Linear Regression for Face Recognition

Article *in* IEEE Transactions on Software Engineering · November 2010

DOI: 10.1109/TPAMI.2010.128 · Source: PubMed

3 authors:

Imran Naseem
University of Western Australia
33 PUBLICATIONS **678** CITATIONS

SEE PROFILE

Roberto Togneri
University of Western Australia
170 PUBLICATIONS **1,739** CITATIONS

SEE PROFILE

Mohammed Bennamoun
University of Western Australia
353 PUBLICATIONS **5,131** CITATIONS

SEE PROFILE

Linear Regression for Face Recognition

Imran Naseem,
Roberto Togneri, *Senior Member*, *IEEE*, and
Mohammed Bennamoun

Abstract—In this paper, we present a novel approach of face identification by formulating the pattern recognition problem in terms of linear regression. Using a fundamental concept that patterns from a single-object class lie on a linear subspace, we develop a linear model representing a probe image as a linear combination of class-specific galleries. The inverse problem is solved using the least-squares method and the decision is ruled in favor of the class with the minimum reconstruction error. The proposed Linear Regression Classification (LRC) algorithm falls in the category of *nearest subspace* classification. The algorithm is extensively evaluated on several standard databases under a number of exemplary evaluation protocols reported in the face recognition literature. A comparative study with state-of-the-art algorithms clearly reflects the efficacy of the proposed approach. For the problem of contiguous occlusion, we propose a *Modular LRC* approach, introducing a novel *Distance-based Evidence Fusion* (DEF) algorithm. The proposed methodology achieves the best results ever reported for the challenging problem of scarf occlusion.

Index Terms—Face recognition, linear regression, nearest subspace classification.

1 Introduction

FACE recognition systems are known to be critically dependent on manifold learning methods. A gray-scale face image of an order $a \times b$ can be represented as an ab-dimensional vector in the original *image space*. However, any attempt at recognition in such a high-dimensional space is vulnerable to a variety of issues often referred to as the *curse of dimensionality*. Therefore, at the feature extraction stage, images are transformed to low-dimensional vectors in the *face space*. The main objective is to find such a basis function for this transformation which could distinguishably represent faces in the face space. A number of approaches have been reported in the literature, such as Principle Component Analysis (PCA) [1], [2], Linear Discriminant Analysis (LDA) [3], and Independent Component Analysis (ICA) [4], [5]. Primarily, these approaches are classified in two categories, i.e., *reconstructive* and *discriminative* methods. Reconstructive approaches (such as PCA and ICA) are reported to be robust for the problem of contaminated pixels [6], whereas discriminative approaches (such as LDA) are known to yield better results in clean conditions [7]. Apart from these traditional approaches, it has been shown recently that unorthodox features, such as downsampled images and random projections, can serve equally well. In fact, the choice of the feature space may no longer be so critical [8]. What really matters is the dimensionality of feature space and the design of the classifier.

In this research, we propose a fairly simple but efficient linear regression-based classification (LRC) for the problem of face identification. Samples from a specific object class are known to lie on a linear subspace [3], [9]. We use this concept to develop class-specific models of the registered users simply using the downsampled gallery images, thereby defining the task of face recognition as a problem of linear regression. Least-squares estimation is used to estimate the vectors of parameters for a given probe against all class models. Finally, the decision rules in favor of the class with the most precise estimation. The proposed classifier can be categorized as a Nearest Subspace (NS) approach.

An important relevant work is presented in [8], where downsampled images from all classes are used to develop a *dictionary* matrix during the training session. Each probe image is represented as a linear combination of all gallery images, thereby resulting in an ill-conditioned inverse problem. With the latest research in compressive sensing and sparse representation, sparsity of the vector of coefficients is harnessed to solve the ill-conditioned problem using the l_1-norm minimization. In [10], the concept of *Locally Linear Regression* (LLR) is introduced specifically to tackle the problem of pose. The main thrust of the research is to indicate an approximate linear mapping between a nonfrontal face image and its frontal counterpart; the estimation of linear mapping is further formulated as a prediction problem with a regression-based solution. For the case of severe pose variations, the nonfrontal image is sampled to obtain many overlapped local segments. Linear regression is applied to each small patch to predict the corresponding virtual frontal patch; the LLR approach has shown some good results in the presence of coarse alignment. In [11], a two-step approach has been adopted, fusing the concept of wavelet decomposition and discriminant analysis to design a sophisticated feature extraction stage. These discriminant features are used to develop feature planes (for Nearest Feature Plane—NFP classifier) and feature spaces (for Nearest Feature Space—NFS classifier). The query image is projected onto the subspaces and the decision rules in favor of the subspace with the minimum distance. However, the proposed LRC approach, for the first time, simply uses the downsampled images in combination with the linear regression classification to achieve superior results compared to the benchmark techniques.

Further, for the problem of severe contiguous occlusion, a modular representation of images is expected to solve the problem [12]. Based on this concept, we propose an efficient *Modular LRC Approach*. The proposed approach segments a given occluded image and reaches individual decisions for each block. These intermediate decisions are combined using a novel *Distance-based Evidence Fusion* (DEF) algorithm to reach the final decision. The proposed DEF algorithm uses the distance metrics of the intermediate decisions to decide about the "goodness" of a partition. There are two major advantages of using the DEF approach. First, the nonface partitions are rejected dynamically; therefore, they do not take part in the final decision making. Second, the overall recognition performance is better than the best individual result of the combining partitions due to the efficient decision fusion of the face segments.

Algorithm: Linear Regression Classification (LRC)
Inputs: Class models $\mathbf{X}_i \in \mathbb{R}^{q \times p_i}, i = 1, 2, \ldots, N$ and a test image vector $\mathbf{y} \in \mathbb{R}^{q \times 1}$.
Output: Class of $\mathbf{y}$

1. $\hat{\beta}_i \in \mathbb{R}^{p_i \times 1}$ is evaluated against each class model, $\hat{\beta}_i = (\mathbf{X}_i^T \mathbf{X}_i)^{-1} \mathbf{X}_i^T \mathbf{y}, \;\; i = 1, 2, \ldots, N$
2. $\hat{\mathbf{y}}_i$ is computed for each $\hat{\beta}_i$, $\hat{\mathbf{y}}_i = \mathbf{X}_i \hat{\beta}_i, \;\; i = 1, 2, \ldots, N$
3. Distance calculation between original and predicted response variables $d_i(\mathbf{y}) = \|\mathbf{y} - \hat{\mathbf{y}}_i\|_2, \;\; i = 1, 2, \ldots, N$
4. Decision is made in favor of the class with the minimum distance $d_i(\mathbf{y})$

The rest of the paper is organized as follows: In Section 2, the proposed LRC and Modular LRC algorithms are described. This is

- *I. Naseem and R. Togneri are with the School of Electrical, Electronic and Computer Engineering, M018, The University of Western Australia, 35 Stirling Highway, Crawley, Western Australia 6009, Australia. E-mail: {imran.naseem, roberto}@ee.uwa.edu.au.*
- *M. Bennamoun is with the School of Computer Science and Software Engineering, The University of Western Australia, 35 Stirling Highway, Crawley, Western Australia 6009, Australia. E-mail: m.bennamoun@csse.uwa.edu.au.*

Manuscript received 10 Oct. 2008; revised 13 Jan. 2009; accepted 11 July 2009; published online 1 July 2010.
Recommended for acceptance by S. Li.
For information on obtaining reprints of this article, please send e-mail to: tpami@computer.org, and reference IEEECS Log Number TPAMI-2008-10-0684.
Digital Object Identifier no. 10.1109/TPAMI.2010.128.

followed by extensive experiments using standard databases under a variety of evaluation protocols in Section 3. The paper concludes in Section 4.

2 Linear Regression for Face Recogniton

2.1 Linear Regression Classification Algorithm

Let there be N number of distinguished classes with p_i number of training images from the ith class, $i = 1, 2, \ldots, N$. Each gray-scale training image is of an order $a \times b$ and is represented as $\overset{(m)}{u_i} \in \mathbb{R}^{a \times b}$, $i = 1, 2, \ldots, N$ and $m = 1, 2, \ldots, p_i$. Each gallery image is downsampled to an order $c \times d$ and transformed to vector through column concatenation such that $\overset{(m)}{u_i} \in \mathbb{R}^{a \times b} \rightarrow \overset{(m)}{\mathbf{w}_i} \in \mathbb{R}^{q \times 1}$, where $q = cd$, $cd < ab$. Each image vector is normalized so that maximum pixel value is 1. Using the concept that patterns from the same class lie on a linear subspace [9], we develop a class-specific model $\mathbf{X}_i$ by stacking the q-dimensional image vectors,

$$\mathbf{X}_i = \left[\overset{(1)}{\mathbf{w}_i} \; \overset{(2)}{\mathbf{w}_i} \; \ldots\ldots \; \overset{(p_i)}{\mathbf{w}_i} \right] \in \mathbb{R}^{q \times p_i}, \quad i = 1, 2, \ldots, N \tag{1}$$

Each vector $\overset{(m)}{\mathbf{w}_i}$, $m = 1, 2, \ldots, p_i$, spans a subspace of $\mathbb{R}^q$ also called the column space of $\mathbf{X}_i$. Therefore, at the training level, each class i is represented by a vector subspace, $\mathbf{X}_i$, which is also called the *regressor* or *predictor* for class i. Let z be an unlabeled test image and our problem is to classify z as one of the classes $i = 1, 2, \ldots, N$. We transform and normalize the gray-scale image z to an image vector $\mathbf{y} \in \mathbb{R}^{q \times 1}$ as discussed for the gallery. If $\mathbf{y}$ belongs to the ith class, it should be represented as a linear combination of the training images from the same class (lying in the same subspace), i.e.,

$$\mathbf{y} = \mathbf{X}_i \beta_i, \quad i = 1, 2, \ldots, N, \tag{2}$$

where $\beta_i \in \mathbb{R}^{p_i \times 1}$ is the vector of parameters. Given that $q \geq p_i$, the system of equations in (2) is well conditioned and β_i can be estimated using least-squares estimation [13], [14], [15]:

$$\hat{\beta}_i = \left(\mathbf{X}_i^T \mathbf{X}_i\right)^{-1} \mathbf{X}_i^T \mathbf{y}. \tag{3}$$

The estimated vector of parameters, $\hat{\beta}_i$, along with the predictors $\mathbf{X}_i$ are used to predict the response vector for each class i:

$$\begin{aligned} \hat{\mathbf{y}}_i &= \mathbf{X}_i \hat{\beta}_i, \quad i = 1, 2, \ldots, N \\ \hat{\mathbf{y}}_i &= \mathbf{X}_i \left(\mathbf{X}_i^T \mathbf{X}_i\right)^{-1} \mathbf{X}_i^T \mathbf{y} \\ \hat{\mathbf{y}}_i &= \mathbf{H}_i \mathbf{y}, \end{aligned} \tag{4}$$

where the predicted vector $\hat{\mathbf{y}}_i \in \mathbb{R}^{q \times 1}$ is the projection of $\mathbf{y}$ onto the ith subspace. In other words, $\hat{\mathbf{y}}_i$ is the closest vector, in the ith subspace, to the observation vector $\mathbf{y}$ in the euclidean sense [16]. $\mathbf{H}$ is called a *hat matrix* since it maps $\mathbf{y}$ into $\hat{\mathbf{y}}_i$. We now calculate the distance measure between the predicted response vector $\hat{\mathbf{y}}_i$, $i = 1, 2, \ldots, N$, and the original response vector $\mathbf{y}$,

$$d_i(\mathbf{y}) = \|\mathbf{y} - \hat{\mathbf{y}}_i\|_2, \quad i = 1, 2, \ldots, N \tag{5}$$

and rule in favor of the class with minimum distance, i.e.,

$$\underbrace{\min}_{i} d_i(\mathbf{y}), \quad i = 1, 2, \ldots, N. \tag{6}$$

2.2 Modular Approach for the LRC Algorithm

The problem of identifying partially occluded faces could be efficiently dealt with using the modular representation approach [12]. Contiguous occlusion can safely be assumed local in nature in a sense that it corrupts only a portion of conterminous pixels of the image, the amount of contamination being unknown. In the modular approach, we utilize the neighborhood property of the contaminated pixels by dividing the face image into a number of subimages. Each subimage is now processed individually and a final decision is made by fusing information from all of the subimages. A commonly reported technique for decision fusion is majority voting [12]. However, a major pitfall with majority voting is that it treats noisy and clean partitions equally. For instance, if three out of four partitions of an image are corrupted, majority voting is likely to be erroneous no matter how significant the clean partition may be in the context of facial features. The task becomes even more complicated by the fact that the distribution of occlusion over a face image is never known a priori and therefore, along with face and nonface subimages, we are likely to have face portions corrupted with occlusion. Some sophisticated approaches have been developed to filter out the potentially contaminated image pixels (for example, [17]). In this section, we make use of the specific nature of distance classification to develop a fairly simple but efficient fusion strategy which implicitly deemphasizes corrupted subimages, significantly improving the overall classification accuracy. We propose using the distance metric as evidence of our belief in the "goodness" of intermediate decisions taken on the subimages; the approach is called "Distance-based Evidence Fusion."

To formulate the concept, let us suppose that each training image is segmented in M partitions and each partitioned image is designated as v_n, $n = 1, 2, \ldots, M$. The nth partition of all p_i training images from the ith class is subsampled and transformed to vectors, as discussed in Section 2, to develop a class-specific and partition-specific subspace $\mathbf{U}_i^{(n)}$:

$$\mathbf{U}_i^{(n)} = \left[\overset{(1)^{(n)}}{\mathbf{w}_i} \; \overset{(2)^{(n)}}{\mathbf{w}_i} \; \ldots\ldots \; \overset{(pi)^{(n)}}{\mathbf{w}_i} \right], \quad i = 1, 2, \ldots, N. \tag{7}$$

Each class is now represented by M subspaces and altogether we have $M \times N$ subspace models. Now a given probe image is partitioned into M segments accordingly. Each partition is transformed to an image vector $\mathbf{y}^{(n)}$, $n = 1, 2, \ldots, M$. Given that i is the true class for the given probe image, $\mathbf{y}^{(n)}$ is expected to lie on the nth subspace of the ith class $\mathbf{U}_i^{(n)}$ and should satisfy

$$\mathbf{y}^{(n)} = \mathbf{U}_i^{(n)} \beta_i^{(n)}. \tag{8}$$

The vector of parameters and the response vectors are estimated as discussed in this section:

$$\hat{\beta}_i^{(n)} = \left[\left(\mathbf{U}_i^{(n)}\right)^T \mathbf{U}_i^{(n)}\right]^{-1} \left(\mathbf{U}_i^{(n)}\right)^T \mathbf{y}^{(n)}, \tag{9}$$

$$\hat{\mathbf{y}}_i^{(n)} = \mathbf{U}_i^{(n)} \hat{\beta}_i^{(n)}; \quad i = 1, 2, \ldots, N. \tag{10}$$

The distance measure between the estimated and the original response vector is computed

$$d_i\left(\mathbf{y}^{(n)}\right) = \left\|\mathbf{y}^{(n)} - \hat{\mathbf{y}}_i^{(n)}\right\|_2; \quad i = 1, 2, \ldots, N. \tag{11}$$

Now, for the nth partition, an intermediate decision called $j^{(n)}$ is reached with a corresponding minimum distance calculated as

$$d^{j^{(n)}} = \underbrace{\min}_{i} d_i(\mathbf{y}^{(n)}) \qquad i = 1, 2, \ldots, N. \tag{12}$$

Therefore, we now have M decisions $j^{(n)}$ with M corresponding distances $d^{j^{(n)}}$ and we decide in favor of the class with minimum distance:

$$Decision = \arg \underbrace{\min}_{j} d^{j^{(n)}} \quad n = 1, 2, \ldots, M. \tag{13}$$

3 Experimental Results

Extensive experiments were carried out to illustrate the efficacy of the proposed approach. Essentially, five standard databases, i.e., the AT&T [18], Georgia Tech [19], FERET [20], Extended Yale B

Fig. 1. A typical subject from the AT&T database.

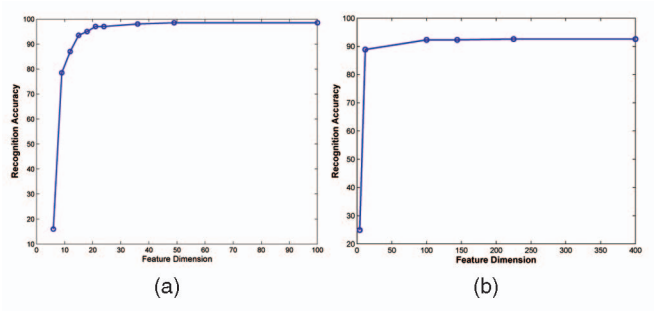

Fig. 2. (a) Recognition accuracy for the AT&T database with respect to feature dimension for a randomly selected leave-one-out experiment. (b) Feature dimensionality curve for the GT database.

[21], [22], and AR [23] have been addressed. These databases incorporate several deviations from the ideal conditions, including pose, illumination, occlusion, and gesture alterations. Several standard evaluation protocols reported in the face recognition literature have been adopted and a comprehensive comparison of the proposed approach with the state-of-the-art techniques has been presented. It is appropriate to indicate that the developed approach has been shown to perform well for the cases of severe gesture variations and contiguous occlusion with little change in pose, scale, illumination, and rotation. However, it is not meant to be robust to other deviations such as severe pose and illumination variations.

3.1 AT&T Database

The AT&T database is maintained at the AT&T Laboratories, Cambridge University; it consists of 40 subjects with 10 images per subject. The database incorporates facial gestures, such as smiling or nonsmiling, open or closed eyes, and alterations like glasses or without glasses. It also characterizes a maximum of 20 degree rotation of the face with some scale variations of about 10 percent (see Fig. 1).

We follow two evaluation protocols as proposed in the literature quite often [24], [25], [26], [27]. Evaluation Protocol 1 (EP1) takes the first five images of each individual as a training set, while the last five are designated as probes. For Evaluation Protocol 2 (EP2), the "leave-one-out" strategy is adopted. All experiments are conducted by downsampling 112×92 images to an order of 10×5. A detailed comparison of the results for the two evaluation protocols is summarized in Table 1. For EP1, the LRC algorithm achieves a comparable recognition accuracy of 93.5 percent in a 50D feature space; the best results are reported for the latest Eigenfeature Regularization and Extraction (ERE) approach, which are 3.5 percent better than the LRC method. Note that recognition error rate is converted to recognition success rate for [27]. Also for EP2, the LRC approach attains a high recognition success of 98.75 percent in a 50D feature space, it outperforms the ICA approach by 5 percent (approximately), and it is fairly comparable to the Fisherfaces, Eigenfaces, Kernel Eigenfaces, 2DPCA, and ERE approaches.

The choice of dimensionality for the AT&T database is dilated in Fig. 2a, which reflects that the recognition rate becomes fairly constant above a 40-dimensional feature space.

3.2 Georgia Tech (GT) Database

The Georgia Tech database consists of 50 subjects with 15 images per subject [19]. It characterizes several variations such as pose, expression, cluttered background, and illumination (see Fig. 3).

Images were downsampled to an order of 15×15 to constitute a 225D feature space; choice of dimensionality is depicted in Fig. 2b, which reflects a consistent performance above a 100D feature space. The first eight images of each subject were used for training, while the remaining seven served as probes [27]; all experiments were conducted on the original database without any cropping/normalization. Table 2 shows a detailed comparison of the LRC with a variety of approaches; all results are as reported in [27] with recognition error rates converted to recognition success rates. Also, results in [27] are shown for a large range of feature dimensions; for the sake of fair comparison, we have picked the best reported results. The proposed LRC algorithm outperforms the traditional PCAM and PCAE approaches by a margin of 12 and 18.57 percent, respectively, achieving a high recognition accuracy of 92.57 percent. It is also shown to be fairly comparable to all other methods, including the latest ERE approaches.

TABLE 1
Results for EP1 and EP2 Using the AT&T Database

Evaluation Protocol	Approach	Recognition Rate
EP1	Fisherfaces [24]	94.50%
	ICA [24]	85.00%
	Kernel Eigenfaces [24]	94.00%
	2DPCA [24]	96.00%
	ERE [27]	97.00%
	LRC	**93.50%**
EP2	Fisherfaces [24]	98.50%
	ICA [24]	93.80%
	Eigenfaces [24]	97.50%
	Kernel Eigenfaces [24]	98.00%
	2DPCA [24]	98.30%
	ERE_$\tilde{\mathbf{S}}^b$ [27]	99.25%
	ERE_$\tilde{\mathbf{S}}^t$ [27]	99.00%
	LRC	**98.75%**

Fig. 3. Samples of a typical subject from the GT database.

TABLE 2
Results for the Georgia Tech Database

Method	PCAM	PCAE	BML	DSL	NLDA
Recognition Rate	80.57%	74.00%	87.43%	90.57%	88.86%
Method	FLDA	UFS	ERE_$\tilde{\mathbf{S}}^b$	ERE_$\tilde{\mathbf{S}}^t$	**LRC**
Recognition Rate	90.71%	90.86	92.86%	93.14%	**92.57%**

TABLE 3
Results for the FERET Database

Experiment	Method	fa	fb	ql	qr	Overall
EP1	PCA	74.22%	73.44%	65.63%	72.66%	71.48%
	ICA I	73.44%	71.09%	65.63%	68.15%	69.57%
	LRC	**91.41%**	**94.53%**	**78.13%**	**84.38%**	**87.11%**
EP2	PCA	80.00%	78.75%	67.50%	71.75%	74.50%
	ICA I	77.50%	77.25%	68.50%	70.25%	73.37%
	LRC	**93.25%**	**93.50%**	**75.25%**	**76.00%**	**84.50%**

3.3 FERET Database

3.3.1 Evaluation Protocol 1

The FERET database is arguably one of the largest publicly available databases [20]. Following [27], [28], we construct a subset of the database consisting of 128 subjects, with at least four images per subject. We, however, used four images per subject [27]. Fig. 4 shows images of a typical subject from the FERET database. It has to be noted that, in [27], the database consists of 256 subjects; 128 subjects (i.e., 512 images) are used to develop the face space, while the remaining 128 subjects are used for the face recognition trials. The proposed LRC approach uses the gallery images of each person to form a linear subspace; therefore, it does not require any additional development of the face space. However, it requires multiple gallery images for a reliable construction of linear subspaces. Using a single gallery image for each person is not substantial in the context of linear regression, as this corresponds to only a single regressor (or predictor) observation, leading to erroneous least-squares calculations.

Cross-validation experiments for LRC were conducted in a 42D feature space; for each recognition trial, three images per person were used for training, while the system was tested for the fourth one. The results are shown in Table 3. The frontal images fa and fb incorporate gesture variations with small pose, scale, and rotation changes, whereas ql and qr correspond to major pose variations (see [20], for details). The proposed LRC approach copes well with the problem of facial expressions in the presence of small pose variations, achieving high recognition rates of 91.41 and 94.53 percent for fa and fb, respectively. It outperforms the benchmark PCA and ICA I algorithms by margins of 17.19 and 17.97 percent for fa and 21.09 and 23.44 percent for fb, respectively. The LRC approach, however, shows degraded recognition rates of 78.13 and 84.38 percent for the severe pose variations of ql and qr, respectively; however, even with such major posture changes, it is substantially superior to the PCA and ICA I approaches. In an overall sense, we achieve a recognition accuracy of 87.11 percent, which is favorably comparable to 83.00 percent recognition achieved by ERE [27] using single gallery images.

3.3.2 Evaluation Protocol 2

In this experimental setup, we validated the consistency of the proposed approach with a large number of subjects. We now have a subset of FERET database consisting of 400 randomly selected persons. Cross-validation experiments were conducted as discussed above; results are reported in Table 3. The proposed LRC approach showed quite agreeable results with the large database as well. It persistently achieved high recognition rates of 93.25 and 93.50 percent, for fa and fb, respectively. For the case of severe pose variations of ql and qr, we note a slight degradation in the performance, as expected. The overall performance is, however, pretty much comparable with an average recognition success of 84.50 percent. For all case studies, the proposed LRC approach is found to be superior to the benchmark PCA and ICA I approaches.

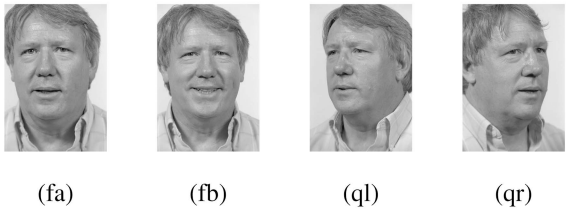

Fig. 4. A typical subject from the FERET database; fa and fb represent frontal shots with gesture variations, while ql and qr correspond to pose variations.

3.4 Extended Yale B Database

Extensive experiments were carried out using the Extended Yale B database [21], [22]. The database consists of 2,414 frontal face images of 38 subjects under various lighting conditions. The database was divided in five subsets; subset 1 consisting of 266 images (seven images per subject) under nominal lighting conditions was used as the gallery, while all others were used for validation (see Fig. 5). Subsets 2 and 3, each consisting of 12 images per subject, characterize slight-to-moderate luminance variations, while subset 4 (14 images per person) and subset 5 (19 images per person) depict severe light variations.

All experiments for the LRC approach were conducted with images downsampled to an order 20×20; results are shown in Table 4. The proposed LRC approach showed excellent performance for moderate light variations, yielding 100 percent recognition accuracy for subsets 2 and 3. The recognition success, however, falls to 83.27 and 33.61 percent, for subsets 4 and 5, respectively. The proposed LRC approach has shown better tolerance for considerable illumination variations compared to benchmark reconstructive approaches comprehensively outperforming PCA and ICA I for all case studies. The proposed algorithm, however, could not withstand severe luminance alterations.

3.5 AR Database

The AR database consists of more than 4,000 color images of 126 subjects (70 men and 56 women) [23]. The database

Fig. 5. Starting from the top, each row illustrates samples from subsets 1, 2, 3, 4, and 5, respectively.

TABLE 4
Results for the Extended Yale B Database

Approach	Subset 2	Subset 3	Subset 4	Subset 5
PCA	98.46%	80.04%	15.79%	24.38%
ICA I	98.03%	80.70%	15.98%	22.02%
LRC	**100%**	**100%**	**83.27%**	**33.61%**

characterizes divergence from ideal conditions by incorporating various facial expressions (neutral, smile, anger, and scream), luminance alterations (left light on, right light on, and all side lights on), and occlusion modes (sunglass and scarf). It has been used by researchers as a testbed to evaluate and benchmark face recognition algorithms. In this research, we address two fundamental challenges of face recognition, i.e., facial expression variations and contiguous occlusion.

3.5.1 Gesture Variations

Facial expressions are defined as the variations in appearance of the face induced by internal emotions or social communications [29]. In the context of face identification, the problem of varying facial expressions refers to the development of the face recognition systems which are robust to these changes. The task becomes more challenging due to the natural variations in the head orientation with the changes in facial expressions, as depicted in Fig. 6. Most of the face detection and orientation normalization algorithms make use of the facial features, such as eyes, nose, and mouth. It has to be noted that for the case of adverse gesture variations such as "scream," the eyes of the subject naturally are closed (see Figs. 6d and 6h). Consequently, under such severe conditions, the eyes cannot be automatically detected and therefore face normalization is likely to be erroneous. Hence, there are two possible configurations for a realistic evaluation of robustness for a given face recognition algorithm: 1) by implementing an automatic face localization and normalization module before the actual face recognition module or 2) by evaluating the algorithm using the original frame of the face image rather than a manually localized and aligned face. With this understanding, we validate the proposed LRC algorithm for the problem of gesture variations on the original, uncropped, and unnormalized AR database.

Out of 125 subjects of the AR database, a subset is generated by randomly selecting 100 individuals (50 males and 50 females). All 576×768 images are downsampled to an order of 10×10.The database characterizes four facial expressions: neutral, smile, anger, and scream.

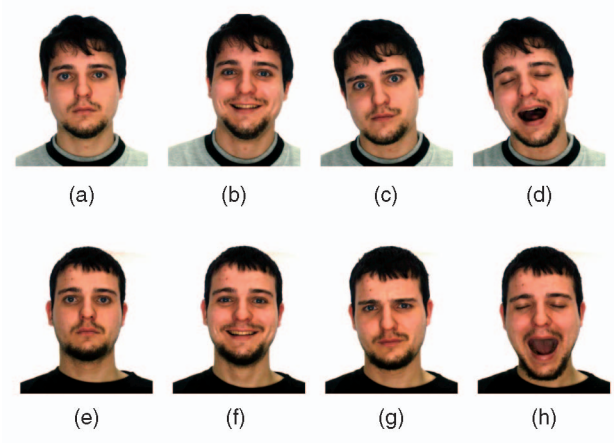

Fig. 6. Gesture variations in the AR database. Note the changing position of the head with different poses. (a)-(d) and (e)-(h) correspond to two different sessions incorporating neutral, happy, angry, and screaming expressions, respectively.

TABLE 5
Recognition Results for Gesture Variations Using the LRC Approach

Gestures	Recognition Accuracy
Neutral	99.00%
Smile	98.50%
Anger	98.50%
Scream	99.50%
Overall	98.88%

Experiments are based on cross-validation strategy, i.e., each time the system is trained using images of three different expressions (600 images in all), while the testing session is conducted using the left-out expression (200 images) [17]. The LRC algorithm achieves a high recognition accuracy for all facial expressions; the results for a 100D feature space are reported in Table 5 with an overall average recognition of 98.88 percent. For the case of screaming, the proposed approach achieves 99.50 percent recognition accuracy, which outperforms the results in [17] by 12.50 percent; noteworthy is the fact that results in [17] are shown on a subset consisting of only 50 individuals.

3.5.2 Contiguous Occlusion

The problem of face identification in the presence of contiguous occlusion is arguably one of the most challenging paradigms in the context of robust face recognition. Commonly used objects, such as caps, sunglasses, and scarves, tend to obstruct facial features, causing recognition errors. Moreover, in the presence of occlusion, the problems of automatic face localization and normalization as discussed in the previous section are even more magnified. Therefore, experiments on manually cropped and aligned databases make an implicit assumption of an evenly cropped and nicely aligned face, which is not available in practice.

The AR database consists of two modes of contiguous occlusion, i.e., images with a pair of sunglasses and a scarf. Fig. 7 reflects these two scenarios for two different sessions. A subset of the AR database consisting of 100 randomly selected individuals (50 men and 50 women) is used for empirical evaluation. The system is trained using Figs. 6a, 6b, 6c, 6d, 6e, 6f, 6g, and 6h for each subject, thereby generating a gallery of 800 images. Probes consist of Figs. 7a and 7b for sunglass occlusion and Figs. 7c and 7d for scarf occlusion. The proposed approach is evaluated on the original database *without any manual cropping and/or normalization.*

For the case of sunglass occlusion, the proposed LRC approach achieves a high recognition accuracy of 96 percent in a 100D feature space. Table 6 depicts a detailed comparison of the LRC approach with a variety of approaches reported in [8], consisting of Principal Component Analysis, Independent Component Analysis-architecture I (ICA I), Local Nonnegative Matrix Factorization (LNMF), least-squares projection onto the subspace spanned by all face images, and Sparse-Representation-based Classification (see [8], for details). NN and NS correspond to Nearest Neighbors and Nearest Subspace-based classification, respectively. The LRC algorithm comprehensively outperforms the best competitor (SRC) by a margin of 9 percent. To the best of our knowledge, the LRC approach achieves the best results for the case of sunglass

Fig. 7. Examples of contiguous occlusion in the AR database.

TABLE 6
Recognition Results for Occlusion

Approach	Recognition Accuracy	
	Sunglass	Scarf
PCA+NN	70.00%	12.00%
ICA I+NN	53.50%	15.00%
LNMF+NN	33.50%	24.00%
l^2+NS	64.50%	12.50%
SRC	87.00%	59.50%
LRC	**96.00%**	**26%**

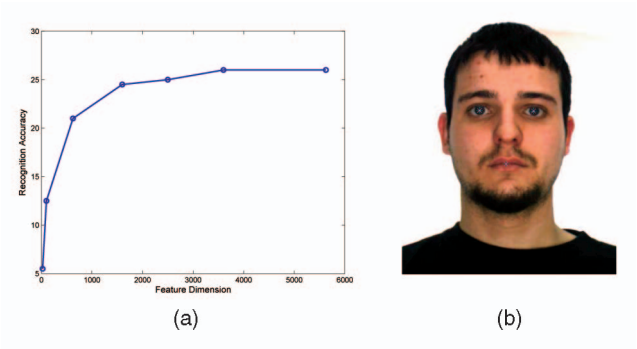

Fig. 8. (a) The recognition accuracy versus feature dimension for scarf occlusion using the LRC approach. (b) A sample image indicating eye and mouth locations for the purpose of manual alignment.

occlusion. Note that in [8], a comparable recognition rate of 97.5 percent has been achieved by a subsequent image partitioning approach.

For the case of severe scarf occlusion, the proposed approach gives a recognition accuracy of 26 percent in a 3,600D feature space. Fig. 8a shows the performance of the system with respect to an increasing dimensionality of the feature space. Although the LRC algorithm outperforms the classical PCA and ICA I approaches by a margin of 14 and 11 percent, respectively, it lags the SRC approach by a margin of 33.5 percent.

We now demonstrate the efficacy of the proposed Modular LRC approach under severe occlusion conditions. As a preprocessing step, the AR database is normalized, both in scale and orientation, generating a cropped and aligned subset of images consisting of 100 subjects. Images are manually aligned using eye and mouth locations, as shown in Fig. 8b [30]; each image is cropped to an order 292×240. Some images from the normalized database are shown in Fig. 9.

All images are partitioned into four blocks, as shown in Fig. 10a. The blocks are numbered in ascending order from left to right, starting from the top; the LRC algorithm for each subimage uses a 100D feature space as discussed in the previous section. Fig. 11a elaborates on the efficacy of the proposed approach for a random probe image. In our proposed approach, we have used the distance measures $d^{j(n)}$ as evidence of our belief in a subimage. The key factor to note in Fig. 11a is that corrupted subimages (i.e., blocks 3 and 4 in Fig. 10a) reach a decision with a low belief, i.e., high distance measures $d^{j(n)}$. Therefore, in the final decision making, these corrupted blocks are rejected, thereby giving a high recognition accuracy of 95 percent. The superiority of the proposed approach is more pronounced by considering the individual recognition rates of the subimages in Fig. 11b. Blocks 1 and 2 yield a high classification accuracy of 94 and 90 percent, respectively, whereas blocks 3 and 4 give 1 percent output each. Note that the effect of the proposed approach is twofold: First it automatically deemphasizes the nonface partitions. Second, the efficient and dynamic fusion harnesses the complementary information of the face subimages to yield an overall recognition accuracy of 95 percent, which is better than the best of the participating face partition.

Fig. 9. Samples of cropped and aligned faces from the AR database.

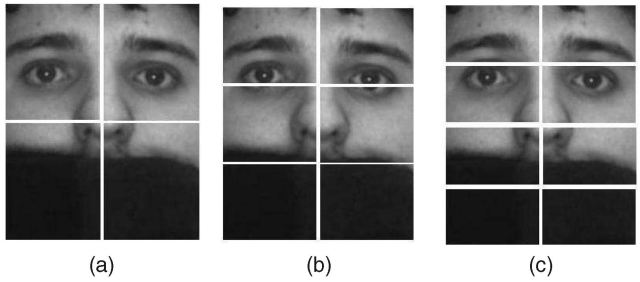

Fig. 10. Case studies for the Modular LRC approach for the problem of scarf occlusion.

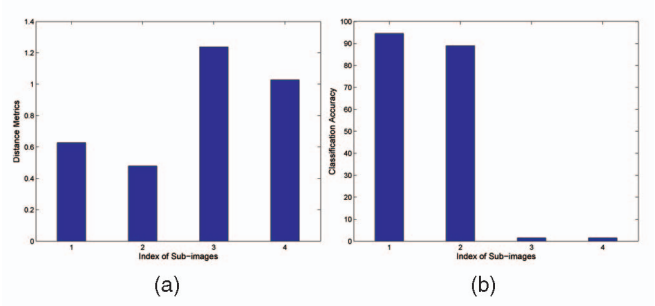

Fig. 11. (a) Distance measures $d^{j(n)}$ for the four partitions; note that nonface components make decisions with low evidences. (b) Recognition accuracies for all blocks.

Note that in Fig. 10a, the partitioning is such that the uncorrupted subimages (blocks 1 and 2) correspond to undistorted and complete eyes, which are arguably one of the most discriminant facial features. Therefore, one can argue that this high classification accuracy is due to this coincidence and the approach might not work well otherwise. To remove this ambiguity, we partitioned the images into six and eight blocks, as shown in Figs. 10b and 10c, respectively. The blocks are numbered left to right, starting from the top. For Fig. 10b, partitions 1 and 2 give high recognition accuracies of 92.5 and 90 percent, respectively, while the remaining blocks, i.e., 3, 4, 5, and 6 yield 8, 7, 0, and 1 percent recognition, respectively. Interestingly, although the best block gives 92.5 percent ,which is 1.5 percent less than the best block of Fig. 10a, the overall classification accuracy comes out to be 95.5 percent.

Similarly, in Fig. 10c, blocks 1, 2, 3, and 4 give classification accuracies of 88.5, 84.5, 80, and 77.5 percent, respectively, while the corrupted blocks 5, 6, 7, and 8 produce 3, 0.5, 1, and 1 percent of classification. The proposed evidence-based algorithm yields a high classification accuracy of 95 percent. A key factor to note is that the best individual result is 88.5 percent, which lags the best individual result of Fig. 10a by 5.5 percent. However, the proposed integration of combining blocks yields a comparable overall recognition. Interestingly, the eyebrow regions (blocks 1 and 2) in Fig. 10c have been found most useful.

To the best of our knowledge, the recognition accuracy of 95.5 percent achieved by the presented approach is the best result ever reported for the case of scarf occlusion, the previous best being 93.5 percent achieved by partitioned SRC approach in [8]; also

TABLE 7
Comparison of the DEF with the Sum Rule for Three Case Studies

Case Studies	Sum Rule	DEF
4-partitions	56.50%	95.00%
6-partitions	62.00%	95.50%
8-partitions	75.00%	95.00%

93 percent classification accuracy is reported for 50 subjects in [17]. Finally, we compare the proposed DEF approach with the weighted sum rule, which is perhaps the major workhorse in the field of combining classifiers [31]. The comparison for three case studies in Fig. 10 is presented in Table 7. Note that, without any prior knowledge of the goodness of a specific partition, we used equal weights for all subimages of a given partitioned image. The DEF approach comprehensively outperforms the sum rule for the three case studies, showing an improvement of 38.5, 33.5, and 20 percent, respectively. The performance of the sum rule improves with the increase in the pure face regions, thereby demonstrating strong dependency on the way partitioning is performed. On the other hand, the proposed DEF approach shows a quite consistent performance for even the worst-case partitioning, i.e., Fig. 10a consisting of an equal number of face and nonface partitions.

4 Conclusion

In this research, a novel classification algorithm is proposed which formulates the face identification task as a problem of linear regression. The proposed LRC algorithm is extensively evaluated using the most standard databases with a variety of evaluation methods reported in the face recognition literature. Specifically, the challenges of varying facial expressions and contiguous occlusion are addressed. Considerable comparative analysis with the state-of-the-art algorithms clearly reflects the potency of the proposed approach. The proposed LRC approach reveals a number of interesting outcomes. Apart from the Modular LRC approach for face identification in the presence of disguise, the LRC approach yields high recognition accuracies without requiring any preprocessing steps of face localization and/or normalization. We argue that in the presence of nonideal conditions such as occlusion, illumination, and severe gestures, a cropped and aligned face is generally not available. Therefore, a consistent reliable performance with unprocessed standard databases makes the LRC algorithm appropriate for real scenarios. For the case of varying gestures, the LRC approach has been shown to cope well with the most severe screaming expression where the state-of-the-art techniques lag behind, indicating consistency for mild and severe changes. For the problem of face recognition in the presence of disguise, the Modular LRC algorithm using an efficient evidential fusion strategy yields the best reported results in the literature.

In the paradigm of view-based face recognition, the choice of features for a given case study has been a debatable topic. Recent research has, however, shown the competency of unorthodox features such as downsampled images and random projections, indicating a divergence from the conventional ideology [8]. The proposed LRC approach in fact conforms to this emerging belief. It has been shown that with an appropriate choice of classifier, the downsampled images can produce good results compared to the traditional approaches. The simple architecture of the proposed approach makes it computationally efficient, therefore suggesting a strong candidacy for realistic video-based face recognition applications. Other future directions include the robustness issues related to illumination, random pixel corruption, and pose variations.

Acknowledgments

This research is partly funded by the Australian Research Council (ARC) grant No. DP0664228.

References

[1] I.T. Jolliffe, *Principal Component Analysis.* Springer, 1986.
[2] M. Turk and A. Pentland, "Eigenfaces for Recognition," *J. Cognitive Neuroscience,* vol. 3, no. 1, pp. 71-86, 1991.
[3] P.N. Belhumeur, J.P. Hespanha, and D.J. Kriegman, "Eigenfaces vs. Fisherfaces: Recognition Using Class Specific Linear Projection," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 19, no. 7, pp. 711-720, July 1997.
[4] P. Comon, "Independent Component Analysis—A New Concept?" *Signal Processing,* vol. 36, pp. 287-314, 1994.
[5] M. Bartlett, H. Lades, and T. Sejnowski, "Independent Component Representations for Face Recognition," *Proc. SPIE Conf. Human Vision and Electronic Imaging III,* pp. 528-539, 1998.
[6] A. Leonardis and H. Bischof, "Robust Recognition Using Eigenimages," *Computer Vision and Image Understanding,* vol. 78, no. 1, pp. 99-118, 2000.
[7] R.O. Duda, P.E. Hart, and D.G. Stork, *Pattern Classification.* John Wiley & Sons, 2000.
[8] 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.
[9] R. Barsi and D. Jacobs, "Lambertian Reflection and Linear Subspaces," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 25, no. 2, pp. 218-233, Feb. 2003.
[10] X. Chai, S. Shan, X. Chen, and W. Gao, "Locally Linear Regression for Pose-Invariant Face Recognition," *IEEE Trans. Image Processing,* vol. 16, no. 7, pp. 1716-1725, July 2007.
[11] J. Chien and C. Wu, "Discriminant Waveletfaces and Nearest Feature Classifiers for Face Recognition," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 24, no. 12, pp. 1644-1649, Dec. 2002.
[12] A. Pentland, B. Moghaddam, and T. Starner, "View-Based and Modular Eigenspaces for Face Recognition," *Proc. IEEE Conf. Computer Vision and Pattern Recognition,* 1994.
[13] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning; Data Mining, Inference and Prediction.* Springer, 2001.
[14] G.A.F. Seber, *Linear Regression Analysis.* Wiley-Interscience, 2003.
[15] T.P. Ryan, *Modern Regression Methods.* Wiley-Interscience, 1997.
[16] R.G. Staudte and S.J. Sheather, *Robust Estimation and Testing.* Wiley-Interscience, 1990.
[17] S. Fidler, D. Skocaj, and A. Leonardis, "Combining Reconstructive and Discriminative Subspace Methods for Robust Classification and Regression by Subsampling," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 28, no. 3, pp. 337-350, Mar. 2006.
[18] F. Samaria and A. Harter, "Parameterisation of a Stochastic Model for Human Face Identification," *Proc. Second IEEE Workshop Applications of Computer Vision,* Dec. 1994.
[19] "Georgia Tech Face Database," http://www.anefian.com/face_reco.htm, 2007.
[20] P.J. Phillips, H. Wechsler, J.S. Huang, and P.J. Rauss, "The FERET Database and Evaluation Procedure for Face-Recognition Algorithms," *Image and Vision Computing,* vol. 16, no. 5, pp. 295-306, 1998.
[21] A. Georghiades, P. Belhumeur, and D. Kriegman, "From Few to Many: Illumination 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.
[22] K.C. Lee, J. Ho, and D. 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.
[23] A. Martinez and R. Benavente, "The AR Face Database," CVC Technical Report 24, June 1998.
[24] J. Yang, D. Zhang, A.F. Frangi, and J. Yang, "Two-Dimensional PCA: A New Approach to Appearance-Based Face Representation and Recognition," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 26, no. 1, pp. 131-137, Jan. 2004.
[25] M.H. Yang, "Kernel Eignefaces vs. Kernel Fisherfaces: Face Recognition Using Kernel Methods," *Proc. Fifth IEEE Int'l Conf. Automatic Face and Gesture Recognition,* pp. 215-220, May 2002.
[26] P.C. Yuen and J.H. Lai, "Face Representation Using Independent Component Analysis," *Pattern Recognition,* vol. 35, no. 6, pp. 1247-1257, 2002.
[27] X. Jiang, B. Mandal, and A. Kot, "Eigenfeature Regularization and Extraction in Face Recognition," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 30, no. 3, pp. 383-394, Mar. 2008.
[28] J. Lu, K.N. Plataniotis, A.N. Venetsanopoulos, and S.Z. Li, "Ensemble-Based Discriminant Learning with Boosting for Face Recognition," *IEEE Trans. Neural Networks,* vol. 17, no. 1, pp. 166-178, Jan. 2006.
[29] *Handbook of Face Recognition,* S.Z. Li, and A.K. Jain, eds. Springer, 2005.
[30] L. Zhang and G.W. Cottrell, "When Holistic Processing Is Not Enough: Local Features Save the Day," *Proc. 26th Ann. Cognitive Science Soc. Conf.,* 2004.
[31] J. Kittler, M. Hatef, R.P.W. Duin, and J. Matas, "On Combining Classifiers," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 20, no. 3, pp. 226-239, Mar. 1998.