Sparse Representation for Face Recognition

Shahnazeer C K
Dept of IT, Anna University, Coimbatore-47, Tamilnadu, India.

Jayavel J
Dept of IT, Anna University, Coimbatore-47, Tamilnadu, India,

Abstract
This paper provides a problem of automatically recognizing human faces from frontal views with various facial expressions, occlusion, illumination and pose. There are two underlying motivations for us to write this paper: the first is to provide an occlusion and various expressions of the existing face recognition and the second is to offer some insights into the studies of pose and illumination of face recognition. We present a mathematical formulation and an algorithmic framework to achieve these goals. The existing framework offers a sparse representation of the test image with respect to the training image. The sparse representation can be accurately and efficiently computed by the l1 minimization. The proposed framework offers an improved sparse representation based classification algorithm. Firstly, for a discriminative representation, a non-negative constraint of sparse coefficient is added to sparse representation problem. Secondly, Mahalanobis distance is employed instead of Euclidean distance to measure the similarity between original data and reconstructed data. The proposed classification algorithm for face recognition has been evaluated under varying illumination and pose. Extensive experiments on publicly available databases verify the efficacy of the proposed method and support the above claims.

Keywords
Face Recognition, Occlusion, Illumination, Pose, Sparse representation, l1-minimization, Mahalanobis distance

I INTRODUCTION

Real-world automatic face recognition systems are confronted with a number of sources of within-class variation, including pose, expression, and illumination, as well as occlusion or disguise. Several decades of intense study within the pattern recognition community have produced numerous methods for handling each of these factors individually.

In this paper, we exploit the discriminative nature of sparse representation [2] to perform classification. We represent the test sample in an over-complete dictionary whose base elements are the training samples themselves. If sufficient training samples are available from each class, it will be possible to represent the test samples as a linear combination of just those training samples from the same class. This representation is naturally sparse, involving only a small fraction of the overall training database. We argue that in many problems of interest, it is actually the sparsest linear representation of the test sample in terms of this dictionary and can be recovered efficiently via l1-minimization [3]. Seeking the sparsest representation therefore automatically discriminates between the various classes present in the training set. Sparse representation also provides a simple and surprisingly effective means of rejecting invalid test samples not arising from any class in the training database: these samples sparsest representations tend to involve many dictionary elements, spanning multiple classes.

We investigate to what extent accurate recognition are possible using only 2D frontal images. More specifically, we address the following problem: Given only frontal images taken under several illuminations recognize faces despite large variation in both pose and illumination.

We propose a non-negative sparse representation based classification algorithm using Mahalanobis distance, and its applications on face recognition. First, we address the problem of sparse representation using the constraint of non-negative sparse coefficient to obtain a discriminative representation. Second, we replace Euclidean distance with Mahalanobis distance to measure similarity between original data and reconstructed data. Then, we reformulate the problem to be an equivalent l1 regularized least square problem for obtaining its solution.

We will motivate and study this new approach to classification within the context of automatic face recognition. Human faces are arguably the most extensively studied object in image-based recognition. This is partly due to the remarkable face recognition capability of the human visual system [4] and partly due to numerous important applications for face recognition technology [5]. In addition, technical issues associated with face recognition are representative of object recognition and even data classification in general.

II REVIEW OF EXISTING SYSTEM

A basic problem in object recognition is to use labeled training samples from k distinct object classes to correctly determine the class to which a new test sample belongs. We arrange the given ni training samples from the ith class as columns of a matrix Ai= [vi,1,vi,2,……,vi,n]εIRmxni . In the context of face recognition, we will identify a w x h grayscale image with the vector vεIRm (m=wh) given by stacking its columns;

Manuscript received July 5, 2014
Manuscript revised July 20, 2014

the columns of Ai are then the training face images of the ith subject.

A. SPARSE REPRESENTATION OF TRAINING SAMPLES

In this section, the training images were presented in matrices form. It performed a linear feature transform. Given sufficient training samples of the ith object class, Ai =[vi,1,vi,2,….,vi,ni]ε IRmxni, any new (test) sample yε IRm from the same class will approximately lie in the linear span of the training samples associated with object i: y= αi,1vi,1+αi,2vi,2+….+αi,nivi,ni for some scalars,αi,jεIR, j = 1, 2, . . . , ni.

Since the membership i of the test sample was initially unknown, and defined a new matrix A for the entire training set as the concatenation of the n training samples of all k object classes: A=[A1,A2,….,Ak]=[v1,1,v1,2,…..,vk,nk]

Then, the linear representation of y can be rewritten in terms of all training samples as y =Ax0 ε IRm, where x0=[0,….,0,αi,1,αi,n,0,….,0]T ε IRn is a coefficient vector whose entries were zero except those associated with the ith class.

B. RECOGNITION WITH FACIAL FEATURES

In this section, the role of feature extraction within the new sparse representation framework for face recognition was reexamined. One benefit of feature extraction, which carried over to the proposed sparse representation framework, was reduced data dimension and computational cost. Our SRC algorithm tested using several conventional holistic face features, namely, Eigenfaces [6], Laplacianfaces [7], and Fisher faces, and compares their performance with two unconventional features: random faces and down-sampled images. In this section, the stable version of SRC in various lower dimensional feature spaces were used for solving the reduced optimization problem with the error tolerance ε=0.05. The Mat lab implementation of the reduced (feature space) version of Algorithm 1 took only a few seconds per test image on a typical 3-GHz PC.

C. HANDLING CORRUPTION AND OCCLUSION

Occlusion poses a significant obstacle to robust real-world face recognition. This difficulty is mainly due to the unpredictable nature of the error incurred by occlusion: it may affect any part of the image and may be arbitrarily large in magnitude.

Now, to show how the proposed sparse representation classification framework can be extended to deal with occlusion. Assume that the corrupted pixels are a relatively small portion of the image. The error vector e0, like the vector x0, then has sparse nonzero entries. Since y0= Ax0 we can rewrite y = y0 +e0 =Ax0+e0 as

$$Y = [A, I]\begin{bmatrix} x_0 \\ e_0 \end{bmatrix} = Bw_0$$

Here, B= [A, I]ε IR mx(n+m), so the system y =Bw is always underdetermined and does not have a unique solution for w. However, from the above discussion about the sparsity of x0 and e0, the correct generating w0=[x0, e0] has at most ni +ρm nonzeros. We might therefore hope to recover w0 as the sparsest solution to the system y =Bw. In fact, if the matrix B is in general position, then as long as y =Bŵ for some ŵ with less than m/2 nonzeros, ŵ is the unique sparsest solution. Thus, if the occlusion e covers less than (m- ni)/2 pixels, ≈50 percent of the image, the sparsest solution ŵ to y =Bw is the true generator, w0=[x0, e0].

More generally, one can assume that the corrupting error e0 has a sparse representation with respect to some basis Ae ε IRmxne . That is, e0 =Aeu0 for some sparse vector u0 ε IRm. Here, choosing the special case Ae =I ε IRmxm as e0 is assumed to be sparse with respect to the natural pixel coordinates. If the error e0 is instead sparser with respect to another basis, the matrix B can simply redefine by appending Ae (instead of the identity I) to A and instead seek the sparsest solution w0 to the equation: y =Bw with B =[A, Ae] ε IR mx(n+ni).

In this way, the same formulation can handle more general classes of (sparse) corruption. As before, to recover the sparsest solution w0 from solving the following extended 11-minimization problem: (l1e) :ŵ1 =arg min||w||1 subject to Bw= y. hThat is, in Algorithm 1, now replace the image matrix A with the extended matrix B =[A, I] and x with w =[x, e].

Clearly, whether the sparse solution w0 can be recovered from the above 11-minimization depends on the neighborliness of the new polytope P =B (P1)=[A, I](P1). This polytope contains vertices from both the training images A and the identity matrix I. The bounds given in imply that if y is an image of subject i, the 11-minimization cannot guarantee to correctly recover w0 = [x0, e0] if ni +|support(e0)| > d/3. Generally, d » ni, so, c .m < t < [(m+1)/3] implies that the largest fraction of occlusion.

Algorithm 1 below summarizes the complete recognition procedure. Our implementation minimizes the l1-norm via a primal-dual algorithm for linear programming.

Algorithm1. Sparse Representation-based Classification (SRC)

---

1. Input: a matrix of training samples A = [A1,A2, . . .,Ak]εIR^{mxn} for k classes, a test sample $x\varepsilon IR^{m}$, (and an optional error tolerance ε> 0.)
2. Normalize the columns of A to have unit l^2-norm.

3. Solve the l^1-minimization problem: $\hat{y}_1$=arg $\min_y$ $\|y\|_1$ subject to Ay=x (Or alternatively, solve $\hat{y}_1$=arg $\min_y$ $\|y\|_1$ subject to $\|Ay-x\|_2 <= \varepsilon$.)
4. Compute the residuals $r_i(x)=\|x-A\delta_i(\hat{y}_1)\|_2$ for i = 1, . . . ,k.
5. Output: identity(x)= arg $\min_i r_i(x)$.

III PROPOSED SYSTEM

In this section, we discuss the difficulties associated with variations in pose and illumination, and why state-of-the-art methods that are quite effective at handling one of these modes of variability tend to fail when both are present simultaneously [8].

The figure 1 represents the system architecture of face recognition. It has two phases: (i) Training Phase (ii) Testing Phase. In training phase there can be four or more input face images. These inputs extract the features that are expression, occlusion, illumination and pose. These details were stored in a database. Put one image in the testing phase and it will compare features on the database. If the feature matches with the test image then it will display the identified face. Otherwise it will not display any image, because there is no corresponding image in the database.

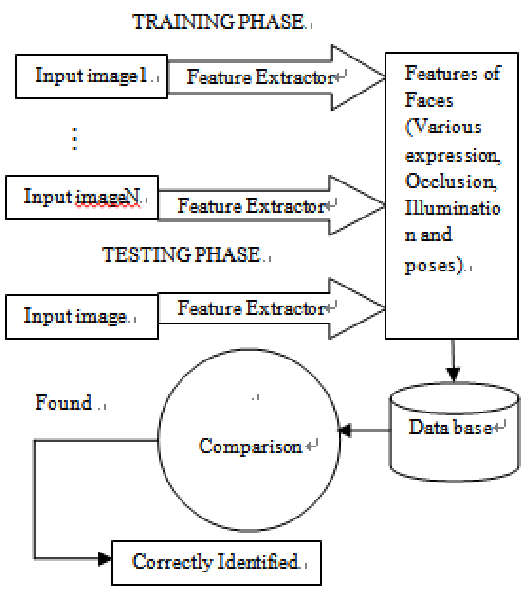

Fig 1: System Architecture

A) NON-NEGATIVE SPARSE REPRESENTATION

For a discriminative representation, we require the sparse coefficient to be non-negative. Therefore, for a given test sample, components of coefficient indicate the contributions of training samples. Furthermore, Mahalanobis distance is employed to measure the similarity between original data and reconstructed data, instead of Euclidean distance.

i) Sparse representation in subspace

Generally, given n high-dimensional data points A = {a1,…, an}, some research on manifold learning, for instance LLE, has proved that these data lie on a lower dimensional manifold. Any data point ai ε A can be approximately represented by the linear combination of its neighboring data points. This kind of linear representation can be generalized to labeled data. Given data points {a1,…, an} in one class, a new data point a^* in the same class can be represented as linear combination of {a1,…, an}, $a^* = \beta_1 a_1 + \ldots + \beta_n a_n$

ii) Nonnegative constraint for sparse coefficient

Sparse representation for classification is different from that for signal reconstruction. In signal processing, an original signal y should be reconstructed as accurately as possible. However, in classification, a discriminative representation is more important than reconstruction accuracy.

For a discriminative representation, we require that coefficient x should indicate contributions of all training samples to a given test sample. Therefore, we add constraint x≥0 to

$$\hat{x} = \arg\min_x \|x\|_1 + \gamma\|Ax - y\|_2^2$$ to get

$$\hat{x} = \arg\min_{x:x\geq 0} \|x\|_1 + \gamma\|Ax - y\|_2^2 \quad (1)$$

Where $\hat{x}$ is sparse, in which all elements are non-negative. The sparse representation from Eq. (1) avoids "negative" contribution of some training samples. In this way, for a given test sample, the similar training samples can be found from sparse representation.

iii) Similarity measure using Mahalanobis distance

For measure the similarity between original data and reconstructed data, we employ Mahalanobis distance instead of Euclidean distance. By introducing Mahalanobis distance, we obtain a generalized distance measure for face recognition, which can embody different weights on different components of feature vector. Mahalanobis Distance has been proved as a better similarity measure than Euclidean distance, when it comes to pattern recognition problems, for instance, face recognition [9].

Given two data points v1, v2 εRm, their Mahalanobis distance is given by:

$$d_M(v_1, v_2) = \sqrt{(v_1 - v_2)^T M (v_1 - v_2)},$$

where M ε Rmxm is a positive definite matrix.

Using the definition of Mahalanobis distance, the distance between original data y and reconstructed data Ax is

$$d_M(Ax, y) = \| Ax - y \| M = \sqrt{(Ax - y)^T M (Ax - y)}$$

The objective function with Mahalanobis distance can be formulated as follows:

$$\hat{x} = \arg\min_{x:x\geq 0} \| x \|_1 + \gamma \| Ax - y \|_M^2 \quad (2)$$

B) CLASSIFICATION ALGORITHM

Using the Cholesky factorization, the problem of Mahalanobis distance based non-negative sparse representation can be solved by a standard optimization algorithm. Then, the classification algorithm is designed based on the idea of finding the minimal reconstruction error [1].

i) Nonnegative l1 regularized least square

Since M is a positive definite matrix, the Cholesky factorization of M is M = LT L (3), where L is an upper triangular matrix with positive diagonal entries. From Eq. (3), the objective function in Eq. (2) can be formulated as:

$$\hat{x} = \arg\min_{x:x\geq 0} \| x \|_1 + \gamma \| LAx - Ly \|_2 \quad (4)$$

Set A'= LA and y'= Ly. Given parameter γ> 0, the problem is equal to the following problem:

$$\hat{x} = \arg\min_{x:x\geq 0} \lambda \| x \|_1 + \| A' x - y' \|_2^2 \quad (5)$$

Where λ= γ^{-1} . Eq. (5) is a non-negative l1-regularized least square problem, which can be solved by second-order cone programming [10, 11]

ii) Recognition algorithm

The recognition algorithm is inspired by the SRC algorithm proposed in [1]. Given a test sample y, we first compute its sparse coefficient $\hat{x}$. Then, we determine the class of this test sample from its reconstruction error between this test sample and the training samples of class k,

$$E_k(\hat{x}) = \|A\,\delta_k(\hat{x}) - y\|_M \quad (6)$$

where residual error is computed using Mahalanobis distance. For each class k, δ_k (x) : Rn → Rn is the characteristic function which selects the coefficients associated with the kth class. The class C(y) which test sample y belongs to is determined by

$$C(y) = \arg\min_k E_k(\hat{x}) \quad (7)$$

The whole algorithm of our method is summarized in algorithm 2.

Algorithm 2.Our proposed algorithm

Input: Test sample y, training matrix A, parameter γ

1. Normalize the columns of A using l_2 norm
2. Solve
   $$\hat{x} = \arg\min_{x:x\geq 0} \|x\|_1 + \gamma \|Ax - y\|_M^2$$
   Using an equivalent non-negative l_1-regularized least square problem
3. Compute reconstruction error E_k (k = 1,….,K):
   $$E_k(\hat{x}) = \|A\delta_k(\hat{x}) - y\|_M$$
4. Output: C(y), where C(y) = $\arg\min_k E_k(\hat{x})$

IV SIMULATION RESULTS

In this work, we have used the GTAV database. Recently, a face database has been created with the main purpose of testing the robustness of face recognition algorithms against strong pose and illumination variations. This database includes a total of 44 persons with 27 pictures per person which correspond to different pose views (0º, ±30º, ±45º, ±60º and v90º) under three different illuminations (environment or natural light, strong light source from an angle of 45º, and finally an almost frontal mid-strong light source. Furthermore, at least 10 more additional frontal view pictures are included with different occlusions and facial expresion variations. The resolution of the images are 240x320 and they are in BMP format.

1) OUTPUT SCREEN

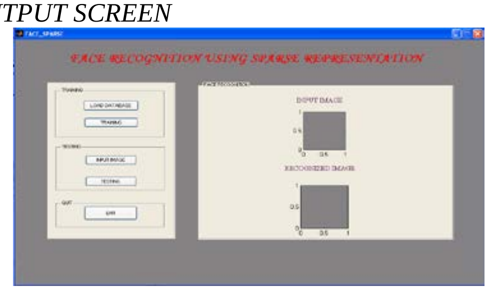

Fig 2: Form

The figure 2 represents the output screen of the recognition system. This form contains load database, training, input image, testing and exit buttons.

2) ILLUMINATION

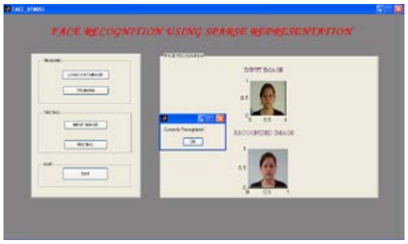

Fig 3: Illumination

The figure 3 represents the output of an illumination face image. Click testing, then recognized image will display in the screen. The recognition rate of the NSRC algorithm is 96.74%.

3) POSE DETECTION

Fig 4: Pose Detection

The figure 4 represents the output of a pose face image. Click testing, then recognized image will display in the screen. The recognition rate of the NSRC algorithm is 96.51%.

4) INCORRECTLY RECOGNIZED IMAGE

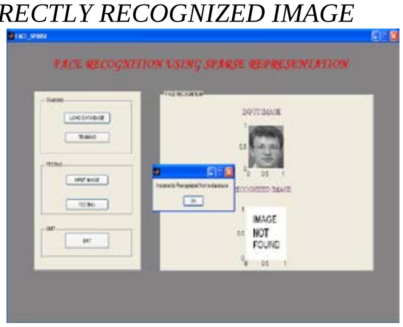

Fig 5: Incorrectly recognized image

The figure 5 represents the output of the incorrectly recognized image. Select an image which is not in the database. Then image not found will display in the screen. Images only recognized in the database.

V CONCLUSION AND FUTURE WORK

We have presented here a method of computing sparse representations of facial images that preserve the information required to estimate expression, occlusion, pose and illumination with SRC and our proposed algorithm. These both reduce the computation required to compute SRC, our proposed algorithm and improves the accuracy of the results.

An intriguing question for future work is whether this framework can be useful for object detection, in addition to recognition. The usefulness of sparsity in detection has been noticed in the work in [12]. We believe that the full potential of sparsity in robust object detection and recognition together is yet to be uncovered. From a practical standpoint, it would also be useful to extend the algorithm to less constrained conditions, especially variations in object pose. Robustness to occlusion allows the algorithm to tolerate small pose variation or misalignment.

REFERENCES

[1] John Wright, Allen Y Yang, Arvind Ganesh, and S Shankar Sastry,"Robust Face Recognition Via Sparse Representation ", IEEE Transaction on Pattern Analysis And Machine Intelligence, February 2009.

[2] K. Huang and S. Aviyente, "Sparse Representation for Signal Classification," Neural Information Processing Systems, 2006

[3] D.Donoho, "For Most Large Underdetermined Systems of Linear Equations the Minimal l1-Norm Solution Is Also the Sparsest Solution," Comm. Pure and Applied Math., vol. 59, no. 6, pp. 797 829, 2006

[4] P. Sinha, B. Balas, Y. Ostrovsky, and R. Russell, "Face Recognition by Humans: Nineteen Results All Computer Vision Researchers Should Know about," Proc. IEEE, 2006.

[5] W. Zhao, R. Chellappa, J. Phillips, and A. Rosenfeld, "Face Recognition: A Literature Survey," ACM Computing Surveys, pp. 399-458, 2003.

[6] M. Turk and A. Pentland, "Eigenfaces for Recognition," Proc. IEEE Int'l Conf. Computer Vision and Pattern Recognition, 1991.

[7] X. He, S. Yan, Y. Hu, P. Niyogi, and H. Zhang, "Face Recognition Using Laplacianfaces," IEEE Trans. Pattern Analysis and Machine Intelli, Mar.2005

[8] Yangfeng Ji, Tong Lin, Hongbin Zha," Mahalanobis Distance Based Non-negative Sparse Representation for Face Recognition", International Conference on Machine Learning and Applications,2009

[9] N. Ramanathan, R. Chellappa, and A. R. Chowdhury. Facial similarity across age, disguise, illumination and pose. In Proceedings of the International Conference on Image Processing(ICIP), 2004

[10] S. Boyd and L. Vandenberghe. Convex Optimization. Cambridge University Press, 2004.

[11] M. Lobo, L. Vandenberghe, S. Boyd, and H. Lebret. Applications of second-order cone programming. Linear Algebra and its Applications, 284, Nov. 1998.

[12] D. Geiger, T. Liu, and M. Donahue, "Sparse Representations for Image Decompositions," Int'l J. Computer Vision, vol. 33, no. 2, 1999.