A Novel Technique For Face Recognition Across Variable Illuminations And Poses

Ms. Khushboo B. Trivedi,
M.E.Final Year Appearing,
Dept. of Information Technology,
Sipna College of Engineering & Technology,
Amravati.

Prof. V. T. Gaikwad,
Associate Professor,
Dept. of Information Technology,
Sipna College of Engineering & Technology,
Amravati.

Prof. H. N. Datir
Assistant Professor
Dept.of Computer Science and Engineering,
Sipna College of Engineering & Technology,
Amravati.

Abstract

In this paper a face recognition algorithm based on simultaneous sparse approximations under varying illumination and pose is given. A dictionary is learned for each class based on given training examples which minimizes the representation error with a sparseness constraint. A novel test image is projected onto the span of the atoms in each learned dictionary. The resulting residual vectors are then used for classification. To handle variations in lighting conditions and pose, an image relighting technique based on pose-robust albedo estimation is used to generate multiple frontal images of the same person with variable lighting. As a result, the proposed algorithm has the ability to recognize human faces with high accuracy even when only a single or a very few images per person are provided for training.

Index Terms-Biometrics, dictionary learning, illumination variation, Albedo, relighting, simultaneous sparse signal representation.

1. INTRODUCTION

As one of the most successful applications of image analysis and understanding, face recognition has recently received significant attention, especially during the past several years. At least two reasons account for this trend: the first is the wide range of commercial and law enforcement applications, and the second is the availability of feasible technologies after 30 years of research. Even though current machine recognition systems have reached a certain level of maturity, their success is limited by the conditions imposed by many real applications. For example, recognition of face images acquired in an outdoor environment with changes in illumination and/or pose remains a largely unsolved problem. In other words, current systems are still far away from the capability of the human perception system [1]. In face recognition system there are some training images and a test image. Training images are those images which have specific controlled conditions, while test image is that image which is used to match with given training image for recognition. Current systems work very well when the test image is captured under controlled conditions. The performance degrades significantly when the test image contains variations that are not present in the training images. Some of these variations include illumination, pose, expression, cosmetics and aging.

In recent years, the theories of Sparse Representation (SR) and Compressed Sensing (CS) have emerged as powerful tools for efficiently processing data in non-traditional ways. This has led to recovery in interest in the principles of SR and CS for face recognition [2, 3, 4, 5, 6]. Phillips [2] proposed matching pursuit filters for face feature detection and identification. The filters were designed through a simultaneous decomposition of a training set into a 2D wavelet expansion designed to discriminate among faces. It was shown that the resulting algorithm was robust to facial expression and the surrounding environment. Wright *et al.* [3] introduced an algorithm, called Sparse Representation based Classification (SRC), where the training face images constitute the dictionary and a test image is classified by finding its sparse representation with respect to this dictionary. This work was later extended to handle pose and illumination variations [4], [5]. Also, an expression-invariant face recognition method based on ideas from the distributed compressed sensing and joint sparsity models was proposed in [6].

There are a number of hurdles that face recognition systems based on sparse representation must overcome. One is designing algorithms that are robust to changes in illumination; a second is that algorithms need to efficiently scale as the number of people enrolled in the system increases. The SRC approach recognizes faces by solving an optimization problem over the set of images enrolled into the database. This solution trades robustness and size of the database against computational efficiency.

In this paper, an algorithm to perform face recognition across varying illumination based on learning class specific dictionaries. Using a relighting method, many elements to the dictionary can be added so that robustness to illumination changes can be realized. This method consists of two stages. In the first stage, given training samples from each class, class specific dictionaries are trained with some fixed number of atoms. Then, a test image is projected onto the span of the atoms in each learned dictionaries. The residual vectors are then used for classification. The paper is divided in following way. The dictionary based face recognition algorithm is detailed in Section 2. Section 3 presents experimental results and Section 4 concludes the paper with a brief summary and discussion.

2. DICTIONARY-BASED RECOGNITION

Let $\tilde{D} = [d_1, \cdot \cdot \cdot, d_K] \in R^{NxK}$ be a redundant dictionary with K atoms represented as columns $dj \in R^N$ with $K \gg N$. The choice of dictionary usually depends on the specific application. A dictionary may be chosen such that it favors sparse approximations or it can be chosen to resemble the structure that may appear in the input samples. For face recognition, in [2] the dictionary contained steerable wavelet bases elements. while in [3] the dictionary consisted of the gallery images .

Given a data matrix $B = [x_1, \cdot \cdot \cdot, x_m] \in R^{Nxm}$ and a fixed dictionary $\tilde{D} \in R^{NxK}$, simultaneous sparse approximation attempts to find a matrix Γsuch that $B \simeq D\Gamma$. Where $D \in R^{NxP}$, P < N, is a dictionary matrix whose atoms are selected from $\tilde{D}$ and $\Gamma = [\gamma_1, \cdot \cdot \cdot, \gamma_m]$ is the matrix whose columns γ_i are the coefficients corresponding to each data vector x_i. In other words, simultaneous sparse approximation attempts to approximate all the samples in B at once as a linear combination of a common subset of atoms with cardinality much smaller than N. In fact, by keeping the sparsity low enough, one can eliminate the internal variation of the samples in B which may lead to more robust representation. It has been shown that instead of using a predetermined dictionary, learning dictionaries from the training data provides much better representation and hence can improve the performance of reconstructive approach to discrimination.

Learning Class Specific Reconstructive Dictionaries:

Designing dictionaries based on training is a much recent approach to dictionary design which is strongly motivated by the advances in the sparse representation theory [7]. The K-SVD [8] algorithm for learning dictionaries for face images.

Given a set of examples $B = [x_1, \cdot \cdot \cdot, x_m]$, the goal of the K-SVD algorithm is to find a dictionary D and a sparse matrix Γ that minimize the following representation error

$$(\hat{D}, \hat{\Gamma}) = \underset{D,\Gamma}{\arg\min} \left\| B - D\Gamma \right\|_F^2 \text{ subject to } \forall i \ \left\|\gamma i\right\| 0 \le T_0 \quad (1)$$

where γi represent the columns of Γ and the ℓ0 sparsity measure $\|.\|_0$ counts the number of nonzero elements in the representation. Here, $\|A\|_F$ denotes the Frobenius norm defined as $\|A\|_F = \sqrt{\Sigma\, ij Aij}$. The K-SVD algorithm alternates between sparse-coding and dictionary update steps. In the sparse-coding step, D is fixed and the representation vectors γ_is are found for each example x_i. Then, the dictionary is updated atom-by-atom in an efficient way.

Classification based on Learned Dictionaries:

Suppose that C distinct face classes and a set of m_i training images per class, $i \in \{1, \cdot \cdot \cdot, C\}$ are given. An $l \times q$ grayscale image as an N-dimensional vector, x, which can be obtained by stacking its columns, where $N = l \times q$. Let $B_i = [x_{i1}, \cdot \cdot \cdot, x_{imi}] \in R^{N\times m}i$ be an $N \times m_i$ matrix of training images corresponding to the i^{th} class.

For training, C class specific dictionaries is being learned, Di, to represent the training samples in each Bi, with some sparsity level T0, using the K-SVD algorithm. Once the dictionaries have been learned for each class, given a test sample y, to be projected onto the span of the atoms in each Di using the orthogonal projector $Pi = D_i(D_i^T D_i)^{-1} D_i^T$. The approximation and residual vectors can then be calculated as

$$\hat{y}^i = P_i y = D_i \alpha^i \quad (2)$$

and

$$r^i(y) = y - \hat{y}^i = (I - Pi)y, \quad (3)$$

respectively, where I is the identity matrix and $\alpha^i = (D_i^T D_i)^{-1} D_i^T y$ are the coefficients. Since the K-SVD algorithm finds the dictionary, I, that leads to the best representation for each examples in Bi, $\|r^i(y)\|_2$ can be small if y were to belong to the i^{th} class and large for the other classes. Based on this, one can classify y by assigning it to the class, $d \in \{1, \cdot \cdot \cdot, C\}$, that gives the lowest reconstruction error, $\|r^i(y)\|_2$:

$$d = \text{identity}(y) = \arg\min_i \|r^i(y)\|_2 \quad (4)$$

Fig 1. Shows example of how DFR algorithm works.

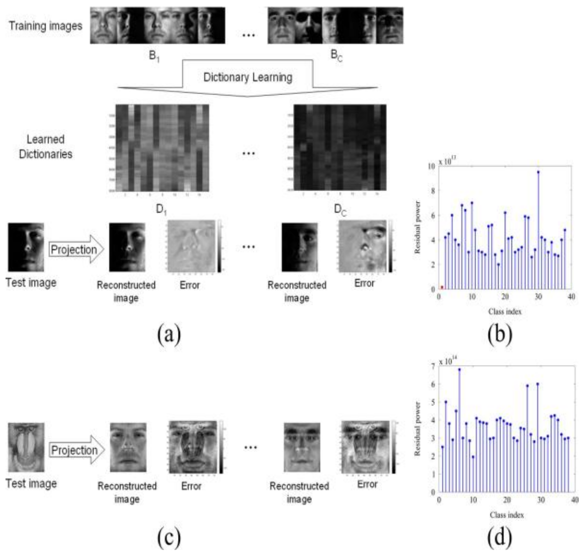

Fig 1. Exapmle of working of DFR algorithm

2.1. Image Relighting

Recognizing faces under varying illumination given a single training image is a difficult problem. In this section, here a method was proposed to deal with this illumination problem. The idea is to capture illumination conditions that might occur in the test sample in the training samples. The Lambertian reflectance model for the facial surface is assumed. The surface normals, albedo and the intensity image are related by an image formation model. For Lambertian objects, the diffused component of the surface reflection is modeled using the Lambert's Cosine Law given by

$$I = \rho max(n^T s, 0), \quad (5)$$

where I is the pixel intensity, s is the light source direction, ρ is the surface albedo and n is the surface normal of the corresponding surface point. Using this model, a nonstationary stochastic filter was recently proposed in [9] to estimate the albedo map from a single face image. We adapt this method to first estimate the albedo map from a given face image. Then, using the estimated albedo map, new images under any illumination condition using the image formation model is generated (5). This can be done by combining the estimated albedo map with the average facial information [10].It was shown in [11] that an image of an arbitrarily illuminated object can be approximated by a linear combination of the image of the same object in the same pose, illuminated by nine different light sources placed at preselected positions. Hence, the image formation equation can be rewritten as $I = \sum_{i=1}^{9} a_i I_i$ where $I_i = \rho max(n^T s_i, 0)$ and {s1, · · · , s9} are the pre-specified illumination directions. Since, the objective is to generate gallery images which will be sufficient to account for any illumination in the new image, images under the nine pre-specified illumination conditions were generated and used in the gallery. As a result, this algorithm has the ability to recognize human faces with good accuracy even when only a single or a very few images are provided for training.

Dictionary-based face recognition (DFR) algorithm is Summarized in Fig. 2.

2.2 *Pose-Robust Albedo Estimation*

The method presented previously can be generalized such that it can handle pose variations . Let $\bar{n}_{i,j}$, $\bar{s}$ and $\bar{\Theta}$ be some initial estimates of the surface normals, illumination direction and initial estimate of surface normals in pose Θ , respectively. Then, the initial albedo at pixel(i,j) can be obtained by

$$\bar{\rho}_{i,j} = \frac{x_{i,j}}{\bar{n}_{i,j}^{\bar{\Theta}} \cdot \bar{s}}$$

where $\bar{n}_{i,j}^{\bar{\Theta}}$ denotes the initial estimate of surface normals in pose $\bar{\Theta}$. Using this model, reformulate the problem of recovering albedo as a signal estimation problem. Formulation for the albedo estimation problem in the presence of pose is as follows:

$$\bar{\rho}_{i,j} = \rho_{i,j} h_{i,j} + \omega_{i,j}$$

Where

$$\omega_{i,j} = \frac{\bar{n}^{\Theta}_{i,j} \cdot s - \bar{n}^{\Theta}_{i,j} \cdot \bar{s}}{\bar{n}^{\Theta}_{i,j} \cdot \bar{s}} \rho_{i,j}$$

$$h_{i,j} = \frac{\bar{n}^{\Theta}_{i,j} \cdot \bar{s}}{\bar{n}^{\Theta}_{i,j} \cdot \bar{s}}$$

$\rho_{i,j}$ is the true albedo and $\bar{\rho}_{i,j}$ is the degraded albedo. In the case when the pose is known accurately, $\bar{\Theta} = \Theta$ and $h_{i,j}=1$. Hence, this can be viewed as a generalization in the case of unknown pose. Using this model, a stochastic filtering framework was recently presented to estimate the albedo from a single nonfrontal face image. Once pose and illumination have been normalized, one can use the relighting method described in the previous section to generate multiple frontal images with different lighting to achieve illumination and pose-robust recognition.

Note that a K-SVD based face recognition algorithm was recently proposed in [12]. Unlike [12], there is no discriminative approach to face recognition. The method is a reconstructive approach to discrimination and does not require multiple images to be available.

Given a test sample y and C training matrices $B_1, \cdots, B_C$ where each $B_i \in R^{N \times m_i}$ contains m_i training samples.
Procedure:
1. For each training image, use the relighting approach described in section 2.1 to generate multiple images with different illumination conditions and use them in the gallery.

2. Learn the best dictionaries Di, to represent the face images in Bi, using the K-SVD algorithm.

3. Compute the approximation vectors, $\hat{y}^i$ and the residual vectors, $r_i(y)$, using (2) and (3), respectively for $i = 1, \cdot \cdot \cdot, C$.

4. Identify y using (4).

Fig. 2. DFR algorithm.

3. RECOGNITION EXPERIMENTS

In this section, an experimental results on some of the publicly available databases for face recognition such as Extended Yale B dataset [13] . The comparison with other existing face recognition methods in [3] suggests that the SRC algorithm is among the best. Hence, it is used as a bench mark for comparisons in this paper. In all of our experiments, the K-SVD [8] algorithm is used to train the dictionaries with 15 atoms. The performance of our algorithm is compared with that of five different methods: SRC, nearest neighbor (NN), nearest subspace (NS), support vector machines (SVM) and class dependent principal component analysis (CDPCA) . This algorithm is also tested using several features, namely, Eigenfaces, Fisherfaces, Randomfaces, and downsampled images.

Results on Extended Yale B Database: There are a total of 2, 414 frontal face images of 38 individuals in the Extended Yale B database. These images were captured under various controlled indoor lighting conditions. They were manually cropped and normalized to the size of 192 × 168 . First set of experiments on the Extended Yale B data set consist of testing the performance of our algorithm with different features and dimensions. The objective is to verify the ability of our algorithm in recognizing faces with different illumination conditions. The experimental setup as considered in [3] is followed. The feature space dimensions of 30, 56, 120, and 504 corresponding to the downsampling ratios of, 1/32, 1/24, 1/16, and 1/8, respectively are computed. Randomly 32 images per subject (i.e. half of the images) for training and the other half for testing is selected. Then train dictionaries on the feature space. The best recognition rates of different methods with different dimensions and features are compared in Table1.

Table 1. Recognition Rates (RR) (in %) of different methods on the Extended Yale B database .

Method	DFR	SRC	NN	NS	SVM	CDPCA
RR	99.17	98.1	90.7	94.1	97.7	98.83

The maximum recognition rates achieved by DFR are 95.99%, 97.16%, 98.58% and 99.17% for all 30, 56, 120 and 504 dimensional feature spaces, respectively. The maximum recognition rate achieved by SRC is 98.1% with 504D randomfaces [3]. Also, NN, NS, SVM and CDPCA achieve the maximum recognition rates of 90.7%, 94.1%, 97.7%, and 98.83%, respectively . As can be seen from this experiment that DFR performs favorably over some of the competitive methods for face recognition on the Extended Yale B database.

Recognition with partial face features: In this section, the ability of our algorithm in recognizing faces from the partial face features is shown. Partial face features have been used in recovering the identity of human faces before [3]. The images in the Extended Yale B database for this experiment was used. The experimental setup of [3] was used. For each subject, 32 images are randomly selected for training, and the remaining images are used for testing. The region of eye, nose and mouth are selected as partial face features. For this experiment, the relighting step of our algorithm was omitted. Examples of these features are shown in Fig. 3. Table 2 compares the results obtained by using our method with other methods presented in [3]. As can be seen from the table, our method achieves recognition rates of 99. 3%, 98. 8% and 99. 8% on eye, nose and mouth region, respectively and it outperforms other methods such as SRC, NN, NS and SVM [3].

Fig. 3. Examples of partial facial features.

Table 3. Recognition results with partial facial features.

	Right Eye	Nose	Mouth
Dimension	5,040	4,270	12,936
DFR	99.30%	98.80%	99.80%
SRC	93.70%	87.30%	98.30%
NN	68.80%	49.20%	72.70%
NS	78.60%	83.70%	94.40%
SVM	85.80%	70.80%	95.30%

4. DISCUSSION AND CONCLUSION

A face recognition algorithm based on dictionary a learning method that is robust to changes in lighting. This entails using a relighting approach based on robust albedo estimation. Various experiments on popular face recognition data sets have shown that this method is efficient and can perform significantly better than many competitive face recognition algorithms. Even though, in this paper, a reconstructive approach to dictionary learning is taken, it is possible to learn discriminative dictionaries for the task of face recognition. It remains an interesting topic for future work to develop a discriminative dictionary learning algorithm that is robust to pose, expression and illumination variations.

5. REFERENCES

[1] W. Zhao, R. Chellappa, J. Phillips, and A. Rosenfeld, "Face recognition: A literature survey," *ACM Computing Surveys*, pp. 399–458, Dec. 2003.

[2] P. J. Phillips, “Matching pursuit filters applied to face identification,”*IEEE Trans. Image Process.*, vol. 7, no. 8, pp. 150–164, 1998.

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

[4] J. Huang, X. Huang, and D.Metaxas, “Simultaneous image transformation and sparse representation recovery,” *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, pp. 1–8, Ancorage, Alaska, June, 2008.

[5] A.Wagner, J. Wright, A. Ganesh, Z. Zhou, and Y.Ma, “Towards a practical face recognition system: Robust registration and illumination by sparse representation,” *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, pp. 597–604, Miami, FL, 2009.

[6] P. Nagesh and B. Li, “A compressive sensing approach for expressioninvariant face recognition,” *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, pp. 1518–1525, Miami, FL, June 2009.

[7] R. Rubinstein, A. M. Bruckstein, and M. Elad, “Dictionaries for sparse representation modeling,” *Proceedings of the IEEE*, submitted 2009.

[8] M. Aharon, M. Elad, and A. M. Bruckstein, “The k-svd: an algorithm for designing of overcomplete dictionaries for sparse representation,” *IEEE Trans. Signal Process.*, vol. 54, no. 11, pp. 4311–4322, 2006.

[9] S. Biswas, G. Aggarwal, and R. Chellappa, “Robust estimation of albedo for illumination-invariant matching and shape recovery,” *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 29, no. 2, pp. 884–899, Mar. 2009.

[10] V. Blanz and T. Vetter, “Face recognition based on fitting a 3d morphable model,” *IEEE Trans. Pattern Analysis andMachine Intelligence*, vol. 25, no. 9, pp. 1063–1074, Sept. 2003.

[11] K. Lee, J. Ho, and D. J. Kriegman, “Acquiring linear subspaces for face recognition under variable lighting,” *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 27, no. 5, pp. 684–698, May 2005.

[12] Q. Zhang and B. Li, “Discriminative k-svd for dictionary learning in face recognition,” *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, pp. 2691–2698, 2010.

[13] A. S. Georghiades, P. N. Belhumeur, and D. J. Kriegman, “From few to many: Ilumination cone models for face recognition under variable lighting and pose,” *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 23, no. 6, pp. 643–660, June 2001.

[14] T. Sim, S. Baker, and M. Bsat, “The cmu pose, illumination, and expression database,” *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 25, no. 12, pp. 1615–1618, Dec. 2003.

[15] V. M. Patel, T. Wu, S. Biswas, R. Chellappa, and P. J. Phillips, “Dictionary-based face recognition,” *UMIACS Tech. Report, UMIACSTR- 2010-07/CAR-TR-1030, University ofMaryland, College Park*, July 2010.

[16] M. Deriche, “A simple face recognition algorithm using eigeneyes and a class-dependent pca implementation,” *International Journal of Soft Computing*, vol. 3, pp. 438–442, 2008.