

A multimodal deep learning framework using local feature representations for face recognition

Alaa S. Al-Waisy¹ · Rami Qahwaji¹ · Stanley Ipson¹ · Shumoos Al-Fahdawi¹

Received: 5 February 2017 / Revised: 19 July 2017 / Accepted: 29 July 2017 / Published online: 4 September 2017
© The Author(s) 2017. This article is an open access publication

Abstract The most recent face recognition systems are mainly dependent on feature representations obtained using either local handcrafted-descriptors, such as local binary patterns (LBP), or use a deep learning approach, such as deep belief network (DBN). However, the former usually suffers from the wide variations in face images, while the latter usually discards the local facial features, which are proven to be important for face recognition. In this paper, a novel framework based on merging the advantages of the local handcrafted feature descriptors with the DBN is proposed to address the face recognition problem in unconstrained conditions. Firstly, a novel multimodal local feature extraction approach based on merging the advantages of the Curvelet transform with Fractal dimension is proposed and termed the Curvelet–Fractal approach. The main motivation of this approach is that the Curvelet transform, a new anisotropic and multidirectional transform, can efficiently represent the main structure of the face (e.g., edges and curves), while the Fractal dimension is one of the most powerful texture descriptors for face images. Secondly, a novel framework is proposed, termed the multimodal deep face recognition (MDFR) framework, to add feature representations by training a DBN on top of the local feature representations instead of the pixel intensity representations. We demonstrate that representations

acquired by the proposed MDFR framework are complementary to those acquired by the Curvelet–Fractal approach. Finally, the performance of the proposed approaches has been evaluated by conducting a number of extensive experiments on four large-scale face datasets: the SDUMLA-HMT, FERET, CAS-PEAL-R1, and LFW databases. The results obtained from the proposed approaches outperform other state-of-the-art of approaches (e.g., LBP, DBN, WPCA) by achieving new state-of-the-art results on all the employed datasets.

Keywords Face recognition · Curvelet transform · Fractal dimension · Fractional Brownian motion · Deep belief network · SDUMLA-HMT database · FERET database · LFW database

1 Introduction

In the recent years, there has been a growing interest in highly secured and well-designed face recognition systems, due to their potentially wide applications in many sensitive places such as controlling access to physical as well as virtual places in both commercial and military associations, including ATM cash dispensers, e-learning, information security, intelligent surveillance, and other daily human applications [1]. In spite of the significant improvement in the performance of face recognition over previous decades, it still a challenging task for the research community, especially when face images are taken in unconstrained conditions due to the large intra-personal variations such as changes in facial expression, pose, illumination, aging, and the small inter-personal differences. Face recognition systems encompass two fundamental stages: feature extraction and classification. The second stage is dependent on the first. Therefore, the

✉ Alaa S. Al-Waisy
king_alaa87@yahoo.com

Rami Qahwaji
R.S.R.Qahwaji@bradford.ac.uk

Stanley Ipson
S.S.Ipson@Bradford.ac.uk

Shumoos Al-Fahdawi
sun_algold@yahoo.com

¹ School of Computing, Informatics and Media,
University of Bradford, Bradford, UK

task of extracting and learning useful and highly discriminating facial features in order to minimize intra-personal variations and maximize interpersonal differences is complicated.

In this regard, a number of approaches have been proposed, implemented, and refined to address all these drawbacks and problems in the face recognition system. These approaches can be divided into two categories: local handcrafted-descriptor approaches and deep learning-based approaches. Local handcrafted-descriptor approaches can be further divided into four groups: feature-based, holistic-based, learning-based and hybrid-based approaches [2]. In the first category, a geometric vector representing the facial features is extracted by measuring and computing the locations and geometric relationships among facial features, such as the mouth, eyes and nose, and using it as an input to a structural classifier. The elastic bunch graph matching (EBGM) system is an example of a features-based method, which uses the responses of Gabor filters at different orientations and frequencies at each facial feature point to extract a set of local features [3,4]. Compared with the feature-based approaches, the holistic methods usually extract the feature vector by operating on the whole face image instead of measuring the local geometric features. The eigenface methods are the best well-known examples of these approaches, which are represented by principal component analysis (PCA), independent component analysis (ICA), etc. [5]. The third learning-based approaches learn features from labeled training samples using machine learning techniques. Finally, the hybrid approaches are based on combinations of two or more of these categories. Some examples of the third and fourth categories can be found in [6–8]. Previous research has demonstrated the efficiency of local handcrafted-descriptor approaches used as robust and discriminative feature detectors to solve the face recognition problem even when relatively few training samples per person are available, as in [9–11]. However, the performance using local handcrafted-descriptors approaches declines dramatically in unconstrained conditions due to fact that the constructed face representations are very sensitive to the highly nonlinear intra-personal variations, such as expression, illumination, pose, and occlusion [12]. To address these drawbacks, considerable attention has been paid to the use of deep learning approaches (e.g., deep neural networks) to automatically learn a set of effective feature representations through hierarchical nonlinear mappings, which can robustly handle the nonlinear variations (intra- and interpersonal variations) of face images. Moreover, in contrast to handcrafted-descriptor approaches, the applications making use of deep learning approaches can generalize well to other new fields [13]. The DBN is one of the most popular unsupervised deep learning methods, which has been

successfully applied to learn a hierarchical representations from unlabeled data in a wide range of fields, including face recognition [14], speech recognition [15], audio classification [16], and natural language understanding [17]. However, a key limitation of the DBN when the pixel intensity values are assigned directly to the visible units is that the feature representations of the DBN are sensitive to the local translations of the input image. This can lead to disregarding local features of the input image known to be important for face recognition. Furthermore, scaling the DBN to work with realistic-sized images (e.g., 128×128) is computationally expensive and impractical. To improve the generalization ability and reduce the computational complexity of the DBN, a novel framework based on merging the advantages of the local handcrafted feature descriptors with the DBN is proposed to address the face recognition problem in unconstrained conditions. We argue that applying the DBN on top of preprocessed image feature representations instead of the pixel intensity representations (raw data), as a way of guiding the learning process, can greatly improve the ability of the DBN to learn more discriminating features with less training time required to obtain the final trained model. To the authors' best knowledge, very few publications can be found in the literature that discuss the potential of applying the DBN on top of preprocessed image feature representations. Huang et al. [18] have demonstrated that applying the convolutional DBN on top of the output of LBP can increase the accuracy rate of the final system. Li et al. [19] have also reached to the same conclusion by applying the DBN on top of center-symmetric local binary pattern (CS-LBP). However, the work in [18] was applied only to the face verification task, while the work in [19] was evaluated on a very small face dataset where the face images were taken in controlled environments. The primary contributions of the work presented here can be summarized as follows:

1. A novel multimodal local feature extraction approach based on merging the advantages of multidirectional and anisotropy transforms, specifically the Curvelet transform, with Fractal dimension is proposed. Termed the Curvelet–Fractal approach, it is different from previously published Curvelet-based face recognition systems, which extract only the global features from the face image. The proposed method has managed to extract the local features along with the face texture roughness and fluctuations in the surface efficiently by exploiting the Fractal dimension properties, such as self-similarity. There are three main differences from the previous conference version [20] of this work. Firstly, unlike [20] which used only the coarse band of the Curvelet transform as an input to the Fractal dimension stage, here we also use the other Curvelet sub-bands features, which rep-

resent the most significant information in the face image (e.g., face curves), which are known to be crucial in the recognition process. Secondly, a new Fractal dimension method is proposed based on an improved differential box counting (IDBC) method in order to calculate the Fractal dimension values from the new added Curvelet sub-bands and handle their high dimensionality. Then, the outputs of the IDBC and fractional Brownian motion (FBM) are combined to build an elementary feature vector. Finally, we propose to use the quadratic discriminant classifier (QDC) instead of K-nearest neighbor (K-NN) because this improves the accuracy of the proposed system.

2. A novel framework is proposed, termed the multimodal deep face recognition (MDFR) framework, to learn additional and complementary features representations by training a deep neural network (e.g., DBN) on top of a Curvelet–Fractal approach instead of the pixel intensity representation. We demonstrate that, the proposed framework can represent large face images, with the time required to obtain the final trained model significantly reduced compared to the direct use of the raw data. Furthermore, the proposed framework is able to efficiently handle the nonlinear variations (intra- and interpersonal variations) of face images and is unlikely to over fit to the training data due to the nonlinearity of a DBN.
3. The performance of the proposed approaches has been evaluated by conducting a number of extensive experiments on four large-scale unconstrained face datasets: SDUMLA-HMT, FERET, CAS-PEAL-R1, and LFW. We are able to achieve a comparable recognition rate to state-of-the-art methods using the Curvelet–Fractal approach. We demonstrate that the feature representations acquired by the DBN as a deep learning approach is complementary to the feature representations acquired by the Curvelet–Fractal approach to handcrafted-descriptors. Thus, the state-of-the-art results on the employed datasets have been farther improved by combining these two representations.

This paper focuses mainly on two different problems in the face recognition system: face identification and face verification. In this paper, the term face recognition will be used in the general case to refer to these two problems. The remainder of the paper is organized as follows: Sect. 2 is devoted to providing an overview of the proposed handcrafted-descriptors and deep learning approaches. Section 3 shows the implementation details of the proposed approaches. The experimental results are presented in Sect. 4. Finally, conclusions and future research directions are stated in the last section.

2 Background

In this section, a brief description of the proposed approaches is presented, including the Curvelet transform and Fractal dimension method used in the proposed multimodal local feature extraction approach. In addition, the proposed deep learning approaches includes the DBN and its building block the restricted Boltzmann machine (RBM) as well. The primary goal here is to review and recognize their strengths and shortcomings to empower the proposal of a novel face recognition framework that consolidates the strengths of these approaches.

2.1 Curvelet transform

In recent years, many multiresolution approaches have been proposed for facial feature extraction at different scales, aiming to improve face recognition performance. The wavelet transform is one of the most popular multiresolution feature extraction methods due to its ability to provide significant features in both space and transform domains. However, according to many studies in the human visual system and image analysis, the wavelet transform is not ideal for facial feature extraction. A feature extraction method cannot be optimal without satisfying conditions relating to the following: multiresolution, localization, critical sampling, directionality, and anisotropy. It is believed that the wavelet transform cannot fulfill the last two conditions due to limitations of its basis functions in specifying direction and the isotropic scale [21]. These restrictions lead to weak representation of the edges and curves which are considered to be the most important facial features. Thus, a novel transform was developed by Candes and Donoho in 1999 known as the Curvelet transform [22]. Their motivation was to overcome the drawbacks and limitations of widely used multiresolution methods such as the wavelet and ridgelet transforms. All the above five conditions can be fulfilled using Curvelet transform.

The Curvelet transform has been successfully applied to solve many problems in the image processing area such as texture classification [23], preserving edges and image enhancement [24], image compression [25], image fusion [26], and image de-noising [27]. Some work has been done to explore the potential of the Curvelet transform to help solve pattern recognition problems, for example by Lee and Chen [28], Mandal and Wu [29] and Xie [30]. These showed that the Curvelet transform can serve as a good feature extraction method for pattern recognition problems like fingerprint and face recognition due to its ability to represent crucial edges and curve features more efficiently than other transformation methods. However, the Curvelet transform suffers from the effects of significant variety in pose, lighting condi-

tions, shadows, and occlusions from wearing glasses or hats. Hence, the Curvelet transform is not able to describe the face texture roughness and fluctuations in the surface efficiently, which will have a significant effect on the recognition rate. All these factors together were behind the adoption here of the Fractal dimension to provide a better description of the face texture under unconstrained environmental conditions.

2.2 Fractal dimension

The term Fractal dimension was first introduced by the mathematician Benoit Mandelbrot as a geometrical quantity to describe the complexity of objects that show self-similarity at different scales [31]. The Fractal dimension has some important properties such as a self-similarity, which means that an object has a similar representation to the original under different magnifications. This property can be used in reflecting the roughness and fluctuation of image’s surface where increasing the scale of magnification provides more and more details of the imaged surface. In addition, the non-integer value of the Fractal dimension gives a quantitative measure of objects that have complex geometry and cannot be well described by an integral dimension (such as the length of a coastline) [32, 33]. Many methods have been proposed to calculate Fractal dimension, such as box counting (BC), differential box counting (DBC) and fractional Brownian motion (FBM), and other methods can be found here [31]. The Fractal dimension has been widely applied in many areas of image processing and computer vision, such as texture segmentation [34] medical imaging [35] face detection [36]. However, not much work has been done to explore and address the potential of using the Fractal dimension to resolve pattern recognition problems. Lin et al. [37] proposed an algorithm for human eye detection by exploiting the Fractal dimension as an efficient approach for representing the texture of facial features. Farhan et al. [38] developed a personal identification system based on fingerprint images using the Fractal dimension as a feature extraction method. Therefore, it appears that the texture of the facial image can be efficiently described by using the Fractal dimension. However, Fractal estimation methods are very time consuming and cannot meet real-time requirements. To address all the limitations and drawbacks in (Sects. 2.1, 2.2), a novel face recognition algorithm based on merging the advantages of a multidirectional and anisotropy transform, specifically the Curvelet transform, with Fractal dimension is proposed.

2.3 Deep learning approaches

In 2006, a new deep neural networks (DNN) was introduced called the deep belief network (DBNs) by Hinton et al. [39]. DBN is a generative probabilistic model that differs from conventional discriminative neural networks. DBNs are com-

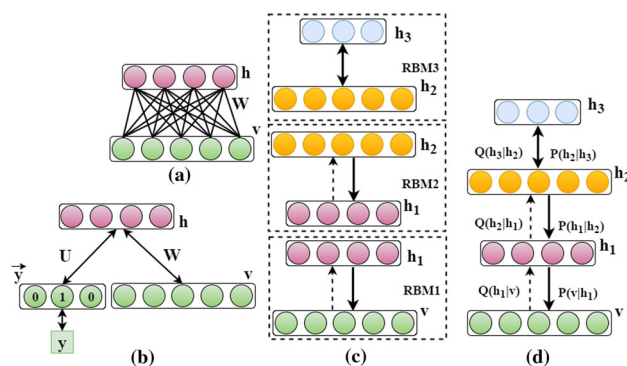


Fig. 1 **a** A typical RBM structure, **b** A discriminate RBM modeling the joint distribution of input variables and target classes, **c** Greedy layer-wise training algorithm for the DBN composed of three stacked RBMs, and **d** Three layers of the DBN as a generative model, where the top-down generative path is represented by the **P** distributions (*Solid arcs*), and bottom-up inference and training path is represented by the **Q** distributions (*Dashed arcs*)

posed of one visible layer (observed data) and many hidden layers that have the ability to learn the statistical relationships between the units in the previous layer. As depicted in Fig. 1c, a DBN can be viewed as a composition of bipartite undirected graphical models each of which is a restricted Boltzmann machine (RBM). RBM is an energy-based bipartite graphical model composed of two fully connected layers via symmetric undirected edges, but there are no connections between units of the same layer. The first layer consists of m visible units $v = (v_1, v_2, \dots, v_m)$ that represent observed data, while the second layer consists of n hidden units $h = (h_1, h_2, \dots, h_n)$ that can be viewed as nonlinear feature detectors to capture higher-order correlations in the observed data. In addition, $W = \{w_{11}, w_{21}, \dots, w_{nm}\}$ is the connecting weights matrix between the visible and hidden units. A typical RBM structure is shown in Fig. 1a. The standard RBM was designed to use only binary stochastic visible units, and is the so-called Bernoulli RBM (BRBM). However, using binary units is not suitable for real-valued data (e.g., pixel intensities values in images). Therefore, a new model has been developed called the Gaussian RBM (GRBM) to address this limitation of the standard RBM [40]. The energy function of the GRBM is defined as follows:

$$E(v, h) = - \sum_{i=1}^m \sum_{j=1}^n w_{ij} h_j \frac{v_i}{\sigma_i} - \sum_{i=1}^m \frac{(v_i - b_i)^2}{2\sigma_i^2} - \sum_{j=1}^n c_j h_j \tag{1}$$

Here, σ_i is the standard deviation of the Gaussian noise for the visible unit v_i , w_{ij} represents the weights for the visible unit v_i and the hidden unit h_j , and b_i and c_j are biases for the visible and hidden units, respectively. The conditional probabilities for the visible units given hidden units and vice versa are defined as follows:

$$p(v_i = v|h) = N\left(v|b_i + \sum_j w_{ij}h_j, \sigma_i^2\right) \tag{2}$$

$$p(h_j = 1|v) = f\left(c_j + \sum_i w_{ij} \frac{v_i}{\sigma_i^2}\right) \tag{3}$$

Here, $N(\cdot|\mu, \sigma^2)$ refers to the Gaussian probability density function with mean μ and standard deviation σ . During the training process, the log-likelihood of the training data is maximized using stochastic gradient descent and the update rules for the parameters are defined as follows:

$$\Delta w_{ij} = \epsilon (\langle v_i h_j \rangle_{data} - \langle v_i h_j \rangle_{model}) \tag{4}$$

Here, ϵ is the learning rate and $\langle v_i h_j \rangle_{data}$ and $\langle v_i h_j \rangle_{model}$ represent the expectations under the distribution specified by the input data and the internal representations of the RBM model, respectively. As reported in the literature, RBMs can be used in two different ways: either as generative models or as discriminative models, as shown in Fig. 1a, b.

Generally, DBNs can be efficiently trained using an unsupervised greedy layer-wised algorithm, in which the stacked RBMs are trained one at a time in a bottom to top manner. For instance, consider training a DBN composed of three hidden layers as shown in Fig. 1c. According to the greedy layer-wised training algorithm proposed by Hinton et al. [39], the first RBM is trained using the contrastive divergence (CD) algorithm to learn a layer (h_1) of feature representations from the visible units, as described in [39]. Then, the hidden layer units (h_1), of the first RBM, are used as visible units to train the second RBM. The whole DBN is trained when the learning of the final hidden layer is completed. A DBN with l layers can model the joint distribution between the observed data vector v and l hidden layers h_k as follows:

$$P(v, h^1, \dots, h^l) = \left(\prod_{k=0}^{l-2} P(h^k|h^{k+1})\right) P(h^{l-1}, h^l) \tag{5}$$

Here, $v = h^0$, $P(h^k|h^{k+1})$ is the conditional distribution for the visible units given hidden units of the RBM associated with level k of the DBN, and $P(h^{l-1}, h^l)$ is the visible-hidden joint distribution in the top-level RBM. An example of a three layers DBN as a generative model is shown in Fig. 1d, where the symbol Q is introduced for exact or approximate posteriors of that model which are used for bottom-up inference. During the bottom-up inference, the Q posteriors are all approximate except for the top level $P(h^l|h^{l-1})$, which is formed as an RBM and then exact inference is possible.

Like any deep learning approach, the DBN is usually applied directly on the pixel intensity representations. However, although DBN has been successfully applied in many

different fields, scaling it to realistic-sized face images still remains a challenging task for several reasons. Firstly, the high dimensionality of the face image leads to increased computational complexity of the training algorithm. Secondly, the feature representations of the DBN are sensitive to the local translations of the input image. This can lead to a disregard of the local features of the input image, which are known to be important for face recognition. To address these issues of the DBN, a novel framework based on merging the advantages of the local handcrafted image descriptors and the DBN is proposed.

3 The proposed framework

As depicted in Fig. 2, a novel face recognition framework named the multimodal deep face recognition (MDFR) framework is proposed to learn high-level facial feature representations by training a DBN on top of a local *Curvelet–Fractal* representation instead of the pixel intensity representation. First, the main stage of the proposed *Curvelet–Fractal* approach is described in detail. This is followed by describing how to learn additional and complementary representations by applying a DBN on top of existing local representations.

3.1 The proposed Curvelet–Fractal approach

The proposed face recognition algorithm starts by detecting the face region using a Viola–Jones face detector [41]. Detecting the face region in a complex background is not one of our contributions in this paper. Then a simple pre-processing algorithm using a sigmoid function is applied. The advantage of the sigmoid function is to reduce the effect of illumination changes by expanding and compressing the range of values of the dark and bright pixels in the face image, respectively. In other words, compressing the dynamic range of the light intensity levels and spreading the pixel values more uniformly. This operation has increased average recognition rate by 6%. After that, the proposed *Curvelet–Fractal* approach is applied to the enhanced face image. As indicated above the Fractal dimension has many important properties, such as its ability to reflect the roughness and fluctuations of a face image’s surface and represent the facial features under different environmental conditions (e.g., illumination changes). However, the Fractal estimation methods can be very time consuming, and the high dimensionality of the face image makes it less suited to meeting the real-time requirements. Therefore, the Fractal dimension approach is applied on the Curvelet’s output to produce an illumination insensitive representation of the face image that can meet the real-time systems demands. Hence, the Curvelet transform is used here as a powerful technique for edge and curve rep-

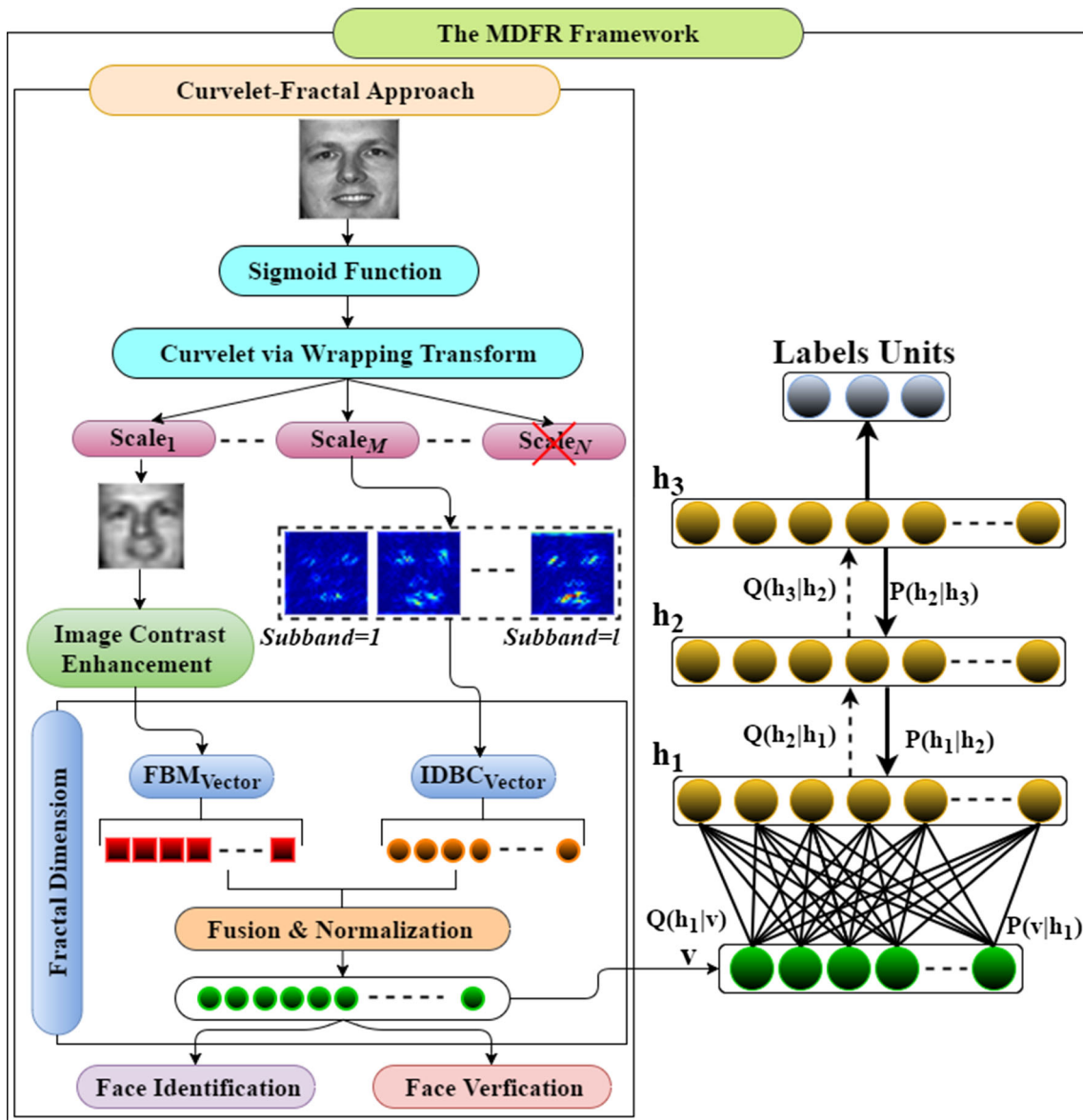


Fig. 2 Illustration of the proposed *Curvelet–Fractal* approach with the MDRF framework

resentation and dimensionality reduction of the face image, to increase the speed of Fractal dimension estimation.

In this work, two different methods to estimate the Fractal dimension are proposed based on the FBM and IDBC methods. The FBM method is used to process only the approximation coefficients (Coarse band) of the Curvelet transform, while the IDBC method is used to process the newly added Curvelet sub-bands and handle their high dimensionality. Then, the output of the FBM and IDBC are combined to build an elementary feature vector of the input image. After the Fractal dimension feature vector FD_{Vector} is obtained, a simple normalization procedure is applied to scale the obtained features to the common range $(0, 1)$, as follow:

$$\tilde{FD}_{Vector} = \frac{FD_{Vector} - \min(FD_{Vector})}{\max(FD_{Vector}) - \min(FD_{Vector})} \tag{6}$$

The main advantage of this scaling is to avoid features with greater numeric ranges dominating those with smaller numeric ranges, which can decrease the recognition accuracy. This procedure has increased average recognition rate by 5%. Finally, the quadratic discriminant classifier (QDC) and correlation coefficients (CC) classifiers are used in the recognition tasks. The main steps of the proposed *Curvelet–Fractal* approach for an input face image can be summarized as follows:

1. The sigmoid function is applied to enhance the face image illumination.

2. The Curvelet transform is applied to the image from 1, so the input image is decomposed into 4 scales and 8 orientations. In this work, the Curvelet sub-bands are divided into three sets, as explained in (Sect. 3.1.1).
3. The FBM method is applied to a contrast enhanced version of the coarse band produced in 2 and the result is then reshaped into a row feature vector FBM_{Vector} , as explained in (Sect. 3.1.2).
4. The IDBC method is applied to the middle frequency bands produced in 2 and a row feature vector $IDBC_{Vector}$ is constructed, as explained in (Sect. 3.1.3).
5. The final facial feature vector $FD_{Vector} = \{FBM_{Vector}, IDBC_{Vector}\}$ is constructed. To obtain a uniform feature vector, a normalization procedure is applied to obtain the normalized feature vector \tilde{FD}_{Vector} .
6. The QDC and CC classifiers are used in the final recognition tasks. The former is used for the identification task, while the latter is used for the verification task.

The next three subsections describe in more detail the Curvelet transform, FBM, and IDBC methods mentioned above.

3.1.1 Curvelet via wrapping transform

In this work, the wrapping based Curvelet transform described below is adopted, because it is faster to compute, more robust and less redundant than the alternative ridgelet- and USFFT-based forms of Curvelet transform. Its ability to reduce the dimensionality of the data and capture the most crucial information within face images, such as edges and curves plays a significant role in increasing the recognition power of the proposed system. The major steps implemented on a face image to obtain the Curvelet coefficients are clearly described in [20].

Based on domain knowledge from literature, suggesting that a higher scale decomposition would only increase the number of Curvelet sub-bands (coefficients) with very marginal or even no improvement in recognition accuracy, the Curvelet coefficients are generated at 4 scale and 8 orientation throughout this work. This maintains an acceptable balance between the speed and performance of the proposed system. Figure 3 shows the Curvelet decomposition coefficients of a face image of size (128×128) pixel taken from the FERET dataset. As indicated in Fig. 3, the output of the Curvelet transform can be divided into three sets:

1. The coarse band, containing only the low frequency (approximation) coefficients, is stored at the center of the display ($Scale_1$). These coefficients represent the main structure of the face.
2. The Cartesian concentric coronae that represents the middle frequency bands of the Curvelet coefficients at

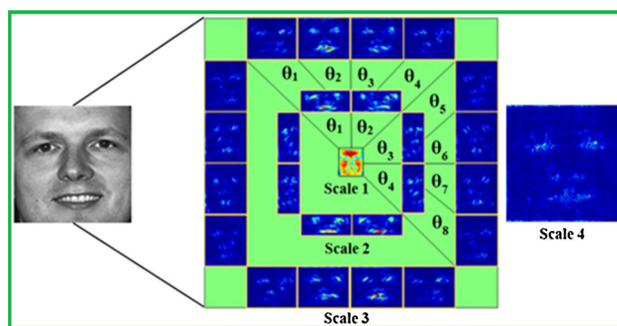


Fig. 3 Illustration of the Curvelet decomposition coefficients obtained from a face image decomposed at scale 4 and orientations 8

different scales, where the outer coronae correspond to the higher frequencies ($Scale_2, \dots, Scale_{N-1}$). Each corona is represented by four strips corresponding to the four cardinal points. These strips are further subdivided in angular panels, which represent the Curvelet coefficients at a specified scale and orientation. The coefficients in these bands represent the most significant information of the face, such as edges and curves.

3. The highest frequency band ($Scale_N$) of the face image, only indicated in Fig. 3, is at scale 4. This band has been discarded due to it being dominated by noise information.

From a practical point of view, the dimensionality of the Curvelet coefficients is extremely high due to the large amount of redundant and irrelevant information in each sub-band, especially in the middle frequency bands. Hence, working on such a large number of Curvelet coefficients is very expensive. A characteristic of the Curvelet transform is that it produces identical sub-bands coefficients at angle θ and $(\pi + \theta)$ for the same scale. Thus, only half of the Curvelet sub-bands need to be considered. In this work, instead of the direct use of the Curvelet coefficients, we analyze and process these coefficients using other methods. For the coarse band (the lowest frequency band), an image contrast enhancement procedure is applied as shown in Fig. 4, to improve the illumination uniformity of the face image stored at the center of the display by stretching the overall contrast of the image between two pre-defined lower and upper cutoffs, which are empirically set to be 0.11, and 0.999, respectively. This is followed by extracting the face texture roughness and fluctuations in the surface using the FBM method. For the middle frequency bands, the IDBC method is applied to reflect the face texture information and reduce the high dimensionality of these bands.

3.1.2 Fractional Brownian motion method

As shown in Fig. 5, the 2D face image can be considered as a 3D spatial surface that reflects the gray-level intensity value

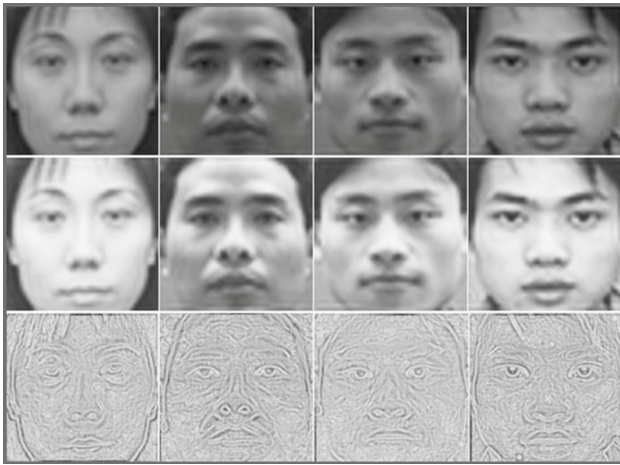


Fig. 4 The *top row* shows coarse band Curvelet approximation coefficients of four images. The *middle row* shows the images after applying the contrast enhancement procedure, and the *bottom row* shows the FBM fractal-transformed images

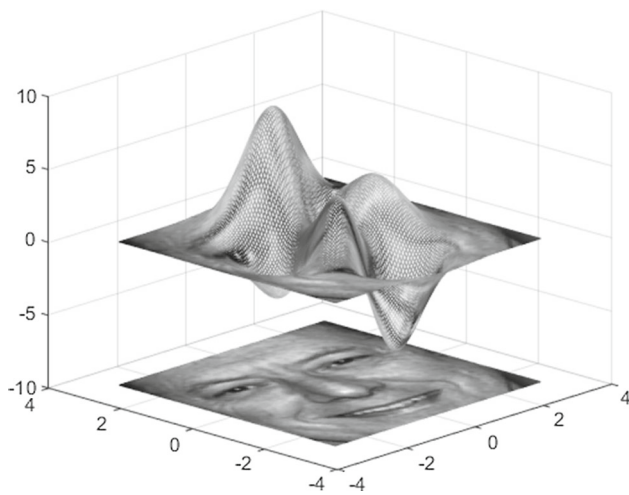


Fig. 5 The spatial surface corresponding to a grayscale face image

at each pixel position where the neighborhood region around each pixel cross the face surface, which covering a varying range of gray levels, can be processed as an FBM surface. The FBM is a nonstationary model and is widely used in medical imaging [33,42] due to its power to enhance the original image and make the statistical features more distinguishable.

For example, in [43] it was found that employing the normalized FBM to extract the feature vectors from surfaces of five ultrasonic liver images improved the classification of the normal and abnormal liver tissues. Moreover, the Fractal dimension for each pixel, calculated over the whole medical image by the normalized FBM method, could be used as a powerful edge enhancement and a detection method, which can enhance the edge representation for the medical images without increasing the noise level. According to Mandelbrot [32], the FBM is statistically self-affine, which means that

the Fractal dimension value of the FBM is not affected by linear transformations such as scaling. Therefore, the FBM is invariant under normally observed transformations of face images.

In this work, the face image of size $(M \times N)$ is transformed to its Fractal dimension form by applying a kernel function $fd(p,q)$ of size (7×7) on the entire face image, using the algorithm summarized in Fig. 6. More information on the mathematical functions of the FBM method can be found in [20]. Figure 4 shows examples of the approximation coefficients of the Curvelet transform and the resulting fractal-transformed images. After, a fractal-transformed image of size $(M \times N)$ has been obtained it is reordered into a row feature vector FBM_{Vector} for further analysis.

3.1.3 Improved differential box counting method

The main purpose of the second Fractal method is estimate the Fractal dimension features from the middle frequency bands of the Curvelet transform, reduce the high dimensionality of these bands, and increase the speed of the proposed system. Face recognition like other pattern recognition systems suffers from the so-called curse of high dimensionality. There are many possible reasons for reducing the feature vector size, such as providing a more efficient way for storing and processing the data related to the increasing number of training samples and increasing the discriminative power of the feature vectors.

The second method to compute the Fractal dimension is based on the improved differential box counting (IDBC) algorithm. The basic approach of the traditional DBC is to treat any image of size $(M \times M)$ as a 3D space where (x, y) denote the pixel position on the image surface, and the third coordinate (z) denotes the pixel intensity. The DBC starts by scaling the image down into nonoverlapping blocks of size $(s \times s)$, where $M/2 > s > 1$ and s is an integer, and then the Fractal dimension is calculated as follows:

$$FD = \lim_{r \rightarrow 0} \frac{\log(N_r)}{\log(1/r)} \quad (7)$$

where $r = s$ is the scale of each block and N_r is the number of boxes required to entirely cover the object in the image, which is counted in the DBC method as follows: On each block there is a column of boxes of size $(s \times s \times s')$, where $s' = s$ and each box assigned with a number $(1, 2, \dots)$ starting from the lowest gray-level value, as shown in Fig. 7. Let the minimum and the maximum gray level of the image in the (i, j) th block fall in box number k and l , respectively. The contribution of n_r in (i, j) th block is calculated as follows:

$$n_r(i, j) = l - k + 1 \quad (8)$$

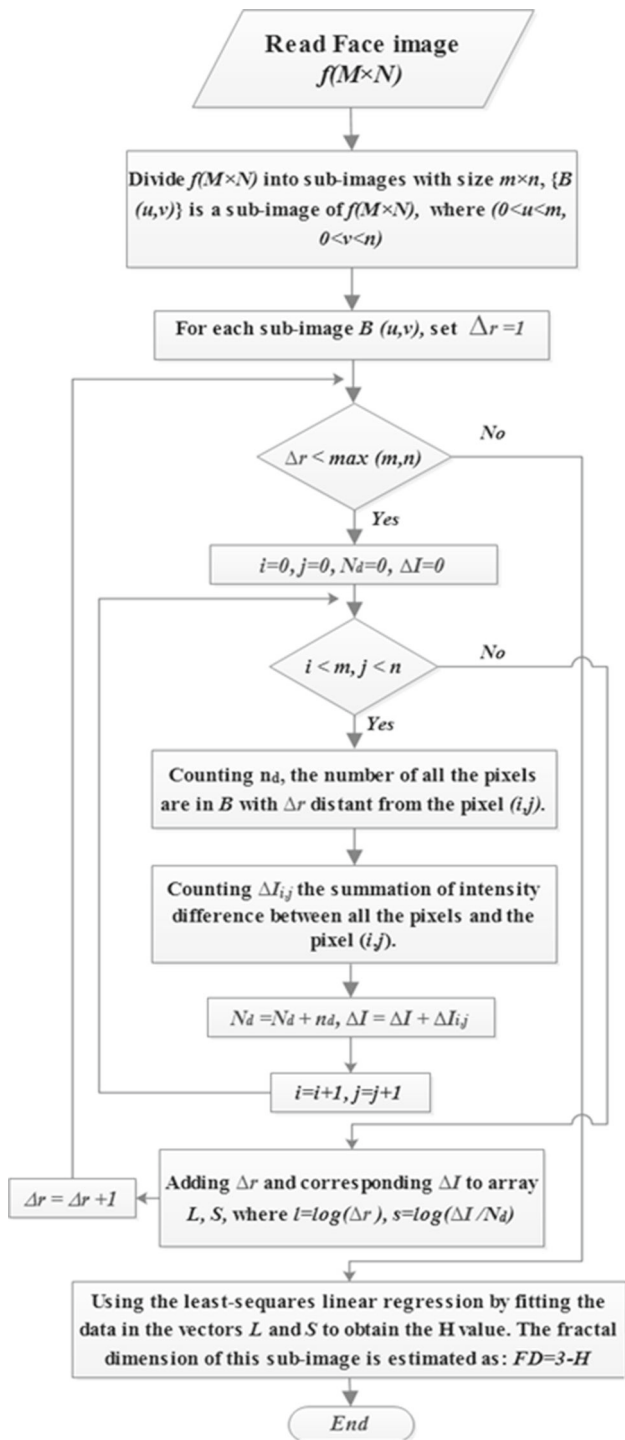


Fig. 6 A block diagram of the implementation of the FBM method

The contributions from all blocks N_r is counted for different values of r as follows:

$$N_r = \sum_{ij} n_r(i, j) \tag{9}$$

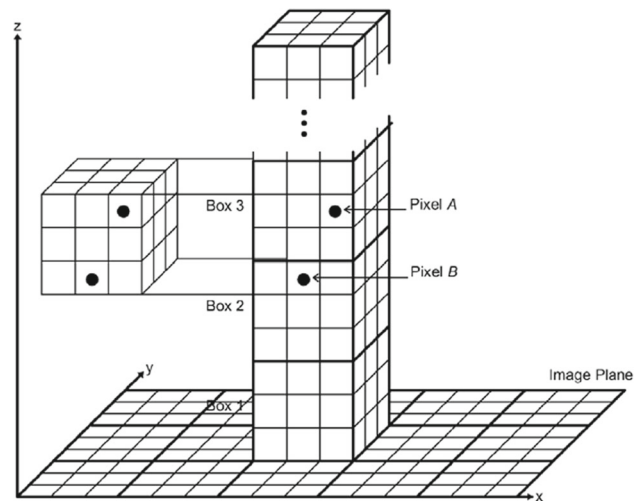


Fig. 7 Calculating the Fractal dimension by using the traditional (DBC)[46]

More information on this technique and its implementation can be found in [31]. The traditional DBC has many issues. The most important is how to choose the best size of the boxes that cover each block on the image surface. This can significantly affect the results of the curve fitting process, and result in inaccurate estimation of the Fractal dimension. Moreover, calculating the Fractal dimension using traditional DBC cannot accurately reflect the local and global facial features of different and similar classes. Finally, the traditional DBC method can suffer from over or under counting of the number of boxes that cover a specific block, which leads to calculating the Fractal dimension inaccurately [44,45]. The Fractal dimension feature is estimated from each block using $(\log_2 4)$ different sizes of boxes. Then, from each sub-image 16 Fractal dimension features are estimated. By combining the features obtained from the four sub-images (4×16) , we construct a sub-row feature vector $V_i = \{Fd_1, Fd_2, \dots, Fd_{64}\}$ for each Curvelet sub-band. As in Eq. (10), the final feature vector $IDBC_{vector}$ of the middle frequency bands is constructed by combining the V_i from 4 and 8 sub-bands located at scale 2 and 3, respectively.

$$IDBC_{vector} = \{V_1, V_2, \dots, V_{12}\} \tag{10}$$

In this work, to ensure the correct division without losing any important information the Curvelet sub-bands at scale 2 and 3 have been resized from their original sizes to (24×24) and (32×32) , respectively. The experimental results have demonstrated that calculating the Fractal dimension features using different sizes of boxes covering the same block can play a significant role in increasing the discriminative power of the final feature vector by efficiently reflecting the face texture information using the edges and curves of the face presented in the middle frequency bands.

3.1.4 Face matching

Classification and decision making are the final steps in the proposed *Curvelet–Fractal* approach. These refer to the process of either classifying the tested samples into N classes based on the identity of the training subjects or deciding whether two faces belong to the same subject or not. In this paper, the QDC and CC classifiers are used in the identification and verification tasks, respectively. The QDC from PRTools¹ is a supervised learning algorithm commonly used for multiclassification tasks. It's a *Bayes-Normal-2* classifier assuming Gaussian distributions, which aims to differentiate between two or more classes using a quadric surface. Using this Bayes rule, a separate covariance matrix is estimated for each class, yielding quadratic decision boundaries. This is done by estimating the covariance matrix \mathbf{C} for the scatter matrix \mathbf{S} as follows:

$$\mathbf{C} = (\mathbf{1} - \alpha - \beta) \mathbf{S} + \alpha \text{diag}(\mathbf{S}) + \frac{\beta}{n} \sum \text{diag}(\mathbf{S}) \quad (11)$$

Here, n refers to the dimensionality of the feature space, α and $\beta \in [0, 1]$ are regularization parameters. In this work, these parameters are determined empirically to be $\alpha = 0.1$ and $\beta = 0.2$, as explained in (Sect. 4.2.1). The decision making is based on calculating the similarity scores between the two face images using the CC classifier, which is defined as follows:

$$C(A, B) = \frac{\sum_m \sum_n (A_{mn} - \bar{A})(B_{mn} - \bar{B})}{\sqrt{(\sum_m \sum_n (A_{mn} - \bar{A})^2)(\sum_m \sum_n (B_{mn} - \bar{B})^2)}} \quad (12)$$

Here, m and n are the dimensions of the sample, and \bar{A} and \bar{B} are the mean values of the testing and training samples, respectively.

3.2 Learning additional features representations

In this study, we argue that applying the DBN on top of local features representations instead of the pixel intensity representations (raw data), as a way of guiding the learning process, can greatly improve the ability of the DBN to learn more discriminating features with a shorter training time required to obtain the final trained model. As shown in Fig. 2, the local facial features are first extracted using the proposed *Curvelet–Fractal* approach. Then, the extracted local features are assigned to the feature extraction units of the DBN to learn additional and complementary representations. In this work, the DBN architecture stacks 3 RBMs

(3 hidden layers). The first two RBMs are used as generative models, while the last one is used as a discriminative model associated with softmax units for the multiclass classification purpose. Finally, the hidden layers of the DBN are trained one at a time in a bottom-up manner, using a greedy layer-wise training algorithm.

In this work, the training methodology to train the DBN model can be divided into three stages: pre-training, supervised, and fine-tuning phases.

1. In the pre-training phase, the first two RBMs are trained in a purely unsupervised way, using a greedy training algorithm, in which each added hidden layer is trained as a RBM (e.g., using the CD algorithm). The activation outputs of a trained RBM can be viewed as feature representations extracted from its input data, which will be the input data (visible units) used to train the next RBM in the stack. The unsupervised pre-training phase is finished when the learning of the second hidden layer is completed. The main advantage of the greedy unsupervised pre-training procedure is the ability to train the DBN using a massive amount of unlabeled training data, which can improve the generalization ability and prevent overfitting. In addition, the degree of complexity is reduced and the speed of training is increased.
2. In the supervised phase, the last RBM is trained as a non-linear classifier using the training and validation set along with their associated labels to observe its performance in each epoch.
3. Finally, the fine-tuning phase is performed in a top-down manner using the back-propagation algorithm to fine-tune parameters (weights) of the whole network for optimal classification.

A difference compared with conventional neural networks is that the DBNs require a massive amount of training data to avoid overfitting during the learning process and achieve satisfactory predictions. Hence, data augmentation is the simplest and most common method, of achieving this, which artificially enlarges the training dataset using techniques such as: random crops, intensity variations, and horizontal flipping. In contrast to previous works that randomly sample a large number of face image patches [12, 47], we propose to uniformly sample a small number of face image patches. To prevent background information from artificially boosting the results of the proposed *Curvelet–Fractal*² approach and to speed up experiments when the DBN is directly applied on the pixel intensity representations, the face region is detected,

¹ <http://www.37steps.com/prhtml/prtools.html>.

² The data augmentation procedure is not implemented during the performance assessment of the proposed *Curvelet–Fractal* approach.

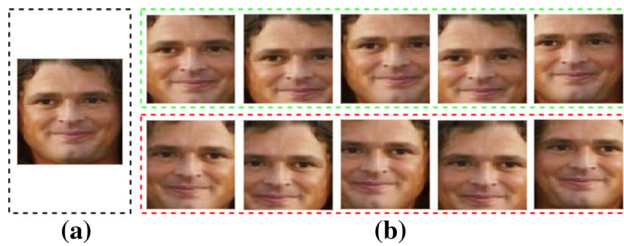


Fig. 8 Data augmentation procedure: **a** detected face image, **b** the normalized face patches used as input for the MDFR where the (*top*) row are patches sampled from (**a**), and the (*bottom*) row their horizontal flipped versions

and the data augmentation procedure³ is implemented on the detected face image. In this work, for a face image of size ($H_{\text{dim}} \times W_{\text{dim}}$), five images patches of the same size are cropped, four starting from the corner and one centered (and their horizontally flipped counterparts), which helps maximize the complementary information contained within the cropped patches. Figure 8 shows the ten image patches generated from a single input image.

4 Experimental results

In this section, comprehensive experiments are described using the proposed approaches for both face identification and verification tasks in order to demonstrate their effectiveness and compare their performance with other existing approaches. First a brief description of the face datasets used in these experiments is given. Then a detailed evaluation and comparison with the state-of-the-art approaches is presented in addition to some insights and findings about learning additional features representations by training a DBN on top of local feature representations.

4.1 Face datasets

In this work, all the experiments were conducted on four large-scale unconstrained face datasets: SDUMLA-HMT [48], Face REcognition Technology (FERET) [49], CAS-PEAL-R1 [50], and Labeled Faces in the Wild (LFW) [51]. Some examples of face images from each dataset are shown in Fig. 9.

- **SDUMLA-HMT dataset** [48] This includes 106 subjects and each has 84 face images taken from 7 viewing angles and under different experimental conditions including, facial expressions, accessories, poses, and illumination. The main purpose of this dataset is to simulate real-world

conditions during face image acquisition. The image size is (640×480) pixel.

- **FERET dataset** [49] This contains a total of 14,126 images taken from 1196 subjects, with at least 365 duplicate sets of images. This is one of the largest publicly available face datasets with a high degree of diversity of facial expression, gender, illumination conditions and age. The image size is (256×384) pixel.
- **CAS-PEAL-R1 dataset** [50] A subset of the CAS-PEAL face dataset has been released for research purposes and named CAS-PEAL-R1. This contains a total of 30,863 images taken from 1040 Chinese subjects (595 are males and 445 are females). The image size is (360×480) pixel.
- **LFW dataset** [51] This contains a total of 13,233 images taken from 5749 subjects where 1680 subjects appear in two or more images. In the LFW dataset, all images were collected from Yahoo! News articles on the Web, with a high degree of intra-personal variations in facial expression, illumination conditions, occlusion from wearing hats and glasses, etc. It has been used to address the problem of unconstrained face verification task in recent years. The image size is (250×250) pixel.

4.2 Face identification experiments

This section describes the evaluation of the proposed approaches to the face identification problem on three different face datasets: SDUMLA-HMT, FERET, and CAS-PEAL-R1. In this work, the SDUMLA-HMT dataset is used as the main dataset to fine-tune the hyper-parameters of the proposed *Curvelet–Fractal* approach (e.g., regularization parameters of the QDC classifier) as well as the proposed MDFR framework (e.g., number of hidden units per layer), because it has more images per person in its image gallery than the other databases. This allowed more flexibility in dividing the face images into training, validation and testing sets.

4.2.1 Parameter settings of the *Curvelet–Fractal* approach

In the proposed *Curvelet–Fractal* approach, the most important thing is to set the regularization parameters of the QDC classifier. In this work, these parameters are determined empirically by varying their values from 0 to 1 in steps of 0.1, starting with $\alpha = 0$ and $\beta = 0$. Hence, 121 experiments were conducted where each time we increase the former by 0.1 and test it with all the possible values of the latter. Figure 10 shows the validation accuracy rate (VAR) generated throughout these experiments. These experiments were carried out using 80% randomly selected samples for

³ In this work, the data augmentation procedure is applied only for the training and validation set.

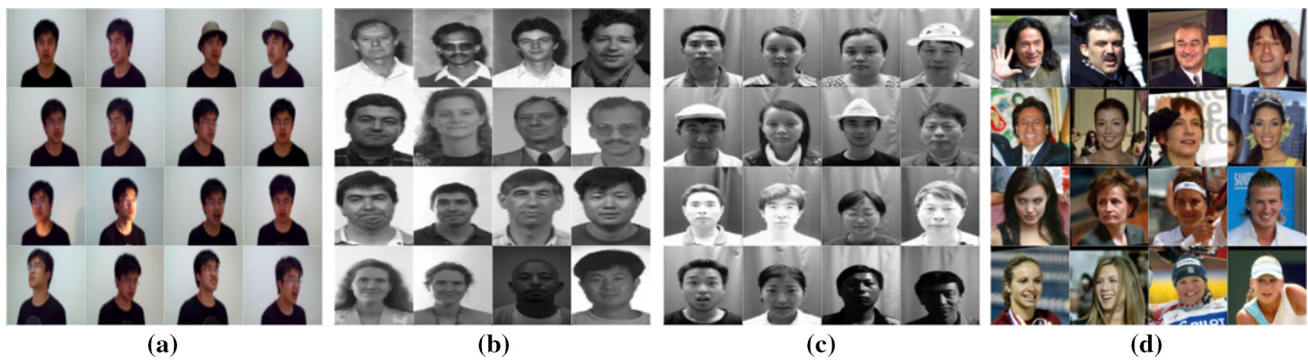


Fig. 9 Examples of face images in four face datasets: **a** SDUMLA-HMT, **b** FERET, **c** CAS-PEAL-R1, and **d** LFW

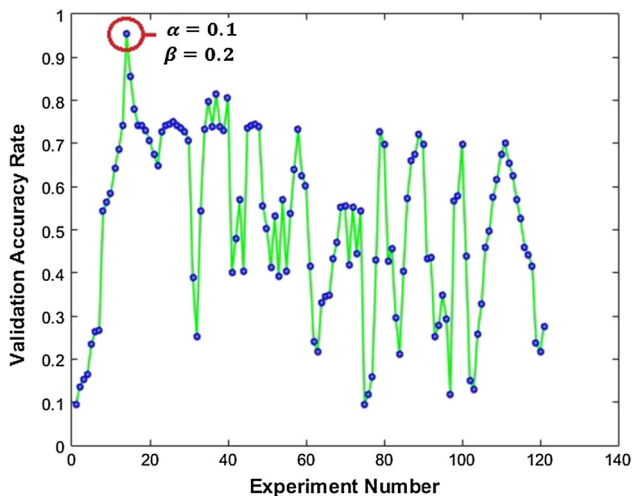


Fig. 10 The validation accuracy rate (VAR) generated throughout 121 experiments of finding the best regularization parameters

training set and the remaining 20% for testing set. In particular, the parameters optimization process is performed on the training set using the tenfold cross-validation procedure that divides the training set into k subsets of equal size. Sequentially, one subset is used to evaluate the performance of the classifier trained on the remaining $k - 1$ subsets. Then, average error rate (AER) over 10 trials is calculated as follows:

$$AER = \frac{1}{K} \sum_{i=1}^k Error_i \quad (13)$$

Here, $Error_i$ refers to the error rate per trial. After finding the best values of the regularization parameters, the QDC

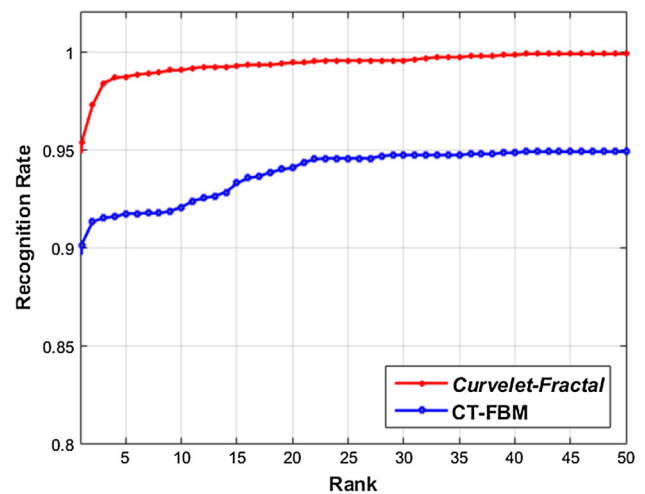


Fig. 11 Performance comparison between the *Curvelet-Fractal* and CT-FBM approaches on the SDUMLA-HMT Dataset

classifier is trained using the whole training set, and its performance in predicting unseen data properly is then evaluated using the testing set. Algorithm 1 shows pseudo-code of the procedure proposed to train the QDC classifier. Figure 11 shows results comparing the present *Curvelet-Fractal* approach with our previous Curvelet Transform-fractional Brownian motion (CT-FBM) approach described in [20] using the cumulative match characteristic (CMC) curve to visualize the performance of both approaches. It can be seen in Figure 11 that the Rank-1 identification rate has dramatically increased from 0.90 to 0.95 using the CT-FBM to more than 0.95 to 1.0 using the *Curvelet-Fractal* approach.

Algorithm 1. Find the optimal QDC classifier.

Input: Labeled dataset: **Training Set** (x_i, y_i) and **Testing Set** (x_i, y_i) .
Output: Optimal QDC Classifier.
Initialization: $\alpha = 0$, $\beta = 0$, and **Max_Accuracy**=0;
for $\alpha_{\text{now}} = 0 : 0.1 : 1$
 for $\beta_{\text{now}} = 0 : 0.1 : 1$
 ➤ Train the QDC classifier using the current values of α_{now} and β_{now} .
 ➤ Apply the 10-folds cross-validation evaluation procedure on the Training set.
 ➤ Calculate the Average Error Rate (AER) over 10-folds using Eq.15.
 ➤ Calculate the Validation_Accuracy = 1-AER;
 ➤ **if** Validation_Accuracy > Max_Accuracy
 Max_Accuracy = Validation_Accuracy;
 Optimal_ α = α_{now} ;
 Optimal_ β = β_{now} ;
 end
 end
end
 ➤ Train the QDC classifier using the **Optimal_ α** and **Optimal_ β** .
 ➤ Assess the final trained QDC classifier using the Testing set.

Table 1 Rank-1 identification rates obtained for different DBN architectures using Validation set

DBN models	Accuracy rate %
600–600–1000	92.19
700–700–1000	92.48
800–800–1000	95.38
900–900–1000	93.68

each one was trained separately as an RBM model in an unsupervised way using the CD learning algorithm with 1 step of Gibbs sampling (CD-1). Each individual model was trained for 300 epochs, momentum of 0.9, a weight decay of 0.0002, and mini-batch size of 100. The weights of each model were initialized with small random values sampled from a zero-mean normal distribution and standard deviation of 0.02. Initially, the learning rate was to be 0.001 for each model as in [52], but we observed this was inefficient as each model took too long to converge due to the learning rate being too small. Therefore, for all the remaining experiments, the learning rate set to be 0.01. The last RBM model was trained in a supervised way as a nonlinear classifier using the training and validation set along with their associated labels to evaluate its discriminative performance. In this phase, the same values of the hyper-parameters used to train the first two models were used, except that the last model was trained for 400 epochs. Finally, in the fine-tuning phase, the whole network was trained in top-down manner using the back-propagation algorithm equipped for dropout compensation to find optimized parameters and to avoid over-fitting. The dropout ratio is set to 0.5, and the number of epochs through the training set was determined using early stopping procedure, in which the training process is stopped as soon as the classification error on the validation set starts to rise again. In these experiments using the validation set, we found (see Table 1; Fig. 12) that hidden layers with sizes **800, 800, 1000** provided considerably better results than the other hidden layer sizes that we trained. This model trained on input data acquired from the *Curvelet–Fractal* approach is termed the MDFR framework. Table 1 shows the Rank-1 identification obtained from the four trained DBNs models over the validation set, while the CMC curves shown in Fig. 12 are used to visualize their performance on the validation set.

4.2.2 MDFR architecture and training details

The major challenge of using DNNs is the number of the model architectures and hyper-parameters that need to be evaluated, such as the number of layers, the number of units per layer, learning rate, the number of epochs. In a DBN, the value of a specific hyper-parameter may mainly depend on the values selected for other hyper-parameters. Moreover, the values of the hyper-parameters set in one hidden layer (RBM) may depend on the values of the hyper-parameters set in other hidden layers (RBMs). Therefore, hyper-parameter tuning in DBNs is very expensive. Given these findings, the best hyper-parameter values are found by performing a coarse search over all the possible values. In this section, all the experiments were carried out using 60% randomly selected samples for training and the remaining 40% samples were divided into two sets of equal size as validation and testing sets. In all experiments, the validation set is used to assess the generalization ability of the MDFR framework during the learning process before using the testing set. Following, the training methodology described in (Sect. 3.2), the MDFR framework was greedily trained using input data acquired from the *Curvelet–Fractal* approach. Once the training of a given hidden layer is accomplished, its weights matrix is frozen, and its activations are served as input to train the next layer in the stack.

As shown in Table 1, four different three-layer DBN models were greedily trained in a bottom-up manner using different numbers of hidden units. For the first two layers,

4.2.3 Comparative study of fractal, Curvelet–Fractal, DBN and MDFR approaches

In this section, to evaluate the feature representations obtained from the MDFR framework, its recognition accuracy was compared with feature representations obtained by

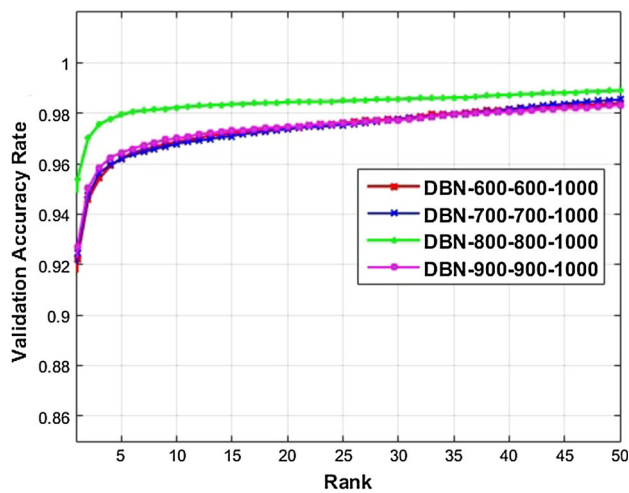


Fig. 12 CMC curves for the four trained DBNs models over the validation set

the Fractal, *Curvelet–Fractal* approach and DBN.⁴ This comparison study was conducted for several reasons: firstly, to demonstrate the efficiency of the proposed *Curvelet–Fractal* approach compared with applying the Fractal dimension individually. Secondly, to demonstrate that the feature representations acquired by the MDRF framework as a deep learning approach is complementary to the feature representations acquired by *Curvelet–Fractal* approach as a handcrafted-descriptors; thirdly, to show that applying the DBN on top of the local feature representations instead of the pixel intensity representations can significantly improve the ability of the DBN to learn more discriminating features with less training time required. Finally, using these complementary feature representations, the MDRF framework was able to efficiently handle the nonlinear variations of face images due to the nonlinearity of a DBN. In this work, the input image rescaled to (32×32) pixel to speed up the experiments when the Fractal dimension approaches are directly applied on the face image. Here, *Fractal_{Vector}* denotes applying both the FBM and IDBC approach directly on the input image.

As shown in Fig. 13, a higher identification rate was obtained using the proposed *Curvelet–Fractal* approach compared to only applying the Fractal dimension. Furthermore, we were able to further improve the recognition rate of the *Curvelet–Fractal* approach by learning additional feature representations through the MDRF framework as well as improve the performance of the DBN by forcing it to learn only the important facial features (e.g., edges and curves). To further examine the robustness of the proposed approaches, a number of experiments were conducted on FERET and CAS-PEAL-R1 dataset, and the results obtained are compared with

⁴ The DBN model was trained on the top of the pixel intensity representation using the same hyper-parameters of the MDRF framework.

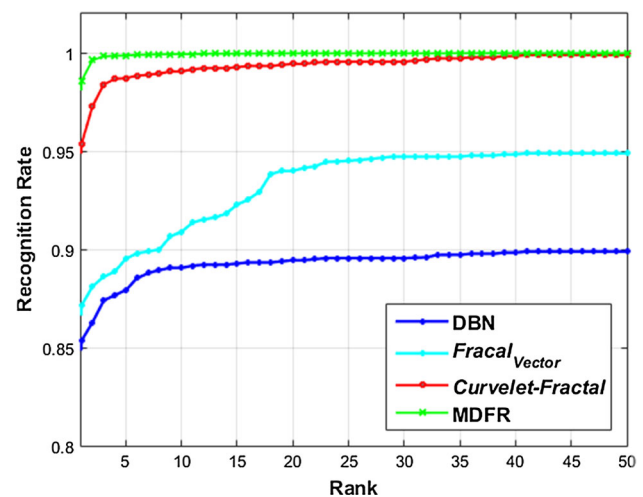


Fig. 13 Performance comparison between the DBN, *Curvelet–Fractal* and MDRF methods on the SDUMLA-HMT Dataset

the state-of-the-art approaches. For a fair comparison, the performance of the *Curvelet–Fractal* approach was evaluated using the standard evaluation protocols of FERET and CAS-PEAL-R1 dataset described in [49,50], respectively. In this work, to prevent overfitting and increase the generalization ability of the MDRF framework, the data augmentation procedure as described in (Sect. 3.2) was applied only to the gallery set of these two datasets. Then, its performance during the learning process was observed on a separate validation set taken from the full augmented gallery set.

According to the standard evaluation protocol, the FERET dataset is divided into five distinct sets: **Fa** contains a total 1196 subjects with one image per subject, which is used as a gallery set. The **Fb** contains 1195 images are taken on the same day and under the same lighting conditions as a **Fa** set, but with different facial expressions. The **Fc** set has 194 images taken on the same day as the **Fa** set, but under different lighting conditions. The **Dup I** set contains 722 images acquired on different days after the **Fa** set. Finally, the **Dup II** set contains 234 images acquired at least 1 year after the **Fa** set. Following the standard evaluation protocol, the last four sets are used as probe sets to address the most challenging problems in the face identification task, such as facial expression variation, illumination changes, and facial aging. Table 2 lists the Rank-1 identification rates of the proposed approaches and the state-of-the-art face recognition approaches on all four probe sets of the FERET dataset.

The standard CAS-PEAL-R1 evaluation protocol divides the dataset into a gallery set and six frontal probe sets without overlap between the gallery set and any of the probe sets. The gallery set consists of 1040 images of 1040 subjects taken under the normal conditions. The six probe sets contain face

Table 2 The Rank-1 identification rates of different methods on the FERET probe sets

Approach	Fb	Fc	Dup.I	Dup.II
DLBP-W [56]	99	99	86	85
G-LQP [57]	99.9	100	93.2	91.0
FHOGC [55]	98.3	98.3	86.3	81.2
Groupwise MRF [54]	98.5	98.8	87.7	86.2
H-Groupwise MRF [54]	99.7	99.2	94.7	93.6
LGOP+WPCA [58]	99.2	99.5	89.5	88.5
DFD(S = 3)+WPCA [53]	99.3	99	88.8	87.6
DFD(S = 5)+WPCA [53]	99.4	100	91.8	92.3
AMF [11]	99.9	100	96.4	93.6
GOM [10]	99.9	100	95.7	93.1
DBN	99.91	100	95.15	93.35
Fractal _{vector}	97.5	96.65	92	90.34
Curvelet–Fractal	100	98.97	97.92	95.72
MDFR framework	100	100	98.40	97.86

images with the following basic types of variations: Expression (PE) consists of 1570 images, accessories (PA) consists of 2285 images, lighting (PL) consists of 2243 images, time (PT) consists of 66 images, background (PB) consists of 553 images, and distance (PS) consists of 275 images. Table 3 lists the Rank-1 identification rates of the proposed approaches and the state-of-the-art face recognition approaches on all six probe sets of the CAS-PEAL-R1 dataset.

It can be seen from the results listed in Tables 2 and 3 that we were able to achieve competitive results with the state-of-the-art face identification results on FERET and CAS-PEAL-R1 dataset using only the *Curvelet–Fractal* approach. Its performance was compared with popular and recent feature descriptors, such as G-LQP, LBP, WPCA. Although that some approaches, such as DFD(S=5)+WPCA [53], GOM [10], AMF [11], and DBN approach achieved a slightly higher identification rate on the **Fc** probe set, they obtained inferior results on the other probe sets of the FERET dataset. In addition, the *Curvelet–Fractal* approach achieved a higher identification rate on all the probe sets of the CAS-PEAL-R1 dataset. Some of existing approaches, such as H-Groupwise MRF [54] and FHOGC [55] also achieved a 100% identification rate on the **PB** and **PT** probe set, respectively, but they obtained inferior results on the other probe sets of the CAS-PEAL-R1 dataset. Finally, a further improvements and a new state-of-the-art recognition accuracy was achieved using the MDFR framework on FERET and CAS-PEAL-R1 dataset, in particular, when the most challenging probe sets are under the consideration, such as **Dup I** and **Dup II** in FERET dataset and **PE**, **PA**, **PL**, and **PS** in the CAS-PEAL-R1 dataset.

4.3 Face verification experiments

In this section, the robustness and the effectiveness of the proposed approaches was examined to address the unconstrained face verification problem using LFW dataset. The face images in the LFW dataset were divided into two distinct Views. “View 1” is used for selecting and tuning the parameters of the recognition model, while “View 2” is used to report the final performance of the selected model. In “View 2”, the face images are paired into 6000 pairs, with 3000 pairs labeled as positive pairs and the rest as negative pairs. The final performance is reported as described in [51] by calculating the mean accuracy rate ($\hat{\mu}$) and the standard error of the mean accuracy (S_E) over tenfold cross-validation, with 300 positive and 300 negative image pairs per each fold. For a fair comparison between all face recognition algorithms, the creators of LFW dataset have pre-defined six evaluation protocols, as described in [63]. In this work, the “Image-Restricted, Label-Free Outside Data” protocol is followed where only the outside data are used to train MDFR framework.

Furthermore, the aligned LFW-a⁵ dataset is used and the face images were resized to (128 × 128) pixel after the face region has been detected using pre-trained Viola–Jones⁶ face detector. For the proposed *Curvelet–Fractal* approach, the feature representation of each test sample is obtained first, and then the similarity score between each pair of face images is calculated using the CC classifier. In the training phase, the *Curvelet–Fractal* approach does not use any data augmentation or outside data (e.g., creating additional positive/negative pairs from any other source), just use of the pre-trained Viola–Jones face detector, which has been trained using outside data. The final results over tenfolds are reported where each of the 10 experiments is completely independent of the others and the decision threshold of the CC classifier is learnt from the training set according to the standard evaluation protocol. Then, the accuracy rate in each round of tenfold cross-validation is calculated as the number of correctly classified pairs of samples divided by the total number of test sample pairs. For further evaluation, the results obtained from the *Curvelet–Fractal* approach were compared to state-of-the-art approaches on LFW dataset, such as DDML [64], LBP, Gabor [65], and MSBSIF-SIEDA [66] using the same evaluation protocol (Restricted), as shown in Table 4. It can be seen that the accuracy rate, 0.9622 ± 0.0272 , of the *Curvelet–Fractal* approach is higher than the best results reported on the LFW dataset, which is 0.9463 ± 0.0095 . In this work, further improvements and a new

⁵ <http://www.openu.ac.il/home/hassner/data/lfwa/>.

⁶ The incorrect face detection results have been detected manually to ensure that all the subjects are contributed in the subsequent evaluation of the proposed approaches.

Table 3 The Rank-1 identification rates of different methods on the CAS-PEAL_R1 probe sets

Approach	PE	PA	PL	PT	PB	PS
RBFNN [59]	84.8	93.4	63.4	96.9	–	–
DT-LBP [60]	98	92	41	–	–	–
DLBP-W [56]	99	92	41	–	–	–
1D-CFA [61]	83.12	74.84	31.43	71.21	98.19	98.55
Groupwise MRF [54]	94.8	90.3	66.9	99.2	98.8	99.5
H-Groupwise MRF [54]	96.4	90.3	66.9	99.8	100	99.6
LGOP+WPCA [58]	99.6	96.8	69.9	–	–	–
DFD($S = 3$)+WPCA [53]	99	96.9	63.9	–	–	–
DFD($S = 5$)+WPCA [53]	99.6	96.9	58.9	–	–	–
FHOGC [55]	94.9	90.3	68.7	100	–	–
LBP [62]	92.93	82.58	32.46	–	–	–
DBN	98.93	75.36	80.60	95.45	96.01	97.09
<i>Fractal</i> _{vector}	95.12	92.55	78.01	92.33	95.23	96.03
<i>Curvelet–Fractal</i>	99.87	98.07	89.48	100	100	99.64
MDFR Framework	100	99.43	89.92	100	100	100

Table 4 Performance comparison between the proposed approaches and the state-of-the-art approaches on LFW dataset under different evaluation protocols

Approach	Acc. ($\hat{\mu} \pm S_E$)	Protocol
DeepFace [67]	0.9735 \pm 0.0025	Unrestricted
DeepID [47]	0.9745 \pm 0.0026	Unrestricted
ConvNet-RBM[68]	0.9252 \pm 0.0038	Unrestricted
Convolutional DBN [18]	0.8777 \pm 0.0062	Restricted
DDML [64]	0.9068 \pm 0.0141	Restricted
VMRS [69]	0.9110 \pm 0.0059	Restricted
HPEN+HD-LBP+DDML [65]	0.9257 \pm 0.0036	Restricted
HPEN+HD-Gabor+DDML [65]	0.9280 \pm 0.0047	Restricted
Sub-SML+Hybrid+LFW3D [70]	0.9165 \pm 0.0104	Restricted
MSBSIF-SIEDA [66]	0.9463 \pm 0.0095	Restricted
DBN	0.9353 \pm 0.0165	Restricted
<i>Curvelet–Fractal</i>	0.9622 \pm 0.0272	Restricted
MDFR framework	0.9883 \pm 0.0121	Restricted

state-of-the-art result was achieved by applying the MDFR framework on LFW dataset. This experiment can be considered as an examination of the MDFR's generalization ability to address the unconstrained face verification problem on LFW dataset. In this work, the final performance of two pre-trained DBNs models was evaluated, while the first model was applied directly on top of pixel intensity representations the second was applied on top of local features representations and referred to the MDFR framework. Following the same evaluation protocol mentioned above, the hyper-parameters of the MDFR framework were fine-tuned using data from the SDUMLA-HMT dataset, as described in (Sect. 4.2.2).

In the MDFR framework, the feature representations \mathbf{f}_x and \mathbf{f}_y of a pair of two images \mathbf{I}_x and \mathbf{I}_y are obtained firstly by applying *Curvelet–Fractal* approach, and then a feature

vector \mathbf{F} for this pair is formed using element-wise multiplication ($\mathbf{F} = \mathbf{f}_x \odot \mathbf{f}_y$). Finally, these feature vectors \mathbf{F} (extracted from pairs of images) are used as input data to the DBN to learn additional features representations and perform face verification in the last layer. The performance of the MDFR framework is reported over tenfolds, each time one-fold was used for testing and the other ninefold for training. For each round of the 10 experiments, the data augmentation procedure was applied for the training set to avoid overfitting and increase the generalization ability of the network. Table 4 lists the mean accuracy of the recent state-of-the-art methods on the LFW dataset and the corresponding ROC curves are shown in Fig. 14. Considering the results of the MDFR framework, it is significantly improved over the mean accuracy rate of the *Curvelet–Fractal* approach and the DBN model applied directly on top of pixel intensity representations by

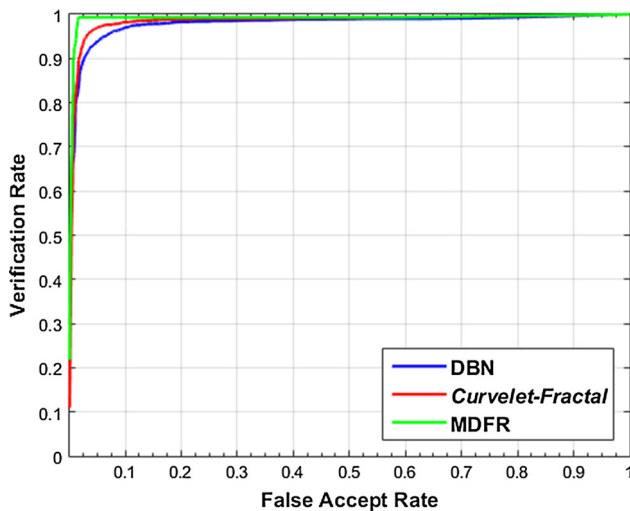


Fig. 14 ROC curves averaged over tenfolds of “View 2” of the LFW-a dataset. Performance comparison between the DBN, Curvelet–Fractal, and MDFR framework on the face verification task

2.6 and 5.3%, respectively. In this work, the performance of the proposed MDFR framework is also compared with several state-of-the-art deep learning approaches, including, DeepFace [67], DeepID [47], ConvNet-RBM [68], convolutional DBN [18] and DDML [64]. The first three approaches were mainly trained using “Unrestricted, Labeled Outside Data” protocol, in which a private dataset consisting of a large number of training images (> 100 K) is employed. The accuracy rate has been improved by 1.38%, compared to the next highest results reported by DeepID [47]. These promising results demonstrate the good generalization ability of the MDFR framework and its feasibility for deployment in real applications.

4.4 Running time

In this section, the running time of the proposed approaches, including the *Curvelet–Fractal*, DBN, and MDFR framework was measured by implementing them on a personal computer with the Windows 8 operating system, a 3.60 GHz Core i7-4790 CPU and 24 GB of RAM. The system code was written to run in MATLAB R2015a. It should be noted that the running time of the proposed approaches is proportional to the number of subjects and their images in the dataset. The training time using the different datasets is given in Table 5.

Table 5 The average training time of the proposed approaches using different datasets

Database	DBN	Curvelet–Fractal	MDFR framework
SDUMLA-HMT	18 h & 35 min	35 min	4 h & 15 min
FERET	16 h & 45 min	17 min	3 h & 33 min
CAS-PEAL-R1	15 h & 27 min	14 min	3 h & 32 min
LFW-a	13 h & 41 min	7 min	2 h & 56 min

It is clear from the table that the training time of the proposed MDFR framework has significantly reduced the training time of the DBN from when it is applied directly on top of the pixel intensity representations. Moreover, the computational efficiency of the proposed MDFR framework can be further improved using graphic processing units (GPUs) and code optimization. The test time per image from image input until the recognition decision for both the *Curvelet–Fractal* approach and MDFR framework is about 1.3 and 1.80 ms, respectively, which is fast enough to be used for real-time applications.

5 Conclusions and future work

In this paper, a novel multimodal local feature extraction approach is proposed based on merging the advantages of multidirectional and anisotropy transforms like the Curvelet transform with Fractal dimension. The main contribution of this approach is to apply the Curvelet transform as a fast and powerful technique for representing edges and curves of the face structure, and then to process the Curvelet coefficients in different frequency bands using two different Fractal dimension approaches to efficiently reflect the face texture under unconstrained environmental conditions. The proposed approach were tested on four large-scale unconstrained face datasets (e.g., SDUMLA-HMT, FERET, CAS-PEAL-R1 and LFW dataset) with high diversity in facial expressions, lighting conditions, noise, etc. The results obtained demonstrated the reliability and efficiency of the *Curvelet–Fractal* approach by achieving competitive results with the state-of-the-art approaches (e.g., G-LQP, LBP, WPCA), especially when there is only one image in the gallery set. Furthermore, a novel MDFR framework is proposed to learn additional and complementary information by applying the DBN on top of the local feature representations obtained from the *Curvelet–Fractal* approach. Extensive experiments were conducted and a new state-of-the-art accuracy rate is achieved by applying the proposed MDFR framework on all the employed datasets. Based on the results, it can be concluded that the proposed *Curvelet–Fractal* approach and MDFR framework can be readily used in real face recognition system for both identification and verification task with different face variations. On the basis of the

promising findings presented in this paper, work on the testing the proposed approaches on more challenging datasets is continuing and will be presented in future papers. In addition, further study of fusing the results obtained from the *Curvelet–Fractal* approach and MDFR framework would be of interest.

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

References

- Bhowmik, M.K., Bhattacharjee, D., Nasipuri, M., Basu, D.K., Kundu, M.: Quotient based multiresolution image fusion of thermal and visual images using daubechies wavelet transform for human face recognition. *Int. J. Comput. Sci.* **7**(3), 18–27 (2010)
- Parmar, D.N., Mehta, B.B.: Face recognition methods & applications. *Comput. Technol. Appl.* **4**(1), 84–86 (2013)
- Jafri, R., Arabnia, H.R.: A survey of face recognition techniques. *J. Inf. Process. Syst.* **5**(2), 41–68 (2009)
- Intiaz, H., Fattah, S.A.: A curvelet domain face recognition scheme based on local dominant feature extraction. *ISRN Signal Process.* **2012**, 1–13 (2012)
- Shreeja, R., Shalini, B.: Facial feature extraction using statistical quantities of curve coefficients. *Int. J. Eng. Sci. Technol.* **2**(10), 5929–5937 (2010)
- Zhang, B., Qiao, Y.: Face recognition based on gradient gabor feature and efficient Kernel Fisher analysis. *Neural Comput. Appl.* **19**(4), 617–623 (2010)
- Cao, Z., Yin, Q., Tang, X., Sun, J.: Face recognition with learning-based descriptor. In: Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pp. 2707–2714 (2010)
- Simonyan, K., Parkhi, O., Vedaldi, A., Zisserman, A.: Fisher vector faces in the wild. In: Proceedings of British Machine Vision Conference 2013, pp. 8.1–8.11 (2013)
- Al Ani, M.S., Al-Waisy, A.S.: Face recognition approach based on wavelet–curvelet technique. *Signal Image Process. Int. J.* **3**(2), 21–31 (2012)
- Chai, Z., Sun, Z., Mendez-Vazquez, H., He, R., Tan, T.: Gabor ordinal measures for face recognition. *IEEE Trans. Inf. Forensics Secur.* **9**(1), 14–26 (2014)
- Li, Z., Gong, D., Li, X., Tao, D.: Learning compact feature descriptor and adaptive matching framework for face recognition. *IEEE Trans. Image Process.* **24**(9), 2736–2745 (2015)
- Sun, Y., Wang, X., Tang, X.: Deep learning face representation by joint identification–verification. In: Advances in Neural Information Processing Systems, pp. 1988–1996 (2014)
- Lee, H., Grosse, R., Ranganath, R., Andrew, Y.N.: Unsupervised learning of hierarchical representations with convolutional deep belief networks. *Commun. ACM* **54**(10), 95–103 (2011)
- Liu, J., Fang, C., Wu, C.: A fusion face recognition approach based on 7-layer deep learning neural network. *J. Electr. Comput. Eng.* **2016**, 1–7 (2016). doi:10.1155/2016/8637260
- Fousek, P., Rennie, S., Dognin, P., Goel, V.: Direct product based deep belief networks for automatic speech recognition. In: Proceedings of IEEE International Conference on Acoustics Speech and Signal Processing ICASSP 2013, pp. 3148–3152 (2013)
- Lee, H., Pham, P., Largman, Y., Ng, A.: Unsupervised feature learning for audio classification using convolutional deep belief networks. In: Advances in Neural Information Processing Systems Conference, pp. 1096–1104 (2009)
- Sarikaya, R., Hinton, G.E., Deoras, A.: Application of deep belief networks for natural language understanding. *IEEE Trans. Audio, Speech Lang. Process.* **22**(4), 778–784 (2014)
- Huang, G.B., Lee, H., Learned-Miller, E.: Learning hierarchical representations for face verification with convolutional deep belief networks. In: Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pp. 2518–2525 (2012)
- Li, C., Wei, W., Wang, J., Tang, W., Zhao, S.: Face recognition based on deep belief network combined with center-symmetric local binary pattern. *Adv. Multimed. Ubiquitous Eng. Springer Singap.* **354**, 277–283 (2016)
- Al-Waisy, A.S., Qahwaji, R., Ipson, S., Al-Fahdawi, S.: A robust face recognition system based on curvelet and fractal dimension transforms. In: 2015 IEEE International Conference on Computer and Information Technology, Ubiquitous Computing and Communications, Dependable, Autonomic and Secure Computing, Pervasive Intelligence and Computing, pp. 548–555 (2015)
- Majumdar, A., Bhattacharya, A.: A comparative study in wavelets, curvelets and contourlets as feature sets for pattern recognition. *Int. Arab J. Inf. Technol.* **6**(1), 47–51 (2009)
- Candes, E., Demanet, L., Donoho, D., Loxton, Y.: Fast discrete curvelet transforms. *Multiscale Model. Simul.* **5**(3), 861–899 (2006)
- Arivazhagan, S., Ganesan, L., Subash Kumar, T.G.: Texture classification using curvelet statistical and co-occurrence features. In: 18th International Conference on Pattern Recognition (ICPR'06), pp. 938–941 (2006)
- Bhutada, G.G., Anand, R.S., Saxena, S.C.: Edge preserved image enhancement using adaptive fusion of images denoised by wavelet and curvelet transform. *Digit. Signal Process.* **21**(1), 118–130 (2011)
- Li, Y., Yang, Q., Jiao, R.: Image compression scheme based on curvelet transform and support vector machine. *Expert Syst. Appl.* **37**(4), 3063–3069 (2010)
- Chen, M.-S., Lin, S.-D.: Image fusion based on curvelet transform and fuzzy logic. In: 2012 5th International Congress on Image and Signal Processing (CISP), pp. 1063–1067. IEEE (2012)
- Ruihong, Y., Liwei, T., Ping, W., Jiajun, Y.: Image denoising based on Curvelet transform and continuous threshold. In: 2010 First International Conference on Pervasive Computing Signal Processing and Applications, pp. 13–16 (2010)
- Lee, Y.-C., Chen, C.-H.: Face recognition based on digital curvelet transform. In: 2008 Eighth International Conference on Hybrid Intelligent System Design and Engineering Application, pp. 341–345 (2008)
- Mandal, T., Wu, Q.M.J.: Face recognition using curvelet based PCA. In: 19th International Conference on Pattern Recognition, ICPR 2008, pp. 1–4 (2008)
- Xie, J.: Face recognition based on curvelet transform and LS-SVM. In: Proceedings of the International Symposium on Information Processing, pp. 140–143 (2009)
- Lopes, R., Betrouni, N.: Fractal and multifractal analysis: a review. *Med. Image Anal.* **13**(4), 634–49 (2009)
- Mandelbrot, B.: *The Fractal Geometry of Nature*, 3rd Edn. W. H. Free. San Fr. Library of Congress Cataloging In publication Data, United States of America (1983)
- Mandelbrot, B.: Self-affinity and fractal dimension. In: *PHYSICA SCRIPTA*, vol. 32, pp. 257–260 (1985)

34. Hsu, T., Hum K.-J.: Multi-resolution texture segmentation using fractal dimension. In: 2008 International Conference on Computer Science and Software Engineering, pp. 201–204 (2008)
35. Al-Kadi, O.S.: A multiresolution clinical decision support system based on fractal model design for classification of histological brain tumours. *Comput. Med. Imaging Graph.* **41**(2015), 67–79 (2015)
36. Zhu, Z., Gao, J., Yu, H.: Face detection based on fractal and complex model in the complex background. In: 2010 International Working Chaos–Fractals Theories and Applications, pp. 491–495 (2010)
37. Lin, K., Lam, K., Siu, W.: Locating the human eye using fractal dimensions. In: *Proceedings International Conference on Image Processing (Cat. No.01CH37205)*, pp. 1079–1082 (2001)
38. Farhan, M.H., George, L.E., Hussein, A.T.: Fingerprint identification using fractal geometry. *Int. J. Adv. Res. Comput. Sci. Softw. Eng.* **4**(1), 52–61 (2014)
39. Hinton, G.E., Osindero, S., Teh, Y.-W.: A fast learning algorithm for deep belief nets. *Neural Comput.* **18**(7), 1527–1554 (2006)
40. Hinton, G.: A practical guide to training restricted Boltzmann machines. *Computer. (Long. Beach. Calif)* **9**(3), 1 (2010)
41. Viola, P., Way, O.M., Jones, M.J.: Robust real-time face detection. *Int. J. Comput. Vis.* **57**(2), 137–154 (2004)
42. Chen, D.-R., Chang, R.-F., Chen, C.-J., Ho, M.-F., Kuo, S.-J., Chen, S.-T., Hung, S.-J., Moon, W.K.: Classification of breast ultrasound images using fractal feature. *Clin. Imaging* **29**(4), 235–45 (2005)
43. Chen, C., Daponte, J.S., Fox, M.D.: Fractal feature analysis and classification in medical imaging. *IEEE Trans. Med. Imaging* **8**(2), 133–142 (1989)
44. Liu, S.: An improved differential box-counting approach to compute fractal dimension of Gray-level image, In: 2008 International Symposium on Information Science and Engineering, pp. 303–306, (2008)
45. Long, M., Peng, F.: A box-counting method with adaptable box height for measuring the fractal feature of images. In: *Radioengineering*, pp. 208–213 (2013)
46. Li, J., Du, Q., Sun, C.: An improved box-counting method for image fractal dimension estimation. *Pattern Recognit.* **42**(11), 2460–2469 (2009)
47. Sun, Y., Wang, X., Tang, X.: Deep learning face representation from predicting 10,000 classes. In: *Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp. 1891–1898 (2014)
48. Yin, Y., Liu, L., Sun, X.: SDUMLA-HMT: a multimodal biometric database. In: *Chinese Conference on Biometric Recognition*, pp. 260–268. Springer-Verlag, Berlin, Heidelberg (2011)
49. Phillips, P.J., Moon, H., Rizvi, S.A., Rauss, P.J.: The FERET evaluation methodology for face-recognition algorithms. *IEEE Trans. Pattern Anal. Mach. Intell.* **22**(10), 1090–1104 (2000)
50. Gao, W., Cao, B., Shan, S., Zhou, D., Zhang, X., Zhao, D., Gao, W., Cao, B., Shan, S., Zhou, D., Zhang, X., Zhao, D.: The CAS-PEAL large-scale Chinese face database and baseline evaluations. *IEEE Trans. Syst. Man Cybern.* **38**(May), 1–24 (2004)
51. Huang, G.B., Mattar, M., Berg, T., Learned-miller, E.: Labeled faces in the wild?: A database for studying face recognition in unconstrained environments (2007)
52. Campilho, A., Kamel, M.: *Image Analysis and Recognition: 11th International Conference, ICIAR 2014 Vilamoura, Portugal, October 22–24, 2014 Proceedings, Part I. Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 8814 (2014)
53. Lei, Z., Pietikainen, M., Li, S.Z.: Learning discriminant face descriptor. *IEEE Trans. Pattern Anal. Mach. Intell.* **36**(2), 289–302 (2014)
54. Liao, S., Shen, D., Chung, A.C.S.: A Markov random field group-wise registration framework for face recognition. *IEEE Trans. Pattern Anal. Mach. Intell.* **36**(4), 657–669 (2014)
55. Tan, H., Ma, Z., Yang, B.: Face recognition based on the fusion of global and local HOG features of face images. *IET Comput. Vis.* **8**(3), 224–234 (2014)
56. Maturana, D., Mery, D., Soto, A.: Learning discriminative local binary patterns for face recognition. In: 2011 IEEE International Conference on Automatic Face & Gesture Recognition Work. FG 2011, pp. 470–475 (2011)
57. Hussain, S.U., Triggs, B., Hussain, S.U., Triggs, B., Recognition, V., Patterns, Q., Lazebnik, S., Perona, P., Sato, Y., Eccv, C. S.: Face recognition using local quantized patterns. In: *British Machine Vision Conference, Sep 2012, Guildford, United Kingdom* (2012)
58. Lei, Z., Yi, D., Li, S. Z.: Local gradient order pattern for face representation and recognition. In: *Proceedings of International Conference on Pattern Recognition*, pp. 387–392 (2014)
59. Farajzadeh, N., Faez, K., Pan, G.: Study on the performance of moments as invariant descriptors for practical face recognition systems. *IET Comput. Vis.* **4**(4), 272–285 (2010)
60. Maturana, D., Mery, D., Alvaro, S.: Face recognition with decision tree-based local binary patterns. In: *European Conference on Computer Vision*, pp. 469–481 (2004)
61. Yan, Y., Wang, H., Li, C., Yang, C., Zhong, B.: A novel unconstrained correlation filter and its application in face recognition. In: *International Conference on Intelligent Science and Intelligent Data Engineering*, pp. 32–39. Springer, Berlin, Heidelberg (2013)
62. Hu, J., Lu, J., Zhou, X., Tan, Y.P.: Discriminative transfer learning for single-sample face recognition. In: *Proceedings of 2015 International Conference on Biometrics, ICB 2015*, pp. 272–277 (2015)
63. Huang, G.B., Learned-miller, E.: Labeled Faces in the Wild?: Updates and New Reporting Procedures. University of Massachusetts, Amherst, UM-CS-2014-003, pp. 1–14 (2014)
64. Hu, J., Lu, J., Tan, Y.P.: Discriminative deep metric learning for face verification in the wild. In: *Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp. 1875–1882 (2014)
65. Zhu, X., Lei, Z., Yan, J., Yi, D., Li, S.Z.: High-fidelity pose and expression normalization for face recognition in the wild. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 787–796 (2015)
66. Ouamane, A., Bengherabi, M., Hadid, A., Cheriet, M.: Side-information based exponential discriminant analysis for face verification in the wild. In: *11th IEEE International Conference and Workshops on*, vol. 2, pp. 1–6 (2015)
67. Taigman, Y., Ranzato, M.A., Aviv, T., Park, M.: DeepFace: closing the gap to human-level performance in face verification. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1701–1708 (2014)
68. Sun, Y., Wang, X., Tang, X.: Hybrid deep learning for face verification. *IEEE Trans. Pattern Anal. Mach. Intell.* **38**(10), 1997–2009 (2016)
69. Barkan, O., Weill, J., Wolf, L., Aronowitz, H.: Fast high dimensional vector multiplication face recognition. In: *Proceedings of IEEE International Conference on Computer Vision*, pp. 1960–1967 (2013)
70. Hassner, T., Harel, S., Paz, E., Enbar, R.: Effective face frontalization in unconstrained images. In: *2015 IEEE Conference on Computer Vision and Pattern Recognition*, pp. 4295–4304 (2015)

Alaa S. Al-Waisy is a PhD student at Bradford University. He received his BSc in 2009 and MSc in 2011 in computer science from Al-Anbar University. He has worked as a lecturer at Al-Ma'aref University College (Iraq) for three years. Through this period, he has taught many subjects, such as the operating system, logic design and object oriented programming (OOP) and he has supervised a number of bachelor's projects at the computer science department in different areas such as image processing, pattern recognition, visual cryptography and texture classification. Al-Waisy served as the head of the laboratory unit at Al-Ma'aref University College. His research interests include pattern recognition, image processing, computer vision, medical imaging, designing and implementing unimodal and multimodal biometric systems.

Professor Rami Qahwaji is a Professor of Visual Computing at Bradford University (UK). He is originally trained as an Electrical Engineer and had MSc in Control and Computer Engineering both from University of Mustansiriyah (Baghdad). He received PhD in Computer Vision Systems in 2002. His research interests include space weather, 2D/3D image processing, machine learning and the design of machine vision systems with proven track record in the fields of solar/satellite imaging, medical imaging, data visualisation and applied data mining working with medical and industrial collaborators. He has raised over million pounds in funding from EPSRC, EU FP7, NHS, ERDF, European Space Agency, Yorkshire Forward, and more. He has over 120 refereed publications including 6 edited books, 10 book chapters, 40 journal papers, and over 65 invited talks and conference presentations and publications. So far, he has supervised 20 completed PhD projects, with another 6 currently in train. He is Fellow of the Institution of Engineering and Technology (FIET), Chartered Engineer (CEng), Fellow of the Higher Education Academy (FHEA) and a member of various professional organisations and has refereed research proposals for various national

and international funding bodies. He has acted as external examiner for research purposes for more than 15 UK and international universities and has a US patent.

Dr. Stanley Ipson is a Senior Lecturer in the Electronic Imaging and Media Communications Department at the University of Bradford. Originally trained as a physicist, with a first-class honours degree in Applied Physics and a PhD in Theoretical Nuclear Physics, his current research expertise includes image processing, pattern recognition and the design of machine vision systems. His imaging research interests have covered a variety of applications including video-rate deblurring of millimetre wave images, vision-based control of industrial machinery, face recognition, recognition of cursive Arabic handwriting, ASIC-based motion detection for surveillance using multiple cameras, photogrammetry applied to rural and urban landscapes and the detection of solar features. He has over 70 book and journal publications. His current research interests include solar imaging, space weather, machine learning and 3D modelling from 2D images.

Shumoos Al-Fahdawi is a PhD student at Bradford University. She earned her BSc in 2009 and MSc in 2011 in computer science from Al-Anbar University. She has worked as a lecturer at Al-Ma'aref University College (Iraq) for one year. Through this period, she has taught three subjects: Software engineering, Logic Design and Computational Theory, and she has supervised two of the bachelor's projects at the computer science department in the image processing field. Her research interests include software engineering, database systems and image processing. Her current research is medical image segmentation and classification.