https://doi.org/10.5755/j02.eie.35897 ELEKTRONIKA IR ELEKTROTECHNIKA, ISSN 1392-1215, VOL. 30, NO. 1, 2024

New Face Recognition System Based on DCT Pyramid and Backpropagation Neural Network

Badreddine Alane*, **Younes Terchi, Saad Bouguezel**
LCCNS, Department of Electronics, Ferhat Abbas University Setif-1,
Setif 19000, Algeria
**badreddine.alane@univ-setif.dz; terchi.younes@univ-setif.dz; sbouguezel@univ-setif.dz*

***Abstract*—Face recognition has emerged as a prominent biometric identification technique with applications ranging from security to human-computer interaction. This paper proposes a new face recognition system by appropriately combining techniques for improved accuracy. Specifically, it incorporates a discrete cosine transform (DCT) pyramid for feature extraction, statistical measures for dimensionality reduction of the features, and a two-layer backpropagation neural network for classification. The DCT pyramid is used to effectively capture both low- and high-frequency information from face images to improve the ability of the system to recognise faces accurately. Meanwhile, the introduction of statistical measures for dimensionality reduction helps in decreasing the computational complexity and provides better discrimination, leading to more efficient processing. Moreover, the two-layer neural network introduced, which plays a vital role in efficiently handling complex patterns, further enhances the recognition capabilities of the system. As a result of these advancements, the system achieves an outstanding 99 % recognition rate on the Olivetti Research Laboratory (ORL) data set, 98.88 % on YALE, and 99.16 % on AR. This performance demonstrates the robustness and potential of the proposed system for real-world applications in face recognition.**

***Index Terms*—Artificial neural networks; Discrete cosine transforms; Face recognition; Feature extraction; Statistical learning.**

I. INTRODUCTION

In recent years, biometrics has gained special attention as an important field with significant implications in multiple sectors [1], [2]. It involves analysing the unique physical or behavioural traits of individuals to achieve identification and authentication. Accurate identification of individuals plays a crucial role in improving security measures, optimising access control systems, and mitigating fraudulent activities. Among the various biometric modalities, facial recognition has emerged as an especially notable and promising approach due to its non-intrusive nature, real-time functionality, and adaptability to diverse environments [1]–[3]. Traditional facial recognition techniques involve two main steps, namely feature extraction and classification [1]–[3]. The former aims to capture the distinguishing characteristics of a face from an image, while the latter uses these extracted features to identify the individual [2], [3]. Various feature extraction methods have been employed in the literature, such as local binary patterns (LBP) [4], [5], eigenfaces [6], [7], Gabor filters [8], and histogram of oriented gradients (HOG) [9]. LBP is a texture descriptor that encodes local patterns by comparing the intensities of the central pixels with those of its neighbours [4]. It is effective for capturing texture variations in the face. HOG focusses on the distribution of local gradients in an image, which can represent shape information [9]. Eigenfaces employ principal component analysis (PCA) to extract the most discriminative features from a set of face images [6], [7]. Finally, Gabor filters are a set of filters that capture spatial frequency information and can represent facial textures [8]. Another approach to facial recognition feature extraction involves the use of established transforms, such as the discrete cosine transform (DCT) [10]–[15] and discrete wavelet transform (DWT) [12], [13], [16]–[19]. On the one hand, the DWT offers the advantage of multiresolution analysis, allowing capture of facial details at different scales [15]. In [12], the authors proposed a hybridisation of the fractional DCT and DWT to normalise light variations in face images. This combination significantly helps the process of extracting features from images. Another approach was introduced in [13], where two sets of extracted feature coefficients were combined using two different feature coefficient extraction techniques: DWT-based Gaussian and DCT-based Gabor filters. The desired class label for individual faces was generated by using a multiclass support vector machine classifier that uses the two combined features. A method combining K-nearest neighbour (KNN) and particle swarm optimisation was proposed in [20]. It starts with extracting features through the local binary pattern. The KNN classifier was improved using the population-based metaheuristic particle swarm optimisation (PSO) algorithm. For example, Setiawan, Sigit, and Rokhana [15] demonstrated the efficiency of employing DCT-based image compression to reduce the training time of their proposed CNN-based facial recognition system without significantly compromising its accuracy. Furthermore, the combination of DWT coherence with various methods, including PCA, linear discriminant analysis (LDA), and convolutional neural networks (CNNs), showed promising performance enhancements, resulting in better recognition rates [16]. In this context, Atta and Ghanbari made a notable contribution in [21] - introduced a novel methodology known as the DCT pyramid. This approach combines the advantageous features of both the DCT and the multiresolution strategy. Although initially developed for compact image coding [22], the DCT

Manuscript received 30 September, 2023; accepted 12 January, 2024.

pyramid demonstrated promising results in the field of facial recognition [21]. For the classification step, several algorithms are commonly used in traditional facial recognition, including nearest neighbour [23], support vector machines (SVM) [24], and decision trees [25]. Nearest neighbour algorithms classify an image based on its similarity to known faces in the training set [20], [23], while SVMs create a hyperplane to separate different classes, maximising the margin between them [24]. On the other hand, decision trees use a hierarchical structure to classify images based on different facial features [25]. However, traditional techniques are less robust when faced with occlusions such as sunglasses, hats, or facial hair, as these may interfere with the feature extraction process and affect classification accuracy [2], [3], [26]. Moreover, traditional methods might struggle to handle noisy or low-quality images, leading to reduced recognition performance in such cases [2], [3]. Deep learning has revolutionised the field of facial recognition by offering significant improvements in accuracy [2], [3]. Deep neural networks, such as CNNs, have shown remarkable performance in facial feature extraction and classification [14], [27], [28]. In fact, CNNs can automatically learn high-level representations of faces by using multiple layers of convolutional and pooling operations. The advantage of deep learning in facial recognition lies in its ability to learn complex features directly from raw data, eliminating the need for handcrafted feature extraction techniques [2], [3], [26]. Additionally, deep learning models can capture subtle variations in facial appearance and deal with variations in pose, lighting conditions, and facial expressions. However, the main disadvantage of deep learning approaches is their hunger for large amounts of labelled training data [2], [3], [26]. Training deep neural networks requires a substantial data set of labelled faces to achieve optimal performance. Collecting and annotating such a data set can be time consuming and expensive. To address this issue, a common approach is to combine traditional feature extraction techniques with deep learning. This coupling involves the use of traditional methods to extract features from facial images, which are then fed into a deep neural network for classification [3]. This hybrid approach benefits from the representational power of deep learning while using the robustness and interpretability of traditional feature extraction methods [3], [19], [24], [29]–[32]. Specifically, in [29], the authors proposed an improved approach for human face recognition using a backpropagation neural network (BPNN) and correlation-based feature extraction. The use of the local binary pattern histogram (LBPH) descriptor and the generation of a new data set (T-data set) are also considered therein. Although the paper claims higher accuracy and reduced computational cost due to the use of a reduced image features. However, employing five combined distance measurement algorithms to generate the T-data set might pose a challenge in terms of the reduced computational efficiency when conducting tests on extensive data sets. The authors in [19] focussed on a technique that combines probabilistic neural networks (PNNs) with an improved version of kernel linear discriminant analysis for facial recognition purposes (IKLDA). Setiawan, Sigit, and Rokhana explore in [15] the use of deep learning (CNN) and the DCT for face recognition. Although it discusses the impact of activation functions and epochs, nevertheless, employing CNNs could lead to intricate data structures and result in an increase of computational coast for classification tasks. The researchers introduced a framework in [31], which involves the use of smart glasses for facial recognition. They harnessed image processing techniques and applied CNNs, specifically employing transfer learning with the pretrained multilayer network AlexNet. Although this framework offers appealing benefits like portability and effective frontal view capture, it is worth noting that utilising the intricate structure of the AlexNet network would necessitate robust hardware specifications. In this paper, we propose a novel approach to enhance the accuracy and efficiency of face recognition systems. It combines the DCT pyramid decomposition technique with an artificial neural network (ANN) employing the widely used backpropagation algorithm. Thus, our objective is to efficiently exploit the advantages offered by each of these two techniques. We show that the resulting face recognition system is highly robust and capable of operating with minimal memory requirements. The DCT pyramid decomposition technique serves as the foundation of our approach. It decomposes the facial image into multiple frequency subbands for capturing both low- and high-frequency components. Using this decomposition, we extract a rich set of features that can significantly improve the discriminative power of the face recognition system. In addition, the pyramid structure allows us to efficiently represent the facial information on different scales, providing robustness against variations in pose, illumination, and facial expressions. However, the efficient representation of facial information provided by the decomposition of the DCT pyramid would result in large vector dimensions. To overcome this drawback, statistical measures such as mean, variance, and entropy are employed to achieve an effective reduction in dimensionality without compromising the essential information contained within the feature vectors. To further enhance recognition accuracy, we integrate an ANN into our system. Specifically, we employ a multilayer perceptron architecture with backpropagation, which is an effective training algorithm [2], [3], [19], [24]. The neural network is trained to learn the complex relationships between the extracted features and the corresponding identities. Through an iterative learning process, the network adapts its internal parameters to optimise recognition performance, improving the accuracy of face identification and verification tasks. By carefully designing the integration of the DCT pyramid decomposition and the artificial neural network, we ensure that the system efficiently utilises available resources without sacrificing accuracy. Thus, another advantage of our approach is its low memory requirement. This is particularly important in real-world applications, where memory constraints are crucial [2], [3], such as embedded systems or mobile devices.

II. BRIEF REVIEW OF DCT PYRAMID DECOMPOSITION

Multiresolution methods have gained widespread use in various domains, including feature extraction, compression, and coding. One approach to achieve multiresolution decomposition is through the discrete cosine transform (DCT) pyramid, which decomposes an input image into multiple levels of approximations and details. A one-level

DCT pyramid decomposition is achieved essentially by employing blockwise DCT decimation as follows. First, the input image is divided into blocks of size N×N. Subsequently, each block is subjected to transformation via the N×N forward DCT. The inverse DCT (IDCT), scaled by a factor of M/N, is then applied to the M×M low-frequency components of the transformed block. Here, M < N denotes the desired down sampling size. This process yields a lower resolution version of the original block, called the "corner", i.e., the original block of size N×N is decimated into a lower resolution corner of size M×M. In addition, the remaining high-frequency components, which represent the fine details within the original block, form a reversed L-shape structure. By gathering the corners in the same order as their corresponding original blocks, the approximation subband, which is a lower resolution version of the input image, is constructed. Conversely, the resulting reversed L-shapes constitute the details subband, which contains the high-frequency components of the input image. Consequently, the DCT pyramid generates a low-resolution version with reduced dimensionality by a factor of M/N, along with a details subband comprising all the resulting reversed L-shapes.

Figure 1 shows the practical implementation of a one-level DCT pyramid decomposition of an 80×80 facial image.

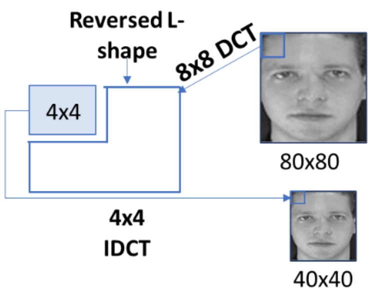

Fig. 1. Decomposition of a DCT pyramid of one level.

In this example (Fig. 1), by employing an 8×8 block size and a down sampling size $M=4$, the image is divided into 100 blocks. Initially, the forward DCT is applied to each block. Subsequently, the low-frequency 4×4 components within the transformed blocks undergo an inverse transformation using the 4×4-IDCT. The resulting values from the IDCT are then appropriately scaled by a factor of $M/N=1/2$, yielding the corresponding corners. At the same time, the remaining high-frequency components manifest themselves as reversed L-shapes. Finally, the corners are combined in the appropriate order to form the approximation subband with size 40×40, whereas the reversed L-shapes are gathered to constitute the details subband. When the DCT pyramid is iteratively applied to the approximation obtained from the previous level, multiple levels of decomposition are achieved. Each level captures finer details and provides a more detailed representation of the image.

III. PROPOSED FACE RECOGNITION SYSTEM

The proposed face recognition system involves several distinct steps: preprocessing, face detection, feature extraction, and classification. First, each face is cropped to its appropriate size using the Viola-Jones classifier [33] for accurate face detection. Subsequently, the detected face is resized to 80×80 and fed into the DCT pyramid algorithm to extract the feature vector. To reduce the dimensionality of the feature vector, we introduce statistical measures such as mean, variance, and entropy. The resulting reduced vector is then associated with the corresponding ID in the face gallery. For the classification stage, a neural network is trained on the feature vectors to distinguish between different subjects. In Fig. 2, a comprehensive block diagram visually illustrates the proposed approach. More detailed explanations concerning each step of this approach are provided in the upcoming subsections.

A. *Preprocessing*

In the preprocessing step, face images are cropped to the size of the face using Haar-cascade face detection [33], which is a face detection method that combines efficient computation with machine learning to achieve accurate results. It has been widely used in face recognition applications [1]. The algorithm uses Haar-like features, which are rectangular filters that are applied to the image at different scales and positions. These features capture patterns of intensity variations in the image, such as edges, corners, and texture. Then, to efficiently compute Haar-like features using simple and fast arithmetic operations, the algorithm uses integral images. An integral image is a representation of the original image, where the value at each pixel is the sum of the pixel intensities in a rectangular region from the top left corner of the image [33]. The algorithm also uses a machine learning technique called Adaboost to train a classifier [33]. Adaboost is an ensemble learning method that combines multiple weak classifiers to form a strong classifier [33]. During training, the algorithm selects the best Haar-like features and their thresholds that can accurately discriminate between faces and nonfaces in the training data. The trained classifier is organised into a cascade of classifiers, where each classifier is trained to be more complex and accurate than the previous one. The cascade is designed to quickly reject regions of the image that are unlikely to contain a face, thus reducing the computational cost of the algorithm. The algorithm applies the trained cascade to the input image by scanning it with a sliding window at different scales and positions. At each window, Haar-like features are computed using integral images, and the cascade of classifiers is applied to determine if the window contains a face or not. If a window is classified as a face by all classifiers in the cascade, it is considered a detection [33]. After the face detection step, each image is then resized to 80×80 for the DCT pyramid decomposition discussed earlier.

B. *Feature Extraction*

For feature extraction, we employ a two-level DCT pyramid, which is performed in two decomposition stages and applied directly to the 80×80 facial image obtained from the preceding preprocessing step. This decomposition is achieved by employing the block dimension $N=8$ and sampling size $M=4$.

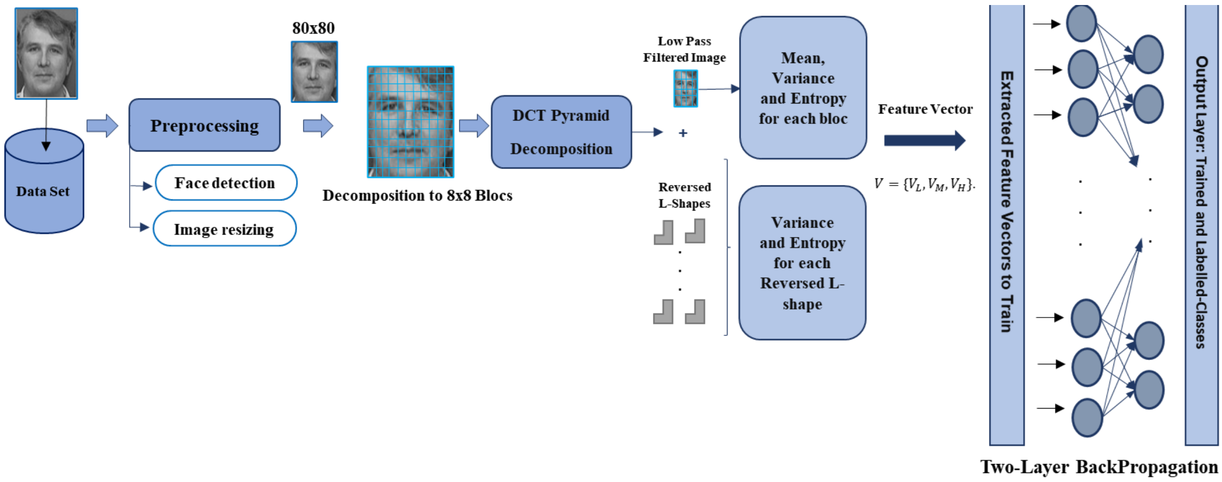

Fig. 2. Block diagram of the proposed face recognition system.

The first decomposition stage produces an approximation band comprising 100 $(M \times M)$-corners, along with their corresponding 100 reversed L-shapes. Subsequently, the second decomposition stage provides another set of 25 $(M \times M)$-corners and 25 reversed L-shapes. These 25 corners effectively describe the low-frequency subband of the facial image, whereas their corresponding 25 reversed L-shapes encapsulate the mid-frequency subband. Conversely, the reversed L-shapes obtained from the first decomposition stage encapsulate the high-frequency subband of the facial image. These resulting low-, mid- and high-frequency subbands span together the entire frequency spectrum of the input image, which provides a comprehensive understanding of its characteristics. However, using these subbands directly as an input of a classification algorithm would involve handling a substantial amount of data. To tackle this challenge, we propose statistical measures such as the mean, variance, and entropy of individual corners to reduce the dimensionality of the low-frequency subband. Accordingly, we compute the corresponding mean, variance, and entropy of the j^{th} corner $\mathbf{C}_j$ using the following:

$$\boldsymbol{\mu}_j = \frac{1}{M^2} \sum_{m=0}^{M-1} \sum_{n=0}^{M-1} \mathbf{C}_j(m,n), \tag{1}$$

$$\boldsymbol{\sigma}_j^2 = \frac{1}{M^2} \sum_{m=0}^{M-1} \sum_{n=0}^{M-1} (\mathbf{C}_j(m,n) - \boldsymbol{\mu}_j)^2, \tag{2}$$

and

$$\mathbf{H}_j = -\sum_{m=0}^{M-1} \sum_{n=0}^{M-1} \mathrm{p}(\mathbf{C}_j(m,n) \times \log_2(\mathrm{p}(\mathbf{C}_j(m,n))), \tag{3}$$

respectively, where $\mathrm{p}(\mathbf{C}_j(m,n))$ denotes the probability associated with each distinct value $\mathbf{C}_j(m,n)$ present within the corner $\mathbf{C}_j$. This process is applied to all corners of the low-frequency subband, resulting in the creation of the low-frequency features subvector $\mathbf{v}_L$ given by

$$\mathbf{v}_L = \bigcup_{j=0}^{L_2-1} \left\{ \boldsymbol{\mu}_j, \boldsymbol{\sigma}_j^2, \mathbf{H}_j \right\}, \tag{4}$$

where $L_2 = 25$ is the total number of corners at the second stage of the two-level DCT pyramid decomposition. To reduce the dimensionality of the mid- and high-frequency subbands, we attempt to use a strategy similar to that used for reducing the dimensionality of the low-frequency subband. However, it is seen that the mid- and high-frequency subbands are constituted of reversed L-shapes, which inherently represent the high-frequency content of the blocks, causing their mean values to converge towards zero. Consequently, we focus solely on using variance and entropy measures for the mid- and high-frequency and hence their corresponding feature subvectors $\mathbf{v}_M$ and $\mathbf{v}_H$ are, respectively, given by

$$\mathbf{v}_M = \bigcup_{j=0}^{L_2-1} \left\{ \hat{\boldsymbol{\sigma}}_{2j}^2, \hat{\mathbf{H}}_{2j} \right\}, \tag{5}$$

and

$$\mathbf{v}_H = \bigcup_{j=0}^{L_1-1} \left\{ \hat{\boldsymbol{\sigma}}_{ij}^2, \hat{\mathbf{H}}_{ij} \right\}, \tag{6}$$

where $\hat{\boldsymbol{\sigma}}_{ij}^2$ and $\hat{\mathbf{H}}_{ij}$ are the variance and entropy of the j^{th} reversed L-shape in the i^{th} decomposition stage. $L_1 = 100$ and $L_2 = 25$ indicate the number of blocks at the first and second decomposition stages. Finally, by concatenating the resulting three feature subvectors, we construct the feature vector representing the input facial image as $\mathbf{v} = \left\{ \mathbf{v}_L, \mathbf{v}_M, \mathbf{v}_H \right\}$. This vector assumes a size of $2L_1 + 5L_2 = 325$, which is of a notably reduced dimensionality compared to the size $80 \times 80 = 3200$ of the original input facial image.

C. Classification

We have implemented and tested different methods for the classification step. Ultimately, we decided to use the artificial neural network (ANN) with backpropagation, which has been proven to perform well, especially with more complex structures [2], [3]. Backpropagation is a popular supervised learning algorithm used in neural networks to train multilayer feedforward ANNs [29]. A backpropagation algorithm for a simple feedforward neural network as illustrated in Fig. 3 can be described as follows.

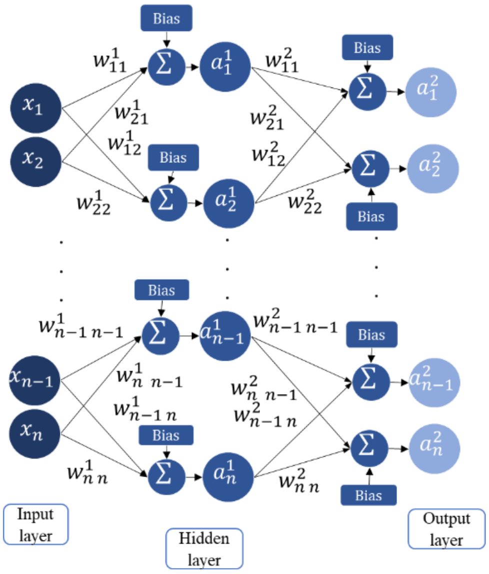

Fig. 3. Two-layer backpropagation neural network.

1. Initialise the weights $\mathbf{w}_j^{[l]}$ and biases $\mathbf{b}_j^{[l]}$ of the neural network randomly or with small values.
2. Feed the set of input features $\mathbf{x}_j, j=1,\dots,n,$ to the input layer of the neural network and compute the weighted sum of inputs and biases for each neuron in each layer. Pass the result through an activation function to obtain the output of each neuron in the hidden and output layers. This is called forward propagation as the input signals flow from the input layer to the output layer through the hidden layers. It is achieved by using

$$\mathbf{a}_j^{[l+1]} = f\left(\sum_{i=1}^{n} \mathbf{w}_{ji}^{[l+1]} \mathbf{a}_j^{[l]} + \mathbf{b}_j^{[l]}\right), \tag{7}$$

where $\mathbf{a}_j^{[0]} = \mathbf{x}_j$ and f is an activation function.

In our simulations, we experimented with various activation functions, including rectified linear unit (ReLu), sigmoid, softmax, and hyperbolic tangent sigmoid (Tansig). After thorough experimentation, we opted for the Tansig activation function for the hidden layer and the softmax activation function for the output layer. This combination delivered the best performance in our study.

3. Compare the output of the neural network with the ground truth (target) values using a loss function, which quantifies the difference between the predicted output and the actual output and serves as a measure of the performance of the neural network. The cross-entropy is used as a loss function in backpropagation neural networks. It measures the difference between predicted probabilities (output) and actual class labels (target) in a probabilistic manner and is used to guide the training process by updating the weights and biases of the neural network during backpropagation. The cross-entropy loss function is defined by

$$E = -\sum_{i} y_i \log(p_i), \tag{8}$$

where y_i is the ground truth for the i^{th} class in multiclass classification, and p_i is the probability of the i^{th} class predicted by the neural network.

4. Use an optimisation algorithm, such as gradient descent or one of its variants, to update the weights and biases of the neural network to minimise the loss based on the computed gradient. The weights and biases are adjusted in the opposite direction of the gradient, with a learning rate controlling the step size of the updates.
5. Iterate forward propagation, loss computation, backward propagation, and weight/bias updates for a number of epochs or until a convergence criterion is met. This allows the neural network to learn from the training data and gradually update its weights and biases to minimise loss and improve prediction accuracy.

IV. TEST RESULTS AND DISCUSSIONS

In this section, we consider the three well-known data sets, namely the Olivetti Research Laboratory data set (ORL), Yale, and AR to evaluate the performance of the proposed face recognition system that is based on the DCT pyramid coupled with ANN. In the following subsections, various experimental results are thoroughly analysed and compared in terms of accuracy and effectiveness with relevant and recent published methods [5], [19], [29], [30], [34]–[41].

A. Tests on ORL

The ORL is a comprehensive collection of facial images, covering ten distinct images for each of the 40 individual subjects present in the data set. To ensure diversity and variability, the images for certain subjects were captured at different times, resulting in variations in lighting conditions, facial expressions (such as opened or closed eyes, smiling or not smiling), and the presence or absence of glasses. Throughout the data set, a common dark and homogeneous background was used, and subjects were consistently positioned upright and frontal, though allowing some tolerance for slight side movements. To facilitate processing and analysis, all the images were uniformly resized to a resolution of 80×80, using a DCT pyramid decimation ratio of $1/2$. This standardised format ensures consistency and simplifies subsequent computations and feature extraction processes. The ORL data set has proven to be a valuable resource for research and development in facial recognition and related fields, due to its diverse and well-organised nature [5], [19], [29], [30], [34], [35], [37], [39], [40]. Several tests are conducted to assess the performance of the proposed system by varying decomposition ratios for training and testing the data set. The data set is considered for many case studies and divided into different proportions, such as 50 % for training and 50 % for testing, 70 % for training and 30 % for testing, and 90 % for training and 10 % for testing. Each labelled vector in the data set comprises 325 extracted features obtained from the application of DCT pyramid on the images of the subjects. These features are then utilised for the classification process within the ANN. The purpose of conducting these tests is to explore the optimal configuration for facial recognition tasks, seeking the most effective decomposition ratio and feature set to achieve high accuracy and reliability in subject classification. This rigorous evaluation process contributes to a deeper understanding of the strengths and limitations of the system, paving the way for further advances in facial recognition research and applications. Table I provides an overview of the perfect performance achieved by the proposed system in the ORL data set. To show the remarkable performance of the proposed system on the ORL data set, comprehensive comparisons are conducted with other state-of-the-art methods, specifically those reported in [5], [29], [30], [34], [35], [37], [39], [40]. Table II presents a detailed breakdown of the recognition rates attained by each of the mentioned methods with five images used for training in the proposed method and clearly shows that the proposed system achieves an impressive recognition rate of 100 %. Upon a deeper examination of this table, it becomes evident that the proposed system outperforms all the other methods, even when using a vector with a dimension of only 325 features. The results obtained underscore the superiority of the proposed system for facial recognition tasks, demonstrating its effectiveness in accurately identifying subjects within the data set. This perfect improvement in recognition rates over existing methods is of great significance as it highlights the power of coupling the DCT pyramid with ANN in providing a robust and reliable solution for facial recognition applications. Furthermore, the compact feature vector dimension of 325 further emphasises the efficiency and effectiveness of this method, making it a highly promising avenue for real-world implementation. As a result, these findings contribute significantly to the advancement of facial recognition technology and its potential deployment in various practical scenarios.

TABLE I. RECOGNITION RATES (%) ON ORL DATA SET USING DIFFERENT DECOMPOSITION RATIOS.

Corresponding Ratios	50:50	70:30	80:20	90:10
Proposed System	99	100	100	100

TABLE II. RECOGNITION RATES (%) ON ORL DATA SET USING DIFFERENT DECOMPOSITION RATIOS.

Method	**Training Samples**	**Recognition Rate (%)**
[35]	5	91
[40]	5	93.10
[37]	5	95.55
[5]	-	95.95
[30]	5	96.5
[34]	5	97.8
[29]	5	98
[41]	5	98.5
Proposed	5	99

B. *Tests on Yale*

In this section, we carry out tests using the Yale face data set, which comprises 165 greyscale images of 15 distinct individuals [36]. For each subject, there are 11 images, each representing a different facial expression or configuration, including centre light, glasses on, happiness, left light, without glasses, neutral expression, right light, sadness, sleepiness, surprise, and winking [36]. The wide range of expressions and configurations in the data set enables an evaluation of the methods strength and success in various situations. In the case of the Yale data set, our tests also encompass various ratios of training and testing sets, specifically 50 %50 %, 60 %40 %, 70 %30 %, and 90 %10 %. Table III provides an overview of the perfect performance achieved by the proposed system in the Yale data set. For a further evaluation of the system in this data set, we consider a comparison with other recently published methods, including the local binary pattern histogram (LBPH) descriptor with multiKNN and backpropagation neural network [29], LBPH for feature extraction with CNN for classification [36], and relative gradient magnitude strength (RGMS) with deep neural networks (DNNs) [30]. The results of this comparison are given in Table IV which clearly shows that the proposed system is outstanding, five images were used for training and six for testing. This is achieved with the reduction of the dimensionality of the feature vector, which is extracted by decomposition of the DCT pyramid with statistical features [21]. These compelling findings reinforce the effectiveness and potential of the DCT pyramid coupled with ANN in facial recognition tasks, particularly when dealing with images from the Yale data set. The use of statistical features in the DCT pyramid decomposition contributes to a more robust and accurate representation of facial characteristics, leading to improved recognition performance and advancing the state-of-the-art in facial recognition research

TABLE III. RECOGNITION RATES (%) ON YALE DATA SET USING DIFFERENT DECOMPOSITION RATIOS.

Corresponding Ratios	50:50	70:30	90:10
Proposed System	98.88	100	100

TABLE IV. RECOGNITION RATES OBTAINED BY DIFFERENT METHODS USING YALE DATA SET.

Method	**Training Samples**	**Recognition Rate (%)**
[40]	8	90.07
[37]	5	92
[5]	-	94.09
[41]	5	94.44
[29]	5	97.7
[36]	11	98.6
[30]	5	98.67
Proposed	5	98.88

C. *Tests on AR*

In this section, we examine the facial data set introduced by Aleix Martinez and Robert Benavente. This data set encompasses an extensive compilation of more than 4000 colour images that show the faces of 126 individuals, consisting of 70 men and 56 women. The images offer a comprehensive showcase of front-facing visages, capturing a wide range of facial expressions, various lighting scenarios, and obstructions such as sunglasses and scarves. The meticulous data collection process was conducted under tightly controlled conditions, ensuring the acquisition of consistent and top quality data. Importantly, participants did not encounter restrictions on their clothing, eyewear, cosmetics, or hairstyles during the image capture phase. For our tests on the AR data set, we adopt two distinct ratios for the training and testing sets, namely 70 %30 % and 90 %10 %. This allows us to conduct a thorough evaluation of the proposed system in varying proportions of training and testing data. The results of these tests are presented in Table V and clearly demonstrate the high performance and effectiveness of our approach for the facial recognition task in the AR data set. These results also provide valuable information on the robustness and accuracy of the method under various facial conditions and occlusions.

TABLE V. RECOGNITION RATES (%) ON AR DATA SET USING DIFFERENT DECOMPOSITION RATIOS.

Corresponding Ratios	70:30	90:10
Proposed System	99.16	100

Additionally, we conduct a separate comparison of the proposed system with [30], using the training and testing ratio of 70%30%. This comparison is provided in Table VI, which clearly shows that the proposed system significantly outperforms the proposed method in [30] on the AR data set. These comparative findings underscore the ability of the proposed system to handle the complexities present in the AR data set, including variations in facial expressions, illumination conditions, and occlusions. The method’s adeptness in extracting relevant features from facial images, combined with its sophisticated neural network architecture, significantly contributes to its outstanding performance.

TABLE VI. RECOGNITION RATES (%) OBTAINED BY DIFFERENT METHODS USING AR DATA SET.

Method	Training Samples	Recognition Rate (%)
[30]	8	92.3
Proposed	7	99.16

D. Comparison Using Mean Recognition Rate

To thoroughly evaluate the effectiveness of our approach, in this section, we conduct a comparative study using the mean recognition rate metric considering the ORL and Yale data sets. For this purpose, we compare the proposed approach with some existing prominent face recognition approaches, specifically, improved kernel linear discriminant analysis (IKLDA) + probabilistic neural networks (PNN) [19], the one based on strings of successive values [38], and that uses a neural network based on a radial basis function [39]. Therefore, we perform various network training iterations in data sets for different values of the training/testing ratio to assess the impact of this ratio on recognition performance. The mean recognition rates obtained by the methods are presented in Table VII for the ORL data set. Table VIII provides a comparison between the proposed method and methods proposed in [19] and [38] on the Yale data set, while Table IX demonstrates a comparison between the proposed method and the method proposed in [39] using different training and testing samples.

TABLE VII. MEAN RECOGNITION RATES (%) OBTAINED BY DIFFERENT METHODS USING ORL DATA SET.

Method	50:50	70:30	80:20
[19]	93.95 ± 0.02	96.35 ± 0.020	97.22 ± 0.02
[38]	95.65 ± 0.6	98.37 ± 0.47	98.75 ± 0.4
[39]	97.75 ± 2.31	-	-
Proposed	97.78 ± 1.29	99.68 ± 0.72	99.78 ± 0.50

TABLE VIII. MEAN RECOGNITION RATES (%) OBTAINED BY THE PROPOSED METHOD AND METHODS PROPOSED IN [38] AND [19] USING YALE DATA SET.

Method	Training Samples	Testing Samples	Mean
[19]	7	4	82.95 ± 0.07
[38]	7	4	96.88 ± 1.55
Proposed	7	4	99.91 ± 0.54

TABLE IX. MEAN RECOGNITION RATES (%) OBTAINED BY THE PROPOSED METHOD AND METHOD PROPOSED IN [39] USING YALE DATA SET.

Method	Training Samples	Testing Samples	Mean
[39]	6	4	99.83 ± 0.53
Proposed	5	6	98.01 ± 1.29
Proposed	6	4	99.58 ± 0.86

Furthermore, a comparison of mean recognition rates is conducted in Table X between the proposed method and methods proposed in [39] and [19] using the AR data set.

TABLE X. MEAN RECOGNITION RATES (%) OBTAINED BY THE PROPOSED METHOD AND METHODS PROPOSED IN [19] AND [39] USING AR DATA SET.

Method	70:30	80:20
[39]	93.15 ± 3.25	-
[19]	98.85 ± 1.08	99.16 ± 0.8
Proposed	99.2 ± 0.8	99.57 ± 0.645

It is seen from these tables that the proposed approach outperforms the existing ones in most cases. Therefore, the results of this comparative analysis underscore the superiority of the proposed system in capturing discriminative features and achieving higher recognition rates.

V. CONCLUSIONS

In this paper, we propose a novel face recognition system by efficiently combining the DCT pyramid and ANN, as well as by advantageously using statistical measures. The DCT pyramid decomposition was used for feature extraction. The process involves obtaining a feature vector that is subsequently reduced using statistical measures such as mean, variance, and entropy. This reduced feature vector serves as an input to a backpropagation neural network, which is trained to perform classification tasks effectively. Throughout the tests carried out on different known benchmark data sets, we have explored various factors and conducted a comprehensive comparison with different state-of-the-art methods. The results have shown promising performance for the proposed system, indicating its ability to achieve impressive recognition rates. In particular, the proposed system has a simple structure and requires minimal memory resources, making it an efficient and practical solution for real-world facial recognition applications. The combination of DCT pyramid decomposition and statistical feature reduction allows the system to capture essential facial characteristics while maintaining computational efficiency. The neural network’s training and classification tasks further contribute to the system’s robustness and adaptability in handling diverse facial conditions and variations. The results of this study would contribute to the advancement of facial recognition research, offering valuable insights and establishing the potential of the proposed approach as a reliable and efficient means of facial recognition. In addition, the remarkable performance of the system further validates its applicability and encourages its adoption in practical scenarios, where reliable and high-performance facial recognition systems are essential.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

References

[1] D. Zeng, R. Veldhuis, and L. Spreeuwers, "A survey of face recognition techniques under occlusion", *IET Biometrics*, vol. 10, no. 6, pp. 581–606, 2021. DOI: 10.1049/bme2.12029.

[2] A. Parashar, A. Parashar, W. Ding, M. Shabaz, and I. Rida, "Data preprocessing and feature selection techniques in gait recognition: A comparative study of machine learning and deep learning approaches", *Pattern Recognition Letters*, vol. 172, pp. 65–73, 2023. DOI: 10.1016/j.patrec.2023.05.021.

[3] M. Adel, S. H. El-Ramly, and B. Abdelhamid, "Adaptive rate polar code for DCT-based digital image transmission", *IEEE Access*, vol. 11, pp. 56196–56213, 2023. DOI: 10.1109/ACCESS.2023.3281197.

[4] T. Ahonen, A. Hadid, and M. Pietikainen, "Face description with local binary patterns: Application to face recognition", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 28, no. 12, pp. 2037–2041, 2006. DOI: 10.1109/TPAMI.2006.244.

[5] S. N. Khanbebin and V. Mehrdad, "Local improvement approach and linear discriminant analysis-based local binary pattern for face recognition", *Neural Computing and Applications*, vol. 33, pp. 7691–7707, 2021. DOI: 10.1007/s00521-020-05512-3.

[6] B. K. Gunturk, A. U. Batur, Y. Altunbasak, M. H. Hayes, R. M. Mersereau, "Eigenface-domain super-resolution for face recognition", *IEEE Trans. on Image Process.*, vol. 12, no. 5, pp. 597–606, 2003. DOI: 10.1109/TIP.2003.811513.

[7] J. Zhang, Y. Yan, and M. Lades, "Face recognition: Eigenface, elastic matching, and neural nets", *Proceedings of the IEEE*, vol. 85, no. 9, pp. 1423–1435, 1997. DOI: 10.1109/5.628712.

[8] J. Wu, G. An, and Q. Ruan, "Independent Gabor analysis of discriminant features fusion for face recognition", *IEEE Signal Processing Letters*, vol. 16, no. 2, pp. 97–100, 2009. DOI: 10.1109/LSP.2008.2010070.

[9] C. Shu, X. Ding, and C. Fang, "Histogram of the oriented gradient for face recognition", *Tsinghua Science & Technology*, vol. 16, no. 2, pp. 216–224, 2011. DOI: 10.1016/S1007-0214(11)70032-3.

[10] S. Naveen and R. S. Moni, "A robust novel method for face recognition from 2D depth images using DWT and DCT Fusion", *Procedia Computer Science*, vol. 46, pp. 1518–1528, 2015. DOI: 10.1016/j.procs.2015.02.072.

[11] B. Nassih, A. Amine, M. Ngadi, and N. Hmina, "DCT and HOG feature sets combined with BPNN for efficient face classification", *Procedia Computer Science*, vol. 148, pp. 116–125, 2019. DOI: 10.1016/j.procs.2019.01.015.

[12] V. P. Vishwakarma and S. Dalal, "Generalized DCT and DWT hybridization based robust feature extraction for face recognition", *Journal of Information and Optimization Sciences*, vol. 41, no. 1, pp. 61–72, 2020. DOI: 10.1080/02522667.2020.1721593.

[13] A. Laucka, D. Andriukaitis, A. Valinevicius, D. Navikas, M. Zilys, V. Markevicius, D. Klimenta, R. Sotner, and J. Jerabek, "Method for Volume of Irregular Shape Pellets Estimation Using 2D Imaging Measurement," *Applied Sciences*, vol. 10, no. 8, p. 2650, Apr. 2020. DOI: 10.3390/app100.

[14] B. M. AlFawwaz, A. AL-Shatnawi, F. Al-Saqqar, and M. Nusir, "Multi-Resolution Discrete Cosine Transform fusion technique Face Recognition Model", *Data*, vol. 7, no. 6, p. 80, 2022. DOI: 10.3390/data7060080.

[15] A. Setiawan, R. Sigit, and R. Rokhana, "Implementation of face recognition using discrete cosine transform on convolutional neural networks", in *Proc. of 2022 International Electronics Symposium (IES)*, 2022, pp. 502–505. DOI: 10.1109/IES55876.2022.9888295.

[16] F. Tabassum, Md. I. Islam, R. T. Khan, and M. R. Amin, "Human face recognition with combination of DWT and machine learning", *Journal of King Saud University - Computer and Information Sciences*, vol. 34, no. 3, pp. 546–556, 2022. DOI: 10.1016/j.jksuci.2020.02.002.

[17] Z.-H. Huang, W.-J. Li, J. Shang, J. Wang, and T. Zhang, "Non-uniform patch based face recognition via 2D-DWT", *Image and Vision Computing*, vol. 37, pp. 12–19, 2015. DOI: 10.1016/j.imavis.2014.12.005.

[18] Z. B. Lahaw, D. Essaidani, and H. Seddik, "Robust face recognition approaches using PCA, ICA, LDA based on DWT, and SVM algorithms", in *Proc. of 2008 41st International Conference on Telecommunications and Signal Processing (TSP)*, 2018, pp. 1–5. DOI: 10.1109/TSP.2018.8441452.

[19] A. Ouyang, Y. Liu, S. Pei, X. Peng, M. He, and Q. Wang, "A hybrid improved kernel LDA and PNN algorithm for efficient face recognition", *Neurocomputing*, vol. 393, pp. 214–222, 2020. DOI: 10.1016/j.neucom.2019.01.117.

[20] K. Sasirekha and K. Thangavel, "Optimization of K-nearest neighbor using particle swarm optimization for face recognition", *Neural Computing and Applications*, vol. 31, no. 11, pp. 7935–7944, 2019. DOI: 10.1007/s00521-018-3624-9.

[21] R. Atta and M. Ghanbari, "Low-memory requirement and efficient face recognition system based on DCT pyramid", *IEEE Transactions on Consumer Electronics*, vol. 56, no. 3, pp. 1542–1548, 2010. DOI: 10.1109/TCE.2010.5606295.

[22] K. H. Tan and M. Ghanbari, "Compact image coding using two-dimensional DCT pyramid", *Electronics Letters*, vol. 28, no. 8, p. 791, 1992. DOI: 10.1049/el:19920499.

[23] S.-H. Hsiao and J.-S. R. Jang, "Improving ResNet-based feature extractor for face recognition via re-ranking and Approximate Nearest Neighbor", in *Proc. of 2019 16th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS)*, 2019, pp. 1–8. DOI: 10.1109/AVSS.2019.8909884.

[24] L. Li, X. Mu, S. Li, and H. Peng, "A review of face recognition technology", *IEEE Access*, vol. 8, pp. 139110–139120, 2020. DOI: 10.1109/ACCESS.2020.3011028.

[25] A. Abozaid, A. Haggag, H. Kasban, and M. Eltokhy, "Multimodal biometric scheme for human authentication technique based on voice and face recognition fusion", *Multimedia Tools and Applications*, vol. 78, no. 12, pp. 16345–16361, 2019. DOI: 10.1007/s11042-018-7012-3.

[26] S. D. Lin, L. Chen, and W. Chen, "Thermal face recognition under different conditions", *BMC Bioinformatics*, vol. 22, suppl. 5, p. 313, 2021. DOI: 10.1186/s12859-021-04228-y.

[27] H. B. Fredj, S. Bouguezzi, and C. Souani, "Face recognition in unconstrained environment with CNN", *The Visual Computer*, vol. 37, pp. 217–226, 2021. DOI: 10.1007/s00371-020-01794-9.

[28] G. Gao, Y. Yu, J. Yang, G.-J. Qi, and M. Yang, "Hierarchical deep CNN feature set-based representation learning for robust cross-resolution face recognition", *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 32, no. 5, pp. 2550–2560, 2022. DOI: 10.1109/TCSVT.2020.3042178.

[29] M. A. Abuzneid and A. Mahmood, "Enhanced human face recognition using LBPH descriptor, multi-KNN, and back-propagation neural network", *IEEE Access*, vol. 6, pp. 20641–20651, 2018. DOI: 10.1109/ACCESS.2018.2825310.

[30] M. N. ElBedwehy, G. M. Behery, and R. Elbarougy, "Face recognition based on relative gradient magnitude strength", *Arabian Journal for Science and Engineering*, vol. 45, no. 12, pp. 9925–9937, 2020. DOI: 10.1007/s13369-020-04538-y.

[31] S. Khan, M. H. Javed, E. Ahmed, S. A. A. Shah, and S. U. Ali, "Facial Recognition using Convolutional Neural Networks and implementation on smart glasses", in *Proc. 2019 International Conference on Information Science and Communication Technology (ICISCT)*, 2019, pp. 1–6. DOI: 10.1109/CISCT.2019.8777442.

[32] I. William, D. R. Ignatius Moses Setiadi, E. H. Rachmawanto, H. A. Santoso, and C. A. Sari, "Face recognition using FaceNet (Survey, performance test, and comparison)", in *Proc. of 2019 Fourth International Conference on Informatics and Computing (ICIC)*, 2019, pp. 1–6. DOI: 10.1109/ICIC47613.2019.8985786.

[33] P. Viola and M. Jones, "Rapid object detection using a boosted cascade of simple features", in *Proc. of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, CVPR 2001*, 2001, pp. I–I. DOI: 10.1109/CVPR.2001.990517.

[34] M. Alajmi, K. Awedat, A. Essa, F. Alassery, and O. S. Faragallah, "Efficient face recognition using regularized adaptive non-local sparse coding", *IEEE Access*, vol. 7, pp. 10653–10662, 2019. DOI: 10.1109/ACCESS.2019.2890845.

[35] S. Liu, L. Li, M. Jin, S. Hou, and Y. Peng, "Optimized coefficient vector and sparse representation-based classification method for face recognition", *IEEE Access*, vol. 8, pp. 8668–8674, 2020. DOI: 10.1109/ACCESS.2019.2960928.

[36] V. B. T. Shoba and I. S. Sam, "Face Recognition using LBPH descriptor and Convolution Neural Network", in *Proc. of 2018 Second International Conference on Intelligent Computing and Control Systems (ICICCS)*, 2018, pp. 1439–1444. DOI: 10.1109/ICCONS.2018.8663009.

[37] A. Singha, M. K. Bhowmik, and D. Bhattacherjee, "Akin-based Orthogonal Space (AOS): A subspace learning method for face recognition", *Multimedia Tools and Applications*, vol. 79, nos. 47–48, pp. 35069–35091, 2020. DOI: 10.1007/s11042-020-08892-9.

[38] H. Zaaraoui, S. El Kaddouhi, A. Saaidi, and M. Abarkan, "Face recognition with a new local descriptor based on strings of successive values", *Multimedia Tools and Applications*, vol. 80, no. 18, pp.

27017–27044, 2021. DOI: 10.1007/s11042-020-09400-9.
[39] V. Agarwal and S. Bhanot, “Radial basis function neural network-based face recognition using firefly algorithm”, *Neural Comput. & Applic.*, vol. 30, pp. 2643–2660, 2018. DOI: 10.1007/s00521-017-2874-2.
[40] T. S. Akheel, V. U. Shree, and S. A. Mastani, “Stochastic gradient descent linear collaborative discriminant regression classification based face recognition”, *Evolutionary Intelligence*, vol. 15, pp. 1729–1743, 2022. DOI: 10.1007/s12065-021-00585-y.
[41] D. Ali, I. Touqir, A. M. Siddiqui, J. Malik, and M. Imran, “Face recognition system based on four state hidden Markov model”, *IEEE Access*, vol. 10, pp. 74436–74448, 2022. DOI: 10.1109/ACCESS.2022.3188717.