the covariance matrix obtained from all the available training images. Hence WD Face transformation of face image with pose variation cannot obtain a lower dimensional subspace which emphasizes the information unique to a person.

This paper deals with the dictionary for sparse coding based pose invariant face recognition. A two stage face recognition system is proposed in sparse coding framework. An overview of the proposed approach is given in Fig. 1. Here the pose of incoming test image is approximated initially. This is carried

978-1-4799-2361-8/14/$31.00

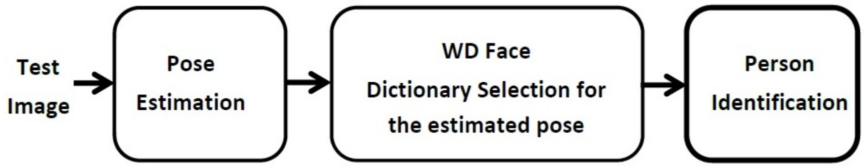

Fig. 1. Flow diagram of the proposed method.

out to choose WD Face dictionary appropriate for the pose of the test image. A pose specific WD Face dictionary is selected for the detected pose. This pose specific WD Face dictionary is generated from the training images belonging to the pose class corresponding to the test image. Person identification is done in the second stage using the selected pose specific WD Face dictionary in sparse coding frame work.

Rest of the paper is organized as follows: Section II explains a pose detection system based on sparse coding, which is the first stage of the proposed face recognition system. Face identification method as a successor of Section II is explained in Section III. Experimental results obtained for the proposed system are evaluated in Section IV. Section V gives a summary of the paper.

II. Pose Detection Based on Sparse Coding

One of the main challenges in face recognition is the deviation of training and test images due to their pose variation. Moreover, in sparse coding based face recognition approach, if the images have pose variations, it is more likely that the test image is classified to a class with similiar pose, that is, while matching, the pose variation plays a major role as compared to the identity of the test image. If one can detect the pose of an incoming test image successfully, the chance of classifying the image in to the correct class is higher. In this context, a classifier based on sparse coding is proposed to detect the pose and classify the test image in to one of the target poses.

Identification of pose of a given test image $\mathbf{y}_p$ is performed by sparse coding based classification as

$$\hat{\mathbf{y}}_p = \hat{\mathbf{D}}_p \mathbf{a}_p, \tag{3}$$

where $\hat{\mathbf{y}}_p = \mathbf{T}\mathbf{y}_p$ and $\hat{\mathbf{D}}_p = \mathbf{T}\mathbf{D}_p$. Here $\mathbf{D}_p$ represents the dictionary containing all the images corresponding to different pose classes and vector $\mathbf{a}_p$ contains zero corresponding to those face images in dictionary which do not belong to the same pose class as the test face image. Here transform operator ($\mathbf{T}$) is deployed to satisfy the under-determined condition of dictionary. The sparse vector $\mathbf{a}_p$ can be obtained by solving,

$$\min \|\mathbf{a}_p\|_1 \text{ subject to } \|\hat{\mathbf{y}}_p - \hat{\mathbf{D}}_p \mathbf{a}_p\|_2 \leq \epsilon_p, \tag{4}$$

where ϵ_p is a small number which represents an error tolerance. A global dictionary ($\hat{\mathbf{D}}_p$) consisting of all the pose variations is used for this purpose. In this formulation, sparse coding framework is used for pose classification rather than person identification.

In our experiments, we have made three classes for pose classification, namely, (i) frontal class, (ii) right side pose and (iii) left side pose. The frontal class is responsible for the frontal images, ideally 0° pose but can accommodate a slight deviation. This class is trained using images containing pose variation from $-6°$ to $+6°$. Class labeled left side pose is trained using images of pose $-25°$ to $-6°$. In a similar manner, training of class named right side pose is done with images of pose $+6°$ to $+25°$. These three classes together can represent the whole database roughly. A total of 320 images were selected from 8 people in 'pose' subset of CUbiC FacePix database for training [11], [12]. In pose classification, dictionary is generated using downsampling transformation as $\mathbf{T}$, which downsamples the training images to 14×18 dimension. Since there is a sufficient number of images for each of the three classes, as discussed earlier, the transformed dictionary created using downsampling transformation will perform as well as any other dictionary.

Face images of the remaining 22 people in the pose subset of CUbiC FacePix database comprising of pose variation $-30°$ to $+30°$ are used to test the proposed pose classification system. The proposed approach produces an average accuracy of 83.3% for pose classification, which is comparable (81.4%) to the methods shown in [13]. It has to be noted that we classified images in to three classes without considering any misclassification tolerance for those images which lie on the boundary of the pose classes. High accuracy of pose classification could not be achieved due to the reason that the images of left side and right side pose which share the boundary with frontal class do not deviate much in pose angle. In other words, the three classes are continuous with 1° interval and hence ideally we can not draw a sharp boundary between frontal class and left pose class as well as frontal class and right pose class. Moreover, the aim here is to estimate the pose roughly in order to choose the appropriate dictionary for the incoming test image, and hence high accuracy in pose classification is not likely to affect the overall performance of the face recognition system.

III. Identity Detection Using Pose Specific Dictionary

Once the pose has been detected, the next challenge is to obtain the identity of the test image. In the work [5], a dictionary is proposed which emphasizes the unique characteristics of a face image, which actually is the identity of the person. This dictionary is termed as WD Face dictionary and can be

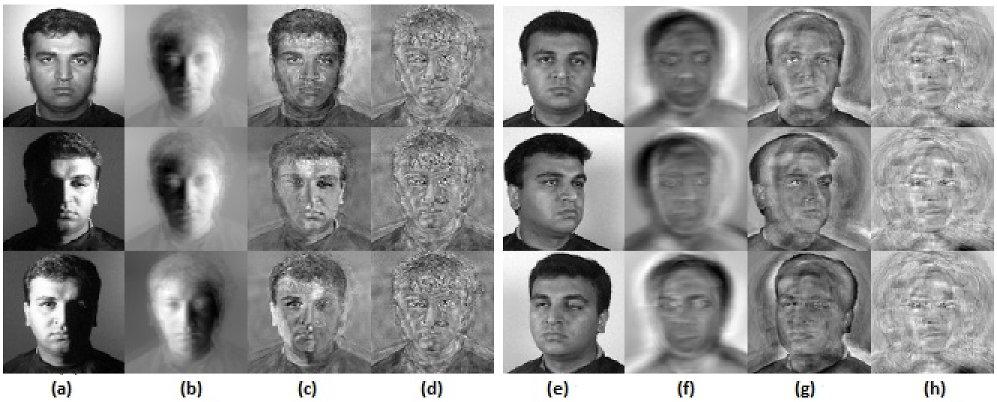

Fig. 2. Illustration of face decomposition in to different components using a pose specific dictionary, (a) Gray level image. The reconstructed face image using (b) first 10 eigenvectors, (c) 11-350 eigenvectors, (d) 351-remaining eigenvectors. Illustration of face decomposition in to different components using a global dictionary: (e) Gray level image. The reconstructed face image using (f) first 10 eigenvectors, (g) 11-350 eigenvectors, (h) 351-remaining eigenvectors.

generated using the transformation $\mathbf{T} = \mathbf{W}\mathbf{\Psi}^T$, that is,

$$\hat{\mathbf{D}} = \mathbf{W}\mathbf{\Psi}^T\mathbf{D}, \tag{5}$$

where $\hat{\mathbf{D}}$ is the generated dictionary, $\mathbf{W}$ is a weight matrix and $\mathbf{\Psi}$ is t he matrix consist of eigen vectors obtained from the eigen analysis of the set of training images $\mathbf{D}$. WD Face dictionary relays on the fact that a face image can be decomposed in to three components namely common component, noise component and the component which carry a person's unique information as shown in Fig. 2. Higher weights are given to these unique components in WD Face dictionary. This can be done with the weight matrix $\mathbf{W} = \mathbf{\Lambda}^{-1/2}$ where $\mathbf{\Lambda}$ is a diagonal matrix contains the eigenvalues corresponding to the eigenvector matrix $\mathbf{\Psi}$. It has to be taken care to remove the least significant eigen components before ascertaining weights. It has been noticed that dictionary created in this manner is highly subject specific and illumination invariant. Moreover, WD Face dictionary is a whitening transform, where the rows of dictionary are uncorrelated [8]. All these properties make WD Face dictionary an ideal candidate for sparse coding based face recognition.

For identity detection, three dictionaries (left pose, frontal and right pose) are generated using WD Face representation. This is because a common dictionary created will contain all ranges of pose variation. WD Face dictionary demands computation of eigenvectors of the covariance matrix obtained from all available training images. Due to wide variety of pose deviation, approximation of these images by projecting in to a lower dimensional subspace won't be accurate. Moreover, while computing the eigenvectors, the variation due to pose deviation will be dominant as compared to the subtle variation of face image. Because of all these reasons, the idea of a global dictionary is not viable in case of large pose variation. The significance of pose specific dictionary can be observed from Fig. 2. Fig. 2(c) shows person specific unique information, which has been given higher weights in WD Face dictionary creation.When we are using a pose specific dictionary as in Fig. 2 (This dictionary accounts for frontal pose), the higher weighted component actually emphasizes person specific information. But in the case of a global dictionary, which contains all the pose variations, the images represented by WD Faces (Fig. 2(g)) don't emphasize much critical information which helps in discriminating the face images of two persons.

The detailed algorithm is shown in Table I. Moreover, a typical example is shown in Fig. 3 for the proposed two stage face recognition system in terms of the sparse coefficients obtained. A test image corresponding to right side pose is given as input to the system. Sparse coefficients obtained in the **STAGE1** clearly show that the image belongs to right side pose class and WD Face dictionary corresponding to right side pose is used. In **STAGE2**, it can be noted that, coefficients corresponding to the true class obtained using WD Face dictionary have higher values while the coefficients corresponding to impostor classes are close to zero and hence, identity of the person is obtained.

IV. Comparison of Results

The importance of the proposed approach is explored in the pose variation subset of CUbiC FacePix Database [11], [12]. WD Face dictionary is generated as explained in [5]. Only 100 significant eigen components are used in the experiments. For identity detection, three separate dictionaries for three different pose classes are generated, namely frontal pose, right side pose and left side pose. Frontal pose dictionary consists of 4 images of each person randomly chosen from his/her images of pose deviation $-6°$ to $+6°$, hence a total of 120 images are used for training from the whole database. This choice of four images for frontal class is based on our previous experiments (detailed in [8]) which show that WD Face dictionary can attain high accuracy of classification in FacePix database with as few as four images per person. Right side pose dictionary was derived by randomly choosing 10 images per person from images of pose deviation from $-25°$ to $-7°$, which will create a total of 300 images for training. Similarly dictionary for left side

TABLE I. Sparse Coding Based Pose Invariant Face Recognition

Algorithm for sparse coding based pose invariant face recognition
STAGE1 Step: Initialize Dictionary $\mathbf{D} = \mathbf{D}_p$, where $\mathbf{D}_p$ is the global pose detection dictionary, test image $\mathbf{y} \in \mathbb{R}^m$ as a column vector, downsampling transform operator $\mathbf{T} \in \mathbb{R}^{r \times m}$, threshold ϵ Label:**MAIN ITERATION** Step: Normalize $\mathbf{y}$ and columns of $\mathbf{D}$ to unit length Step: Transform $\hat{\mathbf{y}} = \mathbf{Ty}$ and $\hat{\mathbf{D}} = \mathbf{TD}$ Step: Solve $\min \|\mathbf{a}\|_1$ subject to $\|\hat{\mathbf{y}} - \hat{\mathbf{D}}\mathbf{a}\|_2 \leq \epsilon$ Step: Compute residual $r_i(\mathbf{y}) = \|\hat{\mathbf{y}} - \hat{\mathbf{D}}\delta_{\mathbf{i}}(\mathbf{a})\|_{\mathbf{2}}$ for $i = 1, 2, ..k$, where $\delta_i(\mathbf{a}) \in \mathbb{R}^n$ is a vector whose non-zero entries are the entries in $\mathbf{a}$ that are associated with class i **MAIN ITERATION** ends Step: Detect the pose, $I_p(\mathbf{y}) = \arg\min_i r_i(\mathbf{y})$, where $I_p(\mathbf{y})$ is the detected pose class **STAGE2** Step: Initialize Dictionary $\mathbf{D}$ as a pose specific dictionary chosen according to the pose detected in **STAGE1**, test image $\mathbf{y} \in \mathbb{R}^m$ as a column vector, WD Face transform operator $\mathbf{T} \in \mathbb{R}^{r \times m}$, threshold ϵ Call **MAIN ITERATION**
$\mathrm{I}(\mathbf{y}) = \arg\min_i r_i(\mathbf{y})$, where $\mathrm{I}(\mathbf{y})$ is the identity of the given test image $\mathbf{y}$

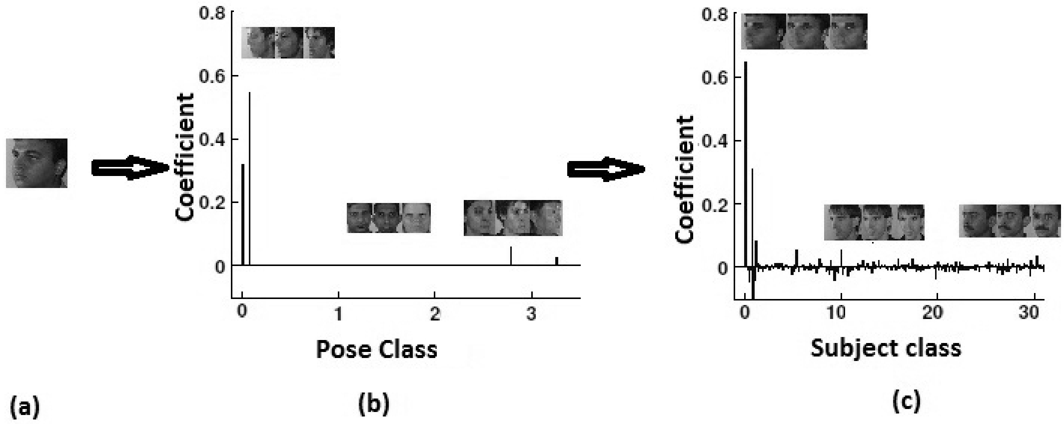

Fig. 3. An example for proposed sparse coding based pose invariant face recognition : (a) Test Image, (b) **STAGE1**: Pose Detection using sparse coding based approach, (c) **STAGE2** : Person Identification using sparse coding based approach.

pose is derived from images of pose deviation +7° to +25°. The purpose of choosing 10 images for right side and left side pose each is to cover the large range of pose variations which are present in these two pose classes as compared to the frontal pose class.

This system was tested against images of pose deviation −30° to +30° excluding training images, i.e., a total of 1110 images were used for testing. For comparing the performance of different dictionaries in this system, the experiments were repeated with dictionaries generated using downsampled face images, Randomfaces and Eigenfaces. DS (Down-sampled) Face dictionary uses images downsampled to 10×10 and randomface dictionary is generated by using transformation matrix $\mathbf{T}$ (as explained in Table I) as a random matrix of size $100 \times m$ (Note that the image of size $\sqrt{m} \times \sqrt{m}$ is converted to a vector of $m \times 1$ prior to projecting in to the random matrix) [2]. Further information related to Randomfaces and Eigenfaces can be obtained from references [2] and [14] respectively.

Further, experiments were extended to the Sheffield (previously UMIST) Face database [15]. Sheffield database consists of 564 images of 20 individuals varying from frontal view to left side pose. Hence this database comprises of pose class of frontal and left side pose only. Seven images per person (total 140 images) are chosen randomly for training and the rest are used for testing. Results obtained for different dictionaries are shown in Table II. From the table it can be observed that WD Face dictionary outperforms its competitors significantly. Further validation of this system in different face databases weren't possible due to the scarcity of publicly available face databases which contains large number of pose variations for the success of our algorithm.

TABLE II. Face Recognition Results for Different Dictionaries under Pose Variation

Dictionary	FacePix	Sheffield
Down Sampled Face	47.7%	97.31%
Randomface	38.9%	92.69%
Eigenface	93.6%	99.23%
WD Face	**95.4%**	**99.62%**

V. Summary

In this paper, the problem of pose variation in sparse coding based face recognition was addressed. The previously proposed WD Face representation could not be used directly to generate a global dictionary which contains all the pose variations, since it requires computation of eigenvectors of the covariance matrix obtained from all the available training images. Hence WD Face transformation of face image with pose variation cannot obtain a lower dimensional subspace which emphasizes the person specific information. This issue is addressed by the proposed two stage approach. In the first stage, the incoming test face image is classified in to one of the target pose classes and WD Face dictionary is chosen for the particular pose class. The second step obtains the identity of test face using sparse coding frame work with the help of the pose specific dictionary chosen in the first stage. The approach of choosing a pose specific WD Face dictionary is supported by the experimental results. The proposed approach can be further improved by generating more pose classes and hence more WD Face dictionaries for classification.

References

[1] W. Zhao, R. Chellappa, P. J. Phillips, and A. Rosenfeld, "Face recognition: A literature survey," *ACM Computer Survey*, vol. 35, no. 4, pp. 399–458, Dec. 2003.

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

[3] W. Deng, J. Hu, and J. Guo, "Extended SRC: Undersampled face recognition via intraclass variant dictionary," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 34, no. 9, pp. 1864–1870, Sep. 2012.

[4] A. Wagner, J. Wright, A. Ganesh, Z. Zhou, H. Mobahi, and Y. Ma, "Toward a practical face recognition system: Robust alignment and illumination by sparse representation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 34, no. 2, pp. 372–386, 2012.

[5] T. Shejin and A. K. Sao, "Significance of dictionary for sparse coding based face recognition," in *Proceedings of 11th International Conference of the Biometrics Special Interest Group*, Darmstadt, Germany, Sep. 2012, pp. 1 –6.

[6] D. Donoho, "Compressed sensing," *IEEE Transactions on Information Theory*, vol. 52, no. 4, pp. 1289 –1306, Apr. 2006.

[7] E. Candes, J. Romberg, and T. Tao, "Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information," *IEEE Transactions on Information Theory*, vol. 52, no. 2, pp. 489 – 509, Feb. 2006.

[8] S. Thavalengal and A. K. Sao, "Analysis of WD face dictionary for sparse coding based face recognition," in *Proceedings of 17th International Conference on Image Analysis and Processing, Lecture Notes in Computer Science*, vol. 8156, Naples, Italy, Sep. 2013, pp. 221–230.

[9] J. Tropp and A. Gilbert, "Signal recovery from random measurements via orthogonal matching pursuit," *IEEE Transactions on Information Theory*, vol. 53, no. 12, pp. 4655 –4666, Dec. 2007.

[10] S. J. Kim, K. Koh, M. Lustig, S. Boyd, and D. Gorinevsky, "An interior-point method for large-scale l_1-regularized least squares," *IEEE Journal of Selected Topics in Signal Processing*, vol. 1, no. 4, pp. 606 –617, Dec. 2007.

[11] J. A. Black, Jr., M. Gargesha, K. Kahol, P. Kuchi, and S. Panchanathan, "A framework for performance evaluation of face recognition algorithms," in *ITCOM, Internet Multimedia Systems II*, Jul. 2002, pp. 163–174.

[12] D. Little, S. Krishna, J. Black, and S. Panchanathan, "A methodology for evaluating robustness of face recognition algorithms with respect to variations in pose angle and illumination angle," in *IEEE International Conference on Acoustics, Speech, and Signal Processing*, vol. 2, Philadelphia, USA, Mar. 2005, pp. 89 – 92.

[13] A. Dahmane, S. Larabi, C. Djeraba, and I. M. Bilasco, "Learning symmetrical model for head pose estimation," in *21st International Conference on Pattern Recognition*, Tsukuba, Japan, Nov. 2012, pp. 3614 –3617.

[14] M. Turk and A. Pentland, "Face recognition using Eigenfaces," in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, Hawaii, USA, Jun. 1991, pp. 586 –591.

[15] U. of Sheffield, "University of sheffield face databse," http://www.sheffield.ac.uk/eee/research/iel/research/face.