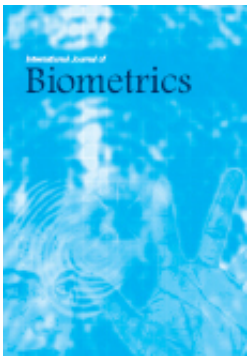


International Journal of Biometrics

ISSN online: 1755-831X - ISSN print: 1755-8301
https://www.inderscience.com/ijbm

Improving face recognition using deep autoencoders and feature fusion

Ali Khider, Rafik Djemili, Ahmed Bouridane, Richard Jiang

DOI: 10.1504/IJBM.2022.10043147

Article History:

Received:	08 June 2021
Accepted:	15 September 2021
Published online:	15 December 2022

Improving face recognition using deep autoencoders and feature fusion

Ali Khider* and Rafik Djemili

PIMIS Lab,
Department of Electronics and Telecommunications,
Université 8 Mai 1945 Guelma,
B.P. 401, Guelma, Algeria
Email: ali.khider@hotmail.com
Email: djemili_rafik@yahoo.fr
*Corresponding author

Ahmed Bouridane

Centre for Data Analytics and Cybersecurity,
University of Sharjah,
Sharjah, UAE
Email: abouridane@sharjah.ac.ae

Richard Jiang

School of Computing and Communication,
Lancaster University,
Lancaster, UK
Email: r.jiang2@lancaster.ac.uk

Abstract: Uncontrolled environments are the main challenges of real face recognition systems, recent success of deep learning and features fusion has led to various performance improvements. This paper proposes a novel scheme called feature autoencoder (FAE), where an autoencoder model is not trained directly from the raw facial images, rather it uses a fusion of features constructed by Gabor filter, local binary pattern and local phase quantisation. For each feature, a linear discriminant analysis is applied to reduce its high dimensionality and a limited adaptive histogram equalisation process is employed for contrast enhancement. The proposed scheme has been evaluated using known datasets such as AR, ORL and YALE, and the experimental results carried out on these databases have been compared using three classifiers: k-nearest neighbour, multiclass support vector machine and softmax classifier, demonstrating the effectiveness of proposed approach and parameters. The experimental results obtained and compared with recent and similar approaches on six databases: ORL, YALE, AR, extended YALE B, CMU PIE, and LFWcrop, suggest that the proposed technique outperforms similar techniques. The recognition rates got from them are 100%, 100%, 99.66%, 99.40%, 97.31%, and 90.68% respectively.

Keywords: uncontrolled environments; face recognition; deep learning; sparse; autoencoder; feature extraction; fusion.

Reference to this paper should be made as follows: Khider, A., Djemili, R., Bouridane, A. and Jiang, R. (2023) 'Improving face recognition using deep autoencoders and feature fusion', *Int. J. Biometrics*, Vol. 15, No. 1, pp.40–58.

Biographical notes: Ali Khider has a Master's in Electrical Engineering from the Medea University in 2013. Currently, he prepares his PhD in Electronics and Biometrics at the Guelma University, Algeria. His research interests are in machine learning, artificial intelligence, biometrics and security.

Rafik Djemili is a Professor of Electronics at the University of Skikda, Algeria. His research interests include: biomedical signal processing, brain-computer-interfaces and biometrics.

Ahmed Bouridane is a Professor of Machine Intelligence and the Director of Cybersecurity and Data Analytics Research Center at the University of Sharjah, UAE. His research interests are in machine learning with applications to imaging for forensics and security, quantitative pathology and bimedical engineering, homeland security and video analytics.

Richard Jiang is a Senior Lecturer in the School of Computing and Communications at Lancaster University, UK. His research interest are artificial intelligence, AI ethics, privacy-preserving machine learning, quantum AI, neuronal computation, AI-automated medical/healthcare services, satellite/aerial image analysis and biodiversity diagnosis.

1 Introduction

Face recognition has gained much interest over the past few decades, this is mainly due to its enormous application areas such as video surveillance, public security and human-computer interaction (Wang et al., 2015; Bowyer, 2004; La Torre et al., 2015) to name a few. The main role of face recognition is to identify a face image from a closed set of several face images stored in a database (Muqeet and Holambe, 2017). Therefore, representing both test face images and those stored in the database would be of a great challenge. The major methods used for face recognition are called holistic methods including principal component analysis (PCA) (Turk and Pentland, 1991), kernel PCA (KPCA) (Schölkopf et al., 1997), linear discriminant analysis (LDA) (Belhumeur et al., 1997) and kernel LDA (KLDA) (Lu et al., 2003). Latterly, based on LDA and PCA, Mandal et al introduced the application of curvelet transform in conjunction with PCA-LDA dimensionality reduction techniques (Mandal et al., 2009). Huang combined both information between rows and columns using two-directional 2DPCA on the fused face images and the optimal discriminative information from column-directional 2DLDA (Huang, 2010). Wen et al proposed an approach for face recognition based on the difference vectors and kernel PCA (DV-KPCA) (Wen et al., 2012). Huang et al proposed a local structure preserving discriminant analysis (LSPDA)

by constructing a local scatter matrix and the between-class scatter matrix to characterise the sub and multi-manifold information, respectively (Huang et al., 2014).

However, PCA and LDA-based methods, also known as eigenfaces and fisherfaces, are statistical linear which are not efficient in uncontrolled environments. Since face images can be interpreted as nonlinear objects (Huang and Guan, 2015), and the performance of these methods deteriorates significantly, especially for multi-view face images (Dora et al., 2017) or under environmental changes such as pose, illuminations, or also when images are acquired in free-uncontrolled environments (Feng, 2016). As a result, many researchers have proposed other feature extraction techniques to mitigate the above limitations. Among these techniques, local binary pattern (LBP) (Ojala et al., 2002), scale-invariant feature transform (SIFT) (Lowe, 2004), Gabor filters (Dora et al., 2017; Štruc and Pavešić, 2010), local phase quantisation (LPQ) (Ojansivu and Heikkilä, 2008), binarised statistical image features (BSIF) (Kannala and Rahtu, 2012), and histograms of oriented gradients (HOG) (Dalal and Triggs, 2005) have been proposed. Nevertheless, these techniques have proven their effectiveness by improving the accuracies in some face recognition problems, they also suffered from a poor representation of nonlinearities inherent to many facial images (Aslan et al., 2017).

Recently, much success has been achieved when judiciously combining local and global methods to provide complementary information for a more effective feature extraction. For example, Zhang et al. (2005) proposed local Gabor binary pattern histogram sequence approach (LGBPHS), which combines Gabor and LBP_s Histogram to capture the appearance variations due to lighting, expression, and ageing. Yu et al. (2010) proposed a Gabor magnitude-phase-based texture representation (GMPTR) based on null space linear discriminant analysis (NLDA). Zhou et al. (2013) described an approach for blurred and low-resolution face images which combine Gabor, LBP_s and LPQ_s (GLL) to ensure that blurring is invariant while able to capture the texture information. Yu et al. (2014) proposed to integrate the mean and standard deviation of the local absolute difference into the feature extraction from the standard LBP_s to improve the classification ability of the extracted features. Based on 2D-DWT, Huang et al. (2015) proposed TWSBF approach to combine pixel and feature-level features using the top-level's wavelet sub-band decomposition in order to make a full use of four top-level's wavelet sub-bands using PCA and LDA for feature dimensionalilty reduction. Guermoui and Mekhalfi (2016) proposed a sparse representation of complete local binary pattern histogram using a sparse representation-based classification and concatenation of complete LBP sing histogram (CLBP_SH) and complete LBP magnitude histogram (CLBP_MH) after the pyramid representation to make the features more global. Fathi et al. (2016) combined global-Gabor-Zernike (GGZ) and the histogram of oriented gradient (HOG) features. However, these methods are not supervised or based on predefined filters, the deep neural networks like the convolutional neural networks (CNN) and the autoencoders (AE) models have recently been providing to learn filters directly and relate with the problem studied (Kan et al., 2014; Sun et al., 2015; Xinhua and Qian, 2015; Vincent et al., 2010). Peng et al. (2015) proposed a discriminative graph regularised extreme learning machine (GELM) for further enhancing the classification performance of extreme learning machine (ELM) neural networks model. Liu et al. (2018) proposed a deep learning framework named enhanced PCA network (EPCANet), based on CNN for image classification used two convolution layers to learn the PCA filters and one spatial pooling in the middle of CNN.

Deep learning usually requires computationally intensive processing and large resources for training to capture patterns, especially when captured under uncontrolled environments to reduce the distortions, in order to learn the patterns of subjects to effectively model the face images (Masi et al., 2016). These conditions may not be achievable in real face recognition systems. This paper proposes a novel face recognition method using a combination of features and deep learning. Rather than using the face images in their original space as an input to the AE, we propose a combination of several features extracted from the original images after a preprocessing stage using a contrast limited adaptive histogram equalisation (CLAHE) (Zuiderveld, 1994) to create a novel representation space to train the AE. The feature employed in this paper is a combination and fusion of local and global features using Gabor and overlapped LBP_s and LPQ_s to ensure more discriminating representation then that the system is invariant under uncontrolled environments. After a reduction of each feature vector independently using the LDA, an AE (Vincent et al., 2010) is trained by the three concatenated and reduced features after a normalisation process using zero mean unit variance and a linear rescaling transformation. Experimental results evaluated on some small and popular benchmark face recognition databases: AR (Martinez and Benavente, 1998), ORL (Samaria and Harter, 1994) and YALE (Belhumeur et al., 1997) are assessed and compared with three classifiers, k-nearest neighbour (KNN), multiclass support vector machine (SVM), and the AE softmax classifier layer to demonstrate the uses of every method in the proposed scheme. Experimental results evaluated on AR, ORL, YALE, extended YALE B (Lee et al., 2005), CMU PIE (Sim et al., 2002), and LFWcrop (Sanderson and Lovell, 2009) compared with the related works in Table 5 shown that the proposed technique is outperforms.

The rest of this paper is organised as follows: the proposed approach including their methods and parameters are presented in Section 2. Section 3 provides the implementation strategy and results of the proposed technique using AR, ORL and YALE databases including a comparative study using KNN, SVM, softmax classifiers, and proposed FAE. Subsection 3.4 presents a comparative analysis of the experimental results against some similar methods using AR, ORL, YALE, extended YALE B, CMU PIE, and LFWcrop databases. Finally, a conclusion and future work are given in Section 4.

2 Proposed approach

The proposed FAE system includes preprocessing, feature extraction, normalisation and the AE model steps as show in Figure 1.

2.1 Preprocessing

A CLAHE algorithm (Zuiderveld, 1994) is used to enhance the contrast locally on small regions of the images, this enhancement is limited by a predefined clipping level (CL) in the range [0, 1] and attempts to overcome the amplification of noise produced by the process under certain conditions. In this paper, CL is fixed to 0.01 as described by Sharma et al. (2015), which provides the best performance after many experiment results, and the best contrast in all databases as shown in Figure 2.

Figure 1 The proposed approach (FAE) for face classification, where N1 and N2 are the zero mean unit variance normalisation and the linear rescaling transformation, respectively (see online version for colours)

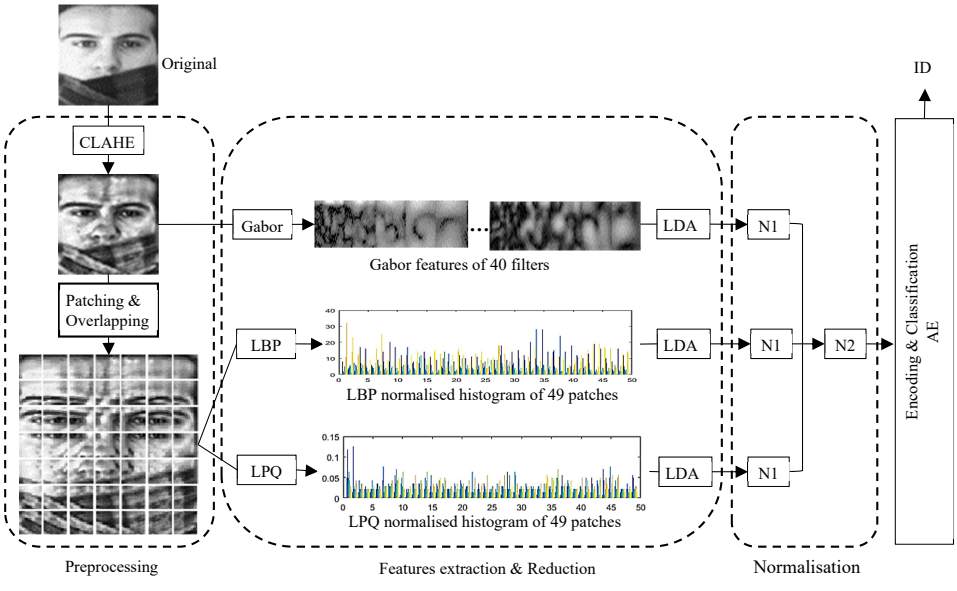
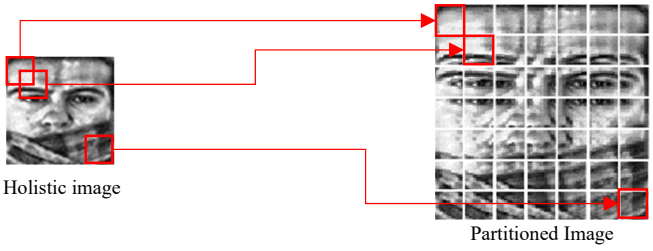


Figure 2 CLAHE enhancement results on, (a) CMU PIE (b) Yale database (c) AR database (see online version for colours)

After the preprocessing stage, a raw image is divided into P patches using a sliding window of $m \times n$ pixels with 50% overlapping as shown in Figure 3.

Figure 3 Example for an image of size 64×64 divided into 49 patches with 50% overlap, where $m \times n = 16 \times 16$ (see online version for colours)



2.2 Feature extraction

In this work, we have chosen to combine Gabor, LBP_s and LPQ_s as input features vector for the AE. In the first step, each raw image is preprocessed with CLAHE method before it is convolved with 40 Gabor filters resulting from five different scales and eight orientations using equation (1) (Štruc and Pavešić, 2010).

$$g(x,y) = \left(\frac{Fu^2}{2\pi}\right) * exp\left(\frac{-Fu^2}{2\pi}\left(xc^2 + yc^2\right)\right) * exp\left(\left(2\pi * Fu * xc\right) i\right) \tag{1}$$

where for each orientation θ and for each scale u,

$$xc = x\cos(\theta) + y\sin(\theta) \tag{2}$$

$$yc = -x\sin(\theta) + y\cos(\theta) \tag{3}$$

and

$$Fu = 0.25/\sqrt{2}^{u} \tag{4}$$

The resulting 40 filtered magnitude images are concatenated to construct a feature vector. In the second step the preprocessed images are divided into 49 patches (Figure 3) and an uniform LBP_s and LPQ_s descriptors with normalised histogram are applied to every patch using the following parameters: for LBP_s settings radius $r = 2$ and number of sample points $n = 16$ and the size of a local uniform window of LPQ_s $w = 5$. Then, all normalised histograms for each feature extraction method are concatenated. Finally, LDA is applied to reduce high dimensionality feature vectors obtained by Gabor, LBP_s and the LPQ_s methods independently, where the feature dimensions become $(Rank - 1)$ of each one, and the final feature dimensions will be $3(Rank - 1)$.

2.3 Normalisation

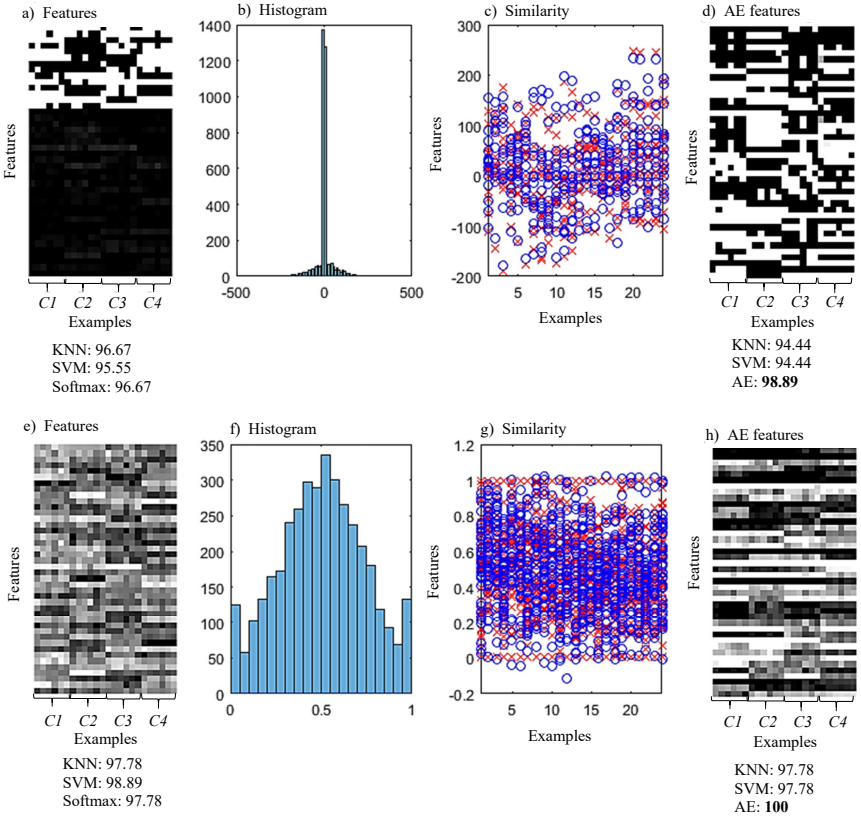
All the three reduced feature vectors are then normalised by applying a zero mean unit variance normalisation on each features vector. Then, these normalised feature vectors are concatenated to one feature referred to x. Once created, a linear rescaling

transformation is applied to x in order to normalise it in the range [0 , 1] using a linear function of equation (5); the resulting feature vector is the fed as input to the AE.

$$y = \frac{x - x_{\min}}{x_{\max} - x_{\min}} \tag{5}$$

This transformation accelerates the encoding process and ensures a similarity between the input and the output of the AE, especially if the logistic sigmoid or the positive saturating linear function are used as transfer functions for the decoder. To show the effect of this normalisation process an evaluation has been carried out on the YALE database as Figure 4.

Figure 4 The results of non-normalised features on top and normalised features on down: shown the first four classes of Yale database, where, (a) (e) are our features, (b) (f) are the histograms of all the database features (c) (g) the similarity between the input and the output of the autoencoder (d) (h) are the encoded features by the autoencoder (see online version for colours)

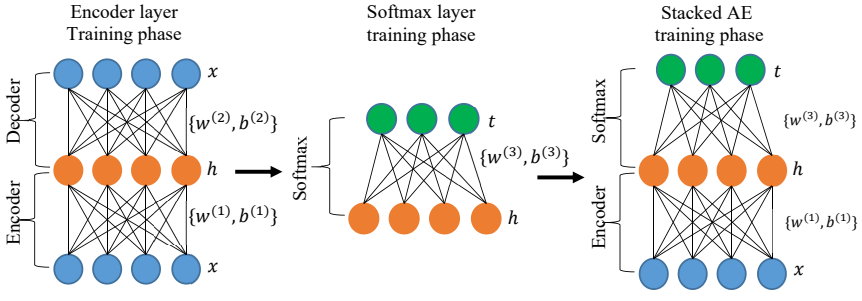


Note: On down of every features the accuracy (%) of all the database obtained by the three classifiers kNN, SVM and softmax.

2.4 Autoencoders

This work proposes to train an AE with two stacked layers: an encoder layer to encode our combined features x in the hidden nodes h and a softmax classifier layer to classify the encoded features h to their class t. Figure 5 shows the three steps for the training process.

Figure 5 The three steps architecture to train the autoencoders (see online version for colours)



2.4.1 Encoder training phase

Firstly, in the encoder layer training phase, we have trained an AE model composed of two layers: an encoder and a decoder are trained to map the input x at its output Figure 5(a). A backpropagation-based scaled conjugate gradient (Møller, 1993) with a cost function based on the mean squared error, the L_2 regularisation and the sparsity regularisation terms have been used to assert the closeness between the input and its reconstructed value. Here, the encoder maps the input x to the hidden nodes through some deterministic mapping function f:

$$f : h = f(x) \tag{6}$$

The decoder maps the hidden nodes back to the original input space through another deterministic mapping function g equation (7):

$$g : \hat{x} = g(h) \tag{7}$$

The global cost function used in the backpropagation (Olshausen and Field, 1997) is:

$$E = MSE + \lambda * \Omega_{weights} + \beta * \Omega_{sparsity} \tag{8}$$

where λ and β are the coefficients of the L_2 regularisation and the sparsity regularisation, respectively. In this work we have fixed $\lambda = 10^{-6}$ as the default value to increase the importance of the sparsity term (Gao et al., 2015) and take β variable in the experiment results. The mean square error in the cost function defined by equation (9):

$$MSE = \frac{1}{N} \sum_{n=1}^{N} \sum_{k=1}^{K} (x_{kn} - \hat{x}_{kn})^2 \tag{9}$$

where N is the number of observations, and K is the number of variables in the output data, in addition, the L_2 regularisation term is:

$$\Omega_{weights} = \frac{1}{2} \sum_{l}^{L} \sum_{n}^{N} \sum_{k}^{K} \left({w_{nk}}^{(l)}\right)^2 \tag{10}$$

where $*^{(l)}$ is the model number l ($l = 1$: the first encoder, $l = 2$: the first decoder), L is the number of hidden layers, K is the number of variables in the input data and $w^{(l)}$ is the l^{th} weight matrix. Finally, the sparsity regularisation term of equation (11) is added to control the sparsity of the output from the hidden layer. The sparsity can be controlled by adding the term Kullback-Leibler divergence (KL) (Olshausen and Field, 1997) that takes a large value when $\hat{p}_i$ of a neuron i and its desired value p are not close in value Figure 6(d); this can cause neurons to be suppressed (Xinhua and Qian, 2015).

$$\Omega_{sparsity} = \sum_{i=1}^{D^{(1)}} KL(p|\,|\hat{p}_i) = \sum_{i=1}^{D^{(1)}} p log\left(\frac{p}{\hat{p}_i}\right) + (1-p)\, log\left(\frac{1-p}{1-\hat{p}_i}\right) \tag{11}$$

where D is the number of neurons in the hidden layer, $\hat{p}_i$ is the average output activation measure the i^{th} neuron of the encoder.

$$\hat{p}_i = \frac{1}{N} \sum_{n=1}^{N} h_{ni} \tag{12}$$

where h_{ni} is the output value of the i^{th} neuron for every input x_n of N training images.

$$h_{ni} = \varphi^{(1)}\left(z_{ni}\right) = \varphi^{(1)}\left({w_i}^{(1)} x_n + {b_i}^{(1)}\right) \tag{13}$$

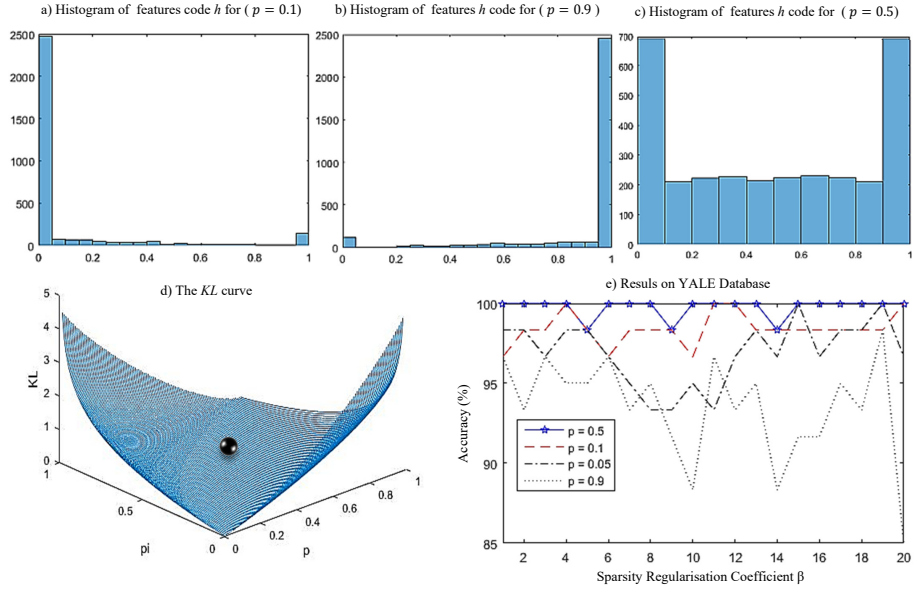
where $\varphi^{(1)}$ is the transfer function for the encoder, ${w_i}^1$ is the i^{th} row of the encoder weight matrix and b_i is the i^{th} entry of the encoder bias vector.

The self-defined parameter p (sparsity proportion) is used as a desired value for every $\hat{p}_i$, its value is in the range [0, 1] to ensure that all the neurons do not fire at values 0 or activated at 1 (Xinhua and Qian, 2015). In this work, we have fixed $p = 0.5$ to make the average output of the neurons close to the centre (0.5) so as to preserve our features to lie between zero and one, Figures 6(a), 6(b) and 6(c). This ensures that our results are stable if the sparsity regularisation coefficient β is varied as shown in Figure 6(e).

In this work, we have chosen a positive saturating linear transfer function for the encoder and a linear transfer function $\varphi(z)^{(2)} = z$ for the decoder. The decoder function g become as shown in equation (14):

$$g : \hat{x}_n = g(h_n) = \varphi(w^{(2)} h_n + b^{(2)})^{(2)} = w^{(2)} h_n + b^{(2)} \tag{14}$$

Figure 6 The evaluation of the sparsity regularisation term in the Yale database (see online version for colours)



2.4.2 Softmax layer training phase

In the second phase, a classification layer (softmax) is trained to map the features code h to the target t (labels) Figure 5(b). The backpropagation based the scaled conjugate gradient (Møller, 1993) has been used to train the classifier with a cross entropy error to calculate the error between the output y and the target t using the following function of equation (15).

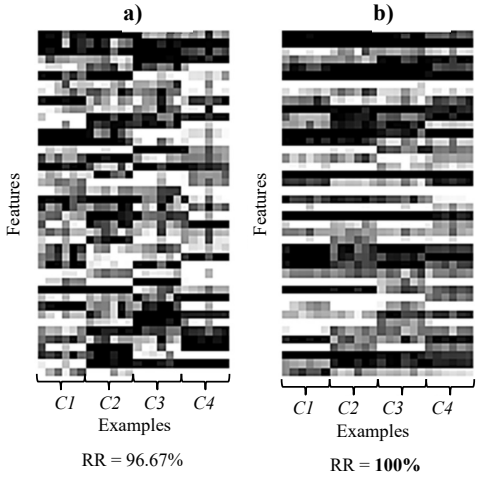
$$E = \frac{1}{N}\sum_{n=1}^{N}\sum_{k=1}^{K} t_{kn} \ln y_{kn} + (1 - t_{kn}) \ln(1 - y_{kn}) \tag{15}$$

where N and K are the number of training examples and the number of classes, respectively.

2.4.3 Stacked autoencoder's training phase

Finally, the encoder layer and the softmax classifier layer are stacked together by retraining them again in the same network by mapping the features x directly into the labels t, to generate the new parameters of the AE network $\{w^{(1)}, b^{(1)}\}$ and $\{w^{(3)}, b^{(3)}\}$ (Vincent et al., 2010). This phase ensures that the features of the hidden layer h are more discriminative for classes as shown in Figure 7.

Figure 7 Feature visualisation from the hidden layer h of the first four classes of YALE database with recognition rates, (a) before retraining the AE (b) after the retraining with the target t (classes)



3 Results and discussion

This section provides the implementation details and parameters setting of the proposed method. We first demonstrate the use of every step in our scheme where KNN, SVM and softmax classifiers have been implemented to validate the power of our proposed AE for feature encoding and classification. In these experiments, all the images from all the datasets are converted to greyscale, cropped and resized to 64 by 64 pixels. The performance of our system is measured using the accuracy of classification calculated by equation (16).

$$Acc = 100 \times \frac{CorrectClassified}{NomberTotal} \tag{16}$$

Three publicly available databases ORL, Yale and AR have been used for evaluating the performance of the proposed system. Tables 1, 2 and 3 show the experimental results, when the proposed: preprocessing (CLAHE), patching (Patch), overlapping (Ovrl) and normalisation (Norm) are used or not. Where in the features column: G, B, Q and L refer to Gabor, LBP_s, LPQ_s, LDA, respectively.

3.1 Experimental results on the ORL face database

The ORL database include 400 face images taken from 40 subjects where each one having ten face images. The images were taken at different time with varying lighting conditions, facial expressions (open/closed eyes, smiling/not smiling), and facial details (glasses/no glasses). For the experiments, as used by most researchers, we have chosen the first four images for training and the rest six images for the test, while the feature dimensions = 117. Table 1 shows the result obtained.

Table 1 The results on ORL database

Features					*Accuracy (%)*			
CLAHE	*Patch*	*Overl*	*Features*	*Norm*	*KNN*	*SVM*	*Softmax*	*AE*
No	No	No	GL+BL+QL	No	98.33	96.25	**99.17**	97.91
No	No	No	GL+BL+QL	Yes	97.5	94.58	**98.33**	97.92
No	16x16	No	GL+BL+QL	No	98.33	96.25	99.17	**99.17**
No	16x16	No	GL+BL+QL	Yes	100	99.17	99.58	**100**
Yes	16x16	No	GL+BL+QL	No	98.33	97.5	98.33	**99.58**
Yes	16x16	No	GL+BL+QL	Yes	97.5	97.08	98.75	**98.75**
No	16x16	50%	GL+BL+QL	No	98.33	96.25	99.17	**99.17**
No	16x16	50%	GL+BL+QL	Yes	**99.58**	99.17	99.58	99.17
Yes	16x16	50%	GL+BL+QL	No	98.33	97.5	98.33	**99.58**
Yes	No	No	GL	Yes	97.91	96.25	98.33	**98.33**
Yes	16x16	50%	BL	Yes	**97.5**	93.75	97.08	95
Yes	16x16	50%	QL	Yes	**100**	97.92	**100**	98.75
Yes	16x16	50%	GL+BL	Yes	97.5	95.83	98.75	98.33
Yes	16x16	50%	GL+QL	Yes	99.58	99.17	99.58	**99.58**
Yes	16x16	50%	BL+QL	Yes	99.17	98.33	99.17	**99.58**
Yes	**16x16**	**50%**	**GL+BL+QL**	**Yes**	100	97.92	99.58	**100**

Table 2 The results on Yale database

Features					*Accuracy (%)*			
CLAHE	*Patch*	*Overl*	*Features*	*Norm*	*KNN*	*SVM*	*Softmax*	*AE*
No	No	No	GL+BL+QL	No	**96.67**	95.56	**96.67**	88.89
No	No	No	GL+BL+QL	Yes	93.33	**94.44**	92.33	93.33
No	16x16	No	GL+BL+QL	No	90	83.33	90	**94.44**
No	16x16	No	GL+BL+QL	Yes	94.44	94.44	94.44	**95.56**
Yes	16x16	No	GL+BL+QL	No	96.67	95.56	96.67	**96.67**
Yes	16x16	No	GL+BL+QL	Yes	94.44	93.33	94.44	**96.67**
No	16x16	50%	GL+BL+QL	No	90	83.33	90	**94.44**
No	16x16	50%	GL+BL+QL	Yes	95.56	94.44	95.56	**95.56**
Yes	16x16	50%	GL+BL+QL	No	96.67	95.56	96.67	**98.89**
Yes	No	No	GL	Yes	**96.67**	95.56	**96.67**	95.56
Yes	16x16	50%	BL	Yes	**93.33**	**93.33**	**93.33**	88.89
Yes	16x16	50%	QL	Yes	93.33	**95.56**	94.44	93.33
Yes	16x16	50%	GL+BL	Yes	97.78	96.67	97.78	**97.78**
Yes	16x16	50%	GL+QL	Yes	97.78	**98.89**	**98.89**	96.67
Yes	16x16	50%	BL+QL	Yes	94.44	**95.56**	94.44	93.33
Yes	**16x16**	**50%**	**GL+BL+QL**	**Yes**	97.78	98.89	97.78	**100**

3.2 Experimental results on the Yale database

The Yale face database contains 165 greyscale images of 15 subjects with each subject having 11 face images which include variations in facial expressions and facial details Figure 2(b). For this dataset, as used by most researchers, we have chosen the first five images for training and the rest six images for the test, while the feature dimensions = 42. Table 2 shows the result obtained.

3.3 Experimental results on the AR database

The AR database consists of over 4,000 frontal images of 126 individual persons. We have chosen a subset containing 50 female and 50 male subjects as used by most researchers. For each individual, 26 images captured in two different sessions with varying facial expressions, illumination conditions and occlusions [Figure 2(c)]. In these experiments, seven non-occluded images are used as the training set with the remaining six occluded images with sunglasses and scarves used as the testing set, while the feature dimensions = 297. Table 3 shows the result obtained.

Table 3 The results on AR S1 dataset

Features					*Accuracy (%)*			
CLAHE	*Patch*	*Overl*	*Features*	*Norm*	*KNN*	*SVM*	*Softmax*	*AE*
No	No	No	GL+BL+QL	No	83.83	81.5	79.83	**88.83**
No	No	No	GL+BL+QL	Yes	66.5	63.67	**82.83**	52.67
No	16x16	No	GL+BL+QL	No	83.83	81.5	79.83	**97.17**
No	16x16	No	GL+BL+QL	Yes	91	85	97.33	**98**
Yes	16x16	No	GL+BL+QL	No	96.5	95.17	96.33	**98.83**
Yes	16x16	No	GL+BL+QL	Yes	98.5	94	**99.33**	98.67
No	16x16	50%	GL+BL+QL	No	83.83	81.5	79.83	**98.5**
No	16x16	50%	GL+BL+QL	Yes	96.5	72.17	98.17	**98.83**
Yes	16x16	50%	GL+BL+QL	No	96.5	95.17	96.33	**99.33**
Yes	No	No	GL	Yes	93.67	89.67	95.5	**95.67**
Yes	16x16	50%	BL	Yes	93.67	90	**98.33**	95.5
Yes	16x16	50%	QL	Yes	**99**	96.83	99.16	98.5
Yes	16x16	50%	GL+BL	Yes	98	94.83	**99.33**	98.83
Yes	16x16	50%	GL+QL	Yes	99	96.83	**99.33**	98.17
Yes	16x16	50%	BL+QL	Yes	98.5	97	99.5	**99.5**
Yes	**16x16**	**50%**	**GL+BL+QL**	**Yes**	98.83	97.67	99.66	**99.66**

As shown in the second part of Tables 1, 2 and 3 the results of combining the three features Gabor + LBP + LPQ consistently outperform the cases when using one or two features separately. The preprocessing also affects the recognition performance especially in the case of Yale Database which has a large illumination variation. In addition, Tables 1, 2 and 3 show that the AE yields higher performances compared by the other classifiers in most cases. To explain these results, Table 4 summarises all the experimental results in terms of the rate of classification improvement. This is illustrated by Acc_{with} which is the average accuracy when using a proposed step, and

$Acc_{without}$ when the proposed step not used, averages accuracies have been computed from Tables 1, 2 and 3.

Table 4 The improvement accuracy of all proposed parts in our approach on all databases (Yale, ORL, and occluded AR)

Proposed steps (methods)	$Acc_{without}$ % *(mean/STD)*	Acc_{wit} % *(mean/STD)*	*Gain %*
Preprocessing	91.83/9.28	97.29/2.25	+5.46
Normalisation	93.76/6.24	96.02/6.29	+2.26
Patching (16x16) and	91.12/10.64 (no patching)	97.06/2.97	+5.94
50% overlapping	95.20/4.97 (no overlapping)		+ 1.86
AE	95.00/6.14	96.41/6.87	+ 1.41
All (FAE)	94.88/6.59	**99.89/0.2**	**+ 5.01**

As shown in Table 4, it can be said that the proposed FAE approach improves the classification rates from 94.88% to 99.89% using these databases. In addition, when compared against KNN, SVM and softmax classifiers the proposed AE encoding outperforms them by at least 1.41%. Furthermore, a performance improvement of 5.46%, 5.94% and 1.84% is achieved when CLAHE, patching and 50% overlapping are used, respectively. Finally, Table 4 shows an improvement gain of +2.26% when a normalisation process is used.

3.4 Results analysis

This section discusses the results of a comparative study of the proposed method against various similar techniques discussed previously in the introduction. Table 5 depicts the results obtained using similar parameters with the following databases.

- *LFWcrop:* Is a cropped version (Sanderson and Lovell, 2009) of Labeled Faces in the Wild (LFW) dataset (Huang et al., 2008), contains 13,233 images of faces collected from the web of 5,749 different individuals. Images in this database exhibit rich intra-personal variations of pose, illumination, and expression. It has been extensively studied for the research of unconstrained face recognition in recent years. Following Liu et al. (2018) in this experiment, a subset of persons who have more than 20 photos but less than 100 photos has been chooses, there are 2,278 images, N = 90% images randomly chosen for train, and the remain T = 10% of each classes for teste.
- *CMU PIE:* This database contains 41,368 images of 68 people where each one is taken under 43 different illuminations, 4 different expressions and 13 different poses, five near frontal poses with all illumination and expression (c05, c07, c09, c27 and c29) are frequently used (Sim et al., 2002).
- *AR1:* The AR database is defined in Subsection 3.3, as described in Huang et al. (2014), Lu et al. (2012) and Fathi et al. (2016) a subset consists of 14 non-occluded images from the two seasons has been selected.

- *AR2:* As described in Lu et al. (2012), eight samples per subject of non-occluded images with various expression from the AR database (Subsection 3.3) are used for training while the others from scarves and sunglasses are used for testing.
- *AR3:* Using the protocol (Wen et al., 2012) we selected 13 images from season 1 to test our method, the original images are normalised to 128×128 we selected $N = 6$ samples from each class for the training, and the remaining ($T = 7$) for testing.
- *AR4:* All the 26 images are used from both seasons 1 and 2 as described in Huang et al. (2015), where each image is cropped and resized to 50×40 pixels, N is set to 9 for training and $T = 17$ for testing.
- *AR5:* In this experiment a subset as described in Peng et al. (2015) and Liu et al. (2018) is used, for each subject only the images with illumination and expression changes have been used. Seven images from session 1 are used for training ($N = 7$), and the other seven images from session 2 for testing ($T = 7$).
- *The extended YALE B database* (Lee et al., 2005) consists of 2,414 face images of 38 individuals. Each one has around 64 near frontal images under different illuminations.

To evaluate the effectiveness of the proposed FAE, different values ranging from 10 to 20 were attributed to the sparsity regularisation coefficients β of equation (8) by randomly selecting N images for training and keeping the remaining T images for testing. The experiments are repeated 20 times independently resulting in 220 sub-tests. The results are then averaged and the standard deviations were computed as shown in column 4. In Table 5, the parameters and the results of related works are shown in column 2 and 3, respectively. F uses the First T images for training while R means a random selection.

As shown in Table 5, the recognition rates (RRs) of the proposed FAE outperforms all other techniques. For example, in the case of illumination and expression changes (using CMU PIE, extended Yale and AR1 datasets). The performance improvement is significant attaining of 99.40% when using extended Yale compared with 95.99% and 93.91% for GELM and EPCANet methods, respectively. Furthermore, a RR of 97.31% using the whole CMU PIE dataset and 98.68% on the subset C29 have been achieved against a RR of GELM 93.47% and 90.60% for GELM and LSPDA, respectively. For the occlusion distortions (AR2), the experiments have been implemented with scarves and a RR of 98.31% and 99.10% when sunglass is included. For the unconstrained environments on LFWcrop dataset, a RR of 90.68% has been achieved by the proposed FAE compared with 88.53% by EPCANet. Finally, the proposed algorithm has achieved the best recognition performances when using datasets with less poses variation as in the case of Yale and ORL datasets. Therefore, these clearly demonstrates that the proposed method outperforms existing methods.

Table 5 The comparison of performances

Authors and year	*Database and parameters*	*Evaluation technique (RR%) and <Features dimensions>*	*Proposed FAE RR ± STD, <Feat-dim>*
Huang et al. (2014)	ORL (112x92), N = F5, T = 5	LDP (91.0) <36> LPP (90.5) <52> LSPDA (95.5) <26>	**100** (64x64) <117>
	AR1 (50x40), N = 5R, T = 9	LDP (74.8) <150> LPP (69.7) <170> LSPDA (86.6) <195>	**99.39** ± 0.2 <297>
	PIE-C29 (64x64), N = 10R, T = 14	LDP (78.7) <95> LPP (78.9) <120> LSPDA (90.6) <195>	**98.68** ± 0.33 <201>
Yu et al. (2014)	ORL (56, 46), N = F5, T = 5	Improved LBP (91.5) </>	**100** <117>
Huang (2010)	ORL (80x80), N = 4R, T = 6	2DPCA (90.27) <20> 2DLDA (91.63) <20> 2DPCA+2DLDA (92.30) <20>	**99.18** ± 1.3 (64x64) <117>
	Yale (100x100), N = 6R, T = 5	2DPCA (92.32) <20> 2DLDA (90.51) <20> 2DPCA+2DLDA (94.88) <20>	**98.11** ± 1.29 (64x64) <42>
Lu et al. (2012)	ORL (32x32), N = 5R, T = 5	DSNPE (96.0) <40>	**98.68** ± 1.04 <117>
	Yale (32x32), N = F6, T = 5	Gabor + DSNPE +SRC (95.1) <20>	**100** <42>
	AR1 (33x24), N = 5R, T = 9	DSNPE (96.1) <300>	**98.04** ± 0.36 <297>
	AR2 Scarves (50x40)	DSNPE+NN (20.5) <300> DSNPE+SRC (59.5) <300>	**98.31** ± 0.49 <297>
	AR2 Sunglass (50x40)	DSNPE+NN (70.0) <300> DSNPE+SRC (87.0) <300>	**99.10** ± 0.33 <297>
Yu et al. (2010)	ORL (64x64), N = 4R, T = 6	GMPTR + NLDA (97.29) <39>	**99.18** ± 1.3 <117>
	Yale (64x64), N = 5R, T = 6	GMPTR + NLDA (93.67) <14>	**97.62** ± 1.64 <42>
Wen et al. (2012)	ORL (112x92), N = 4R, T = 6	DV-KPCA (93.45) <80>	**99.18** ± 1.3 (64x64) <117>
	Yale (100x80), N = 6R, T = 5	DV-KPCA (95, 13) <40>	**98.11** ± 1.29 (64x64) <42>
	AR3 (128x128), N = 6R, T = 7	DV-KPCA (91.45) </>	**99.83** ± 0.16 <297>
Huang et al. (2015)	ORL (112x92), N = 4R, T = 6	TWSBF + LDA (95.04) </>	**99.18** ± 1.3 (64x64) <117>
	AR4 (50x40), N = 9R, T = 17	TWSBF + LDA (94.23) </>	**99.58** ± 0.18 <297>
Mandal et al. (2009)	ORL (/), N = 5R, T = 5	Curvelet+ PCA+ LDA (97.7) <60>	**99.68** ± 0.68 (64x64) <117>
	Yale (/), N = 6R, T = 5	Curvelet+ PCA+ LDA (92.0) <60>	**98.11** ± 1.29 (64x64) <42>
Peng et al. (2015)	ORL (32x32), N = 5R, T = 5	GELM (96.34) <199>	**98.68** ± 1.04 <117>
	Yale (32x32), N = 5R, T = 6	GELM (82.36) <69>	**94.36** ± 2.32 <42>
	Extended Yale B (32x32), N = 20R, T≈44	GELM (95.99) <759>	**99.40** ± 0.14 <111>
	PIE-All (32x32), N = 20R, T≈150	GELM (93.47) <1379>	**97.31** ± 0.2 <201>
	AR5 (60x43), N = 7, T = 7	GELM (93.85) <300>	**98.80** ± 1.6 <297>
Fathi et al. (2016)	ORL (92x92), N = 5R, T = 5	GGZ + HOG (98.0) <90>	**99.68** ± 0.68 (64x64) <117>
	Yale (100x100), N = 5R, T = 6	GGZ + HOG (97.8) <90>	**98.75** ± 1.21 <42>
	AR1 (120x120), N = 7R, T = 7R	GGZ + HOG (97.1) <90>	**99.91** ± 0.13 <297>
Liu et al. (2018)	AR5 (50, 40), N = 7, T = 7	EPCANet (94.40) </>	**98.64** ± 0.18 <297>
	Extended Yale B (32x32), N = 20, T≈44	EPCANet (93.91) </>	**99.40** ± 0.14 <111>
	LFWcrop (64x64), N = 90%, T = 10%	EPCANet (88.53) </>	**90.68** ± 2.45 <183>

4 Conclusions

This paper proposed a face recognition approach using an improved feature with deep learning employing autoencoders as a modified deep learning methodology. The approach consists of combining local and global features extracted using three feature extraction techniques using a simple fusion technique. The autoencoder is trained using two layers one acting as an encoder and the second as a classifier. This allows the technique to map the input using a deterministic approach to generate the best features hence to address the face distortions. Extensive experiments were conducted to validate the approach and demonstrate the discriminative power as shown by the much improved RRs obtained using various face datasets used by the research community. It has been shown that the proposed method is robust to occlusion, pose and illumination variations. The experimental results carried out using six databases demonstrated the efficacy of our proposed approach which uses only the frontal gallery images without any additional information from the unconstrained images. As a recent work, we create a CNN architecture for the big data of face and iris recognition. As a future work, we will investigate the use of face alignment to make the system invariant to large pose variations. We will also investigate a feature selection using different methods that are less invariant to age variations.

References

Aslan, M.S., Hailat, Z., Alafif, T.K. and Chen, X-W. (2017) 'Multi-channel multi-model feature learning for face recognition', *Pattern Recognition Letters*, January, Vol. 85, pp.79–83.

Belhumeur, P.N., Hespanha, J.P. and Kriegman, D.J. (1997) 'Eigenfaces vs. fisherfaces: recognition using class specific linear projection', *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 19, No. 7, pp.711–720.

Bowyer, K.W. (2004) 'Face recognition technology: security versus privacy', *IEEE Technology and Society Magazine*, Vol. 23, No. 1, pp.9–19.

Dalal, N. and Triggs, B. (2005) 'Histograms of oriented gradients for human detection', *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, IEEE, Vol. 1, pp.886–893.

Dora, L., Agrawal, S., Panda, R. and Abraham, A. (2017) 'An evolutionary single Gabor kernel based filter approach to face recognition', *Engineering Applications of Artificial Intelligence*, June, Vol. 62, pp.286–301.

Fathi, A., Alirezazadeh, P. and Abdali-Mohammadi, F. (2016) 'A new global-Gabor-Zernike feature descriptor and its application to face recognition', *Journal of Visual Communication and Image Representation*, July, Vol. 38, pp.65–72.

Feng, S. (2016) 'Robust face recognition under varying illumination and occlusion via single layer networks', *Chinese Conference on Biometric Recognition*, Springer, pp.93–101.

Gao, S., Zhang, Y., Jia, K., Lu, J. and Zhang, Y. (2015) 'Single sample face recognition via learning deep supervised autoencoders', *IEEE Transactions on Information Forensics and Security*, Vol. 10, No. 10, pp.2108–2118.

Guermoui, M. and Mekhalfi, M.L. (2016) *A Sparse Representation of Complete Local Binary Pattern Histogram for Human Face Recognition*, arXiv preprint arXiv:1605.09584.

Huang, G. (2010) 'Fusion (2D)2pcalda: a new method for face recognition', *Applied Mathematics and Computation*, Vol. 216, No. 11, pp.3195–3199.

Huang, G.B., Mattar, M., Berg, T. and Learned-Miller, E. (2008) 'Labeled Faces in the Wild: a database forstudying face recognition in unconstrained environments', in *Workshop on Faces in 'Real-Life' Images: Detection, Alignment, and Recognition*, October.

Huang, P., Chen, C., Tang, Z. and Yang, Z. (2014) 'Feature extraction using local structure preserving discriminant analysis', *Neurocomputing*, 22 September, Vol. 140, pp.104–113.

Huang, Y. and Guan, Y. (2015) 'On the linear discriminant analysis for large number of classes', *Engineering Applications of Artificial Intelligence*, August, Vol. 43, pp.15–26.

Huang, Z-H., Li, W-J., Wang, J. and Zhang, T. (2015) 'Face recognition based on pixel-level and feature-level fusion of the top-level's wavelet sub-bands', *Information Fusion*, March, Vol. 22, pp.95–104.

Kan, M., Shan, S., Chang, H. and Chen, X. (2014) 'Stacked progressive auto-encoders (SPAE) for face recognition across poses', *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp.1883–1890.

Kannala, J. and Rahtu, E. (2012) 'BSIF: binarized statistical image features', in *Proceedings of the 21st International Conference on Pattern Recognition (ICPR'2012)*, IEEE, pp.1363–1366.

La Torre, M.D., Granger, E., Radtke, P.V., Sabourin, R. and Gorodnichy, D.O. (2015) 'Partially-supervised learning from facial trajectories for face recognition in video surveillance', *Information Fusion*, July, Vol. 24, pp.31–53.

Lee, K-C., Ho, J. and Kriegman, D.J. (2005) 'Acquiring linear subspaces for face recognition under variable lighting', *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 27, No. 5, pp.684–698.

Liu, Y., Zhao, S., Wang, Q. and Gao, Q. (2018) 'Learning more distinctive representation by enhanced pca network', *Neurocomputing*, 31 January, Vol. 275, pp.924–931.

Lowe, D.G. (2004) 'Distinctive image features from scale-invariant keypoints', *International Journal of Computer Vision*, Vol. 60, No. 2, pp.91–110.

Lu, G-F., Jin, Z. and Zou, J. (2012) 'Face recognition using discriminant sparsity neighborhood preserving embedding', *Knowledge-Based Systems*, July, Vol. 31, pp.119–127.

Lu, J., Plataniotis, K.N. and Venetsanopoulos, A.N. (2003) 'Face recognition using kernel direct discriminant analysis algorithms', *IEEE Transactions on Neural Networks*, Vol. 14, No. 1, pp.117–126.

Mandal, T., Wu, Q.J. and Yuan, Y. (2009) 'Curvelet based face recognition via dimension reduction', *Signal Processing*, Vol. 89, No. 12, pp.2345–2353, Special Section: Visual Information Analysis for Security.

Martinez, A. and Benavente, R. (1998) *The AR Face Database*, CVC Technical Report, Vol. 24.

Masi, I., Trần, A.T., Hassner, T., Leksut, J.T. and Medioni, G. (2016) 'Do we really need to collect millions of faces for effective face recognition?', in *European Conference on Computer Vision*, Springer, Cham, October, pp.579–596.

Møller, M.F. (1993) 'A scaled conjugate gradient algorithm for fast supervised learning', *Neural Networks*, Vol. 6, No. 4, pp.525–533.

Muqeet, M.A. and Holambe, R.S. (2017) 'Local appearance-based face recognition using adaptive directional wavelet transform', *Journal of King Saud University – Computer and Information Sciences*, April, Vol. 31, No. 2, pp.161–174.

Ojala, T., Pietikainen, M. and Maenpaa, T. (2002) 'Multiresolution gray-scale and rotation invariant texture classification with local binary patterns', *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 24, No. 7, pp.971–987.

Ojansivu, V. and Heikkilä, J. (2008) 'Blur insensitive texture classification using local phase quantization', in *International Conference on Image and Signal Processing*, Springer, Berlin, Heidelberg, July, pp.236–243.

Olshausen, B.A. and Field, D.J. (1997) 'Sparse coding with an overcomplete basis set: a strategy employed by v1?', *Vision Research*, Vol. 37, No. 23, pp.3311–3325.

Peng, Y., Wang, S., Long, X. and Lu, B-L. (2015) 'Discriminative graph regularized extreme learning machine and its application to face recognition', *Neurocomputing*, 3 February, Vol. 149, pp.340–353, Advances in Neural Networks, Advances in Extreme Learning Machines.

Samaria, F.S. and Harter, A.C. (1994) 'Parameterisation of a stochastic model for human face identification', in *Proceedings of 1994 IEEE Workshop on Applications of Computer Vision*, IEEE, December, pp.138–142.

Sanderson, C. and Lovell, B.C. (2009) 'Multi-region probabilistic histograms for robust and scalable identity inference', *International Conference on Biometrics*, Springer, pp.199–208.

Schölkopf, B., Smola, A. and Müller, K-R. (1997) 'Kernel principal component analysis', in Gerstner, W., Germond, A., Hasler, M. and Nicoud, J-D. (Eds.): *Artificial Neural Networks – ICANN'97*, pp.583–588, Springer Berlin Heidelberg, Berlin, Heidelberg.

Sharma, N., Saurav, S., Singh, S., Saini, R. and Saini, A.K. (2015) 'A comparative analysis of various image enhancement techniques for facial images', in *2015 International Conference on Advances in Computing, Communications and Informatics (ICACCI)*, IEEE, August, pp.2279–2284.

Sim, T., Baker, S. and Bsat, M. (2002) 'The CMU pose, illumination, and expression (PIE) database', in *Proceedings of Fifth IEEE International Conference on Automatic Face Gesture Recognition*, IEEE, May, pp.53–58.

Štruc, V. and Pavešić, N. (2010) 'The complete gabor-fisher classifier for robust face recognition', *EURASIP Journal on Advances in Signal Processing*, Vol. 2010, No. 1, p.847680.

Sun, Y., Liang, D., Wang, X. and Tang, X. (2015) *DeepID3: Face Recognition with Very Deep Neural Networks*, arXiv preprint arXiv:1502.00873.

Turk, M. and Pentland, A. (1991) 'Eigenfaces for recognition', *Journal of Cognitive Neuroscience*, Vol. 3, No. 1, pp.71–86.

Vincent, P., Larochelle, H., Lajoie, I., Bengio, Y. and Manzagol, P-A. (2010) 'Stacked denoising autoencoders: learning useful representations in a deep network with a local denoising criterion', *Journal of Machine Learning Research*, December, Vol. 11, No. 12, pp.3371–3408.

Wang, G., Zheng, F., Shi, C., Xue, J-H., Liu, C. and He, L. (2015) 'Embedding metric learning into set-based face recognition for video surveillance', *Neurocomputing*, 3 March, Vol. 151, pp.1500–1506.

Wen, Y., He, L. and Shi, P. (2012) 'Face recognition using difference vector plus KPCA', *Digital Signal Processing*, Vol. 22, No. 1, pp.140–146.

Xinhua, L. and Qian, Y. (2015) 'Face recognition based on deep neural network', *Int. J. Signal Processing, Image Processing and Pattern Recognition*, Vol. 8, No. 10, pp.29–38.

Yu, L., He, Z. and Cao, Q. (2010) 'Gabor texture representation method for face recognition using the gamma and generalized gaussian models', *Image and Vision Computing*, Vol. 28, No. 1, pp.177–187.

Yu, W., Gan, L., Yang, S., Ding, Y., Jiang, P., Wang, J. and Li, S. (2014) 'An improved lbp algorithm for texture and face classification', *Signal, Image and Video Processing*, Vol. 8, No. 1, pp.155–161.

Zhang, W., Shan, S., Gao, W., Chen, X. and Zhang, H. (2005) 'Local gabor binary pattern histogram sequence (LGBPHS): a novel non-statistical model for face representation and recognition', *Tenth IEEE International Conference on Computer Vision, 2005. ICCV 2005*, IEEE, Vol. 1, pp.786–791.

Zhou, S-R., Yin, J-P. and Zhang, J-M. (2013) 'Local binary pattern (LBP) and local phase quantization (LBQ) based on gabor filter for face representation', *Neurocomputing*, 20 September, Vol. 116, pp.260–264, Advanced Theory and Methodology in Intelligent Computing.

Zuiderveld, K. (1994) 'Graphics gems iv', *Contrast Limited Adaptive Histogram Equalization*, pp.474–485, Academic Press Professional, Inc., San Diego, CA, USA.