Adaptive Deep Learning Vector Quantisation for Multimodal Authentication

M. Y. Shams [1*], S. H. Sarhan [1] and A. S. Tolba [1]

[1*] Faculty of Computer and Information Sciences
Mansoura University
myysd2016@mans.edu.eg;sh-sarhan@mans.edu.eg;astolba@mans.edu.eg
*Corresponding author:myysd2016@mans.edu.eg

[1]Department of Computer Science
Faculty of Computer and Information Sciences
35516, Mansoura University, EGYPT.

ABSTRACT. *Deep neural network (DNN) techniques are utilised extensively to handle big data problems as well as predicting missing information in retrieval systems. In this paper, we propose a multimodal biometric retrieval system based on adaptive deep learning vector quantisation (ADLVQ) that resolves big data and prediction problems. Intuitively, each subject enrolled in the system is authenticated according to the required degree of security determined by the administrator. We authenticate using not only one face and fingerprint modality but also multi-sample, multi-instances face and fingerprints. The proposed system utilises local gradient pattern with variance (LGPV) to extract the features of the input modalities that are dynamically enrolled in the system. These enrolled features are classified using DNN after quantisation using the K-means algorithm based on prior learning vector quantisation (LVQ) knowledge. Further, the system assesses the performance of the input features adopted with different scenarios taking the priority of the enrolled features into consideration. The results of experiments conducted using occluded black images from the SDUMLA-HMT and CASIA-V5 public standard datasets with different blocks intercepting face images indicate that the proposed system is superior to state of the art systems.*
Keywords: Multimodal biometrics; Deep neural networks; Local gradient pattern with variance; Adaptive deep learning vector quantisation; Face; Fingerprint.

1. **Introduction.** Security is essential for any system based on multimodal biometrics. Such systems frequently require two or more biometric modalities for authentication. For any expert system requiring high security levels, multi-instances biometrics is preferred to protect the system against adversary attacks [1]. Detection of the evidence used by law enforcement agencies in most countries around the world relies on face and fingerprint biometric modalities. For retrieval systems, retrieving the information for multimodal, multi-instances biometrics is very difficult, especially in large-scale applications [2]-[3]. More specifically, there are two major problems affecting large-scale multi-biometric datasets. The first problem is the search strategy used to retrieve the evidence stored in the database, in which the modalities are queried against every identity stored in the database. The second problem is the false acceptance rate (FAR), which significantly increases with the size of the database. Classification and indexing schemes are typically used to filter biometric databases. For example, Gyaourova and Ross [2] proposed an indexing method
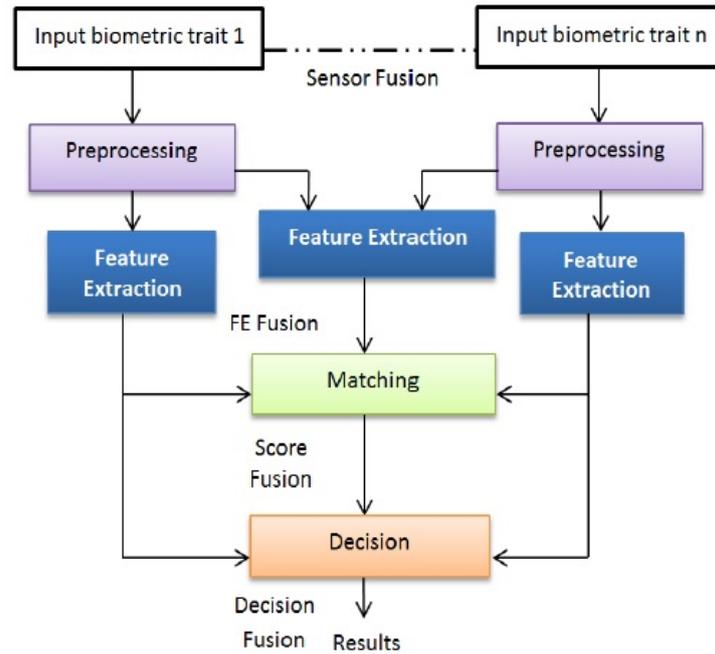
FIGURE 1. General schematic diagram for multimodal biometrics.

that generates an index code for each enrolled template (face and fingerprint) in order to retrieve the probe image from the candidate list. The general structure of the most recent multimodal biometric systems is shown in Figure 1. The structure consists of five fusion levels, proceeding from the sensor level to the decision level in multimodal biometrics. The data are compressed after fusion operations, as shown in Figure 2 [4].

Matching score level fusion (also called confidence or score level) is considered as the most common fusion level in multimodal biometrics owing to the relationship between the contents of the fused information and the classifier combination [5]. In this paper, we propose a multimodal biometric system based on adaptive deep learning vector quantisation (ADLVQ) that addresses the retrieval problem. In addition, we address the occlusion and interclass variation with different scenarios problems. The proposed system is based on the characteristics and structure of the learning vector quantisation (LVQ) algorithm, which is used to handle overfitting and memory problems. DNNs comprise a family of machine learning algorithms that helps abstraction levels in data to employ the architectures of multiple nonlinear transformations [3]. Vector quantisation is used to compress the DNN parameters. This results in improvement of the system efficiency as it creates a balance between model size and recognition rate. In this study, we utilised large-scale databases from SDUMLA-HMT and CASIA-V5 which contain multi-sample face images and multi-instances fingerprints. The main contributions of this paper are as follows:

- We propose a robust adaptive multimodal biometric system based on multi-sample face images and multi-instances fingerprints.
- We integrate both gradient features and the variance histogram of input images to address the problem of training and quantisation using LGPV.
- The proposed algorithm introduces the enrolled templates in both serial and parallel fashion.
- We use vector quantisation based on K-means to tackle memory and overfitting problems.
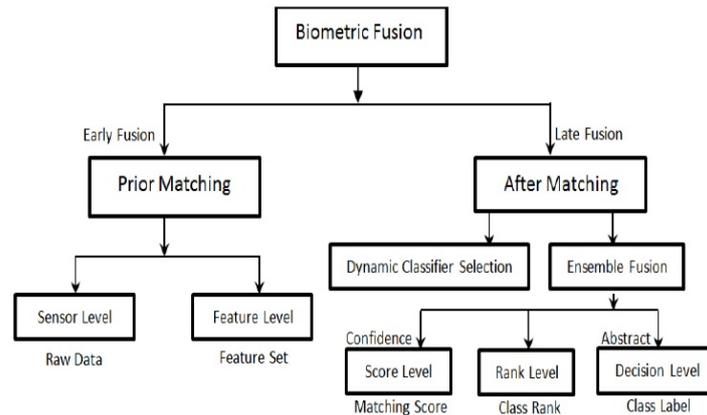
FIGURE 2. Taxonomy of fusion levels in biometrics [4].

- We exploit the advantages of DNNs in both classification and feature extraction to predict missing occluded data using the expectation maximisation (EM) algorithm.
- The proposed system can be utilised for information retrieval and to thwart spoofing attacks.

The remainder of this paper is organised as follows: Section 2 presents related work. Section 3 outlines our methodology. Section 4 discusses the experimental evaluations conducted and analyses the results obtained. Section 5 concludes this paper and outlines possible future work.

2. **Related Work.** Deep learning is defined as the composition of multiple layers based on nonlinear operations by which data can be reconstructed from a number of hidden levels. An architecture can be formulated using classical neural networks [6]. During the training process, two major problems must be taken into consideration: (1) adaptation of lower layers to provide upper layer settings, (2) adaptation of upper layers in order to use the final lower layer that deeply requires unsupervised learning [7]. There are three types of deep learning architectures: feed-forward, feedback, and bidirectional. Examples of feed-forward deep architecture include multilayer neural nets and convolutional nets. Stacked sparse coding and deconvolutional nets are examples of feedback deep architecture. Deep Boltzmann machine (DBM) and stack auto-encoder are examples of bidirectional deep architecture [8]. A DNN can be established using more than one neural network (NN) layer by stacking the NN layers towards the output layers inferred by the overall process iterations [9]. Chen et al. [10] illustrates the construction of deep belief network (DBN) that contains several hidden layers and the connections between each layer. In general, DNNs can be used as feature extractors and/or classifiers. Further, rectified linear unit (ReLU) and maxout functions can be used for network modifications. To handle the optimisation problems, an adaptive learning rate can be introduced; furthermore, dropout can be applied to prevent overfitting problems. Because deep learning strategies need memory in some cases, the output of the hidden layer are stored and considered as another output in recurrent neural networks (RNN) [11]. Bruna and Mallat [12] conducted mathematical model analysis for wavelet scattering networks based on deep convolution networks for classification purposes. In some deep learning networks, the top layers produce feature vectors that are vulnerable to perturbations that need quantisation and dimensionality reduction (DR) [13]. Thus, deep quantisation for image retrieval, as carried out by Cao et al. [14], is necessary. Wu et al. [15] proposed a semi-supervised model for multimodal gesture recognition that depends on deep dynamic neural networks.

They use a Gaussian Bernoulli DBN to extract skeleton features, and 3D convolutional NNs for both depth and RGB extracted features. Further, they obtain emission probabilities using HMM to produce gesture patterns for recognition. Ngiam et al. [16] proposed a multimodal deep learning model based on audio and video sounds and presented three learning settings: multimodal fusion using shallow learning, cross modality learning using restricted Boltzmann machines (RBM), and shared representation learning using a DBN. Ding and Tao [17] used convolutional neural networks for facial expression. Menotti et al. [18] proposed a combination of both architecture and filter optimisation based on convolutional deep representation of iris, face, and fingerprint modalities to counter spoofing attacks. They evaluated the convolutional deep architecture based on linear support vector machine (SVM) scores issued by each convolutional neural network layer. Lumini and Nanni [19] conducted a detailed survey of the most recent combination techniques in biometric matchers. They analysed the methodologies, architectures, and evaluation results for both unimodal and multimodal biometric systems. Subsequently, they suggested deep learning neural networks in multimodal biometric systems as a future direction. In this section, we highlight the most recent multimodal biometrics and DNNs related to our research.

2.1. **Multimodal face and fingerprint modalities.**

2.1.1. *Face modality.* In general, face images are represented as feature vectors that are subject to indexing, searching, and ranking in feature space [3]. Lin et al. [20] proposed orthogonal enhanced linear discriminant analysis (OELDA), which maps the higher dimension subspace of face images to lower dimensions, for face DR. Li and Suen [21] handled the occlusion problem using dynamic rank representation with satisfactory results. Learned-Miller et al. [22] listed recent efforts to recognise and detect labelled face images based on deep learning approaches. In [23], we proposed a multi-view face recognition system based on application of Additive white Gaussian noise with different poses and variation angles. In this paper, we propose an adaptive system for multi-sample face images that handles interclass and poses variations based on DNNs. We also address occlusion problems with different block size images.

2.1.2. *Fingerprint modality.* For fingerprint retrieval tasks, an indexing code is generated using a set of predetermined fingerprints that are used to search the model index codes, as accomplished by Gyaourova and Ross [24]. Nanni et al. [25] presented a combination of fingerprint matchers with likelihood estimation based on statistical approaches. Peralta et al. [26] conducted a detailed survey of fingerprint minutiae-based local matching techniques. Kumari and JayaSuma [27] conducted a comparative analysis of these techniques. Kumar et al. [28] proposed a fingerprint-matching algorithm based on extraction of the orientation features of the ROI fingerprint. Their proposed algorithm uses Euclidean distance to determine the distance between the extracted orientation features and the stored images of fingerprints. Jain and Prasad [29] proposed a dynamic clustering-based fingerprint-indexing scheme. Su et al. [30] proposed a learning-based fingerprint pose estimation algorithm for indexing fingerprints into a common finger coordinate system. Anush et al. [31] proposed an adaptive latent fingerprint segmentation method based on random decision classification. In this paper, we propose using multi-instances fingerprints to protect systems against spoof attacks. In the proposed system, five instances representing one hand for each subject enrolled in the system are indexed based on the extracted LGPV code with the class-enrolled number.

2.1.3. *Multimodal biometrics.* In previous work [32], we conducted a comparative study of the most common biometric traits, specifically, iris, fingerprint, finger vein, and face. Neural classifiers can be separated into two groups: base classifiers and ensemble classifiers. A base classifier is learned using a single feature and is trained using a single fixed training database, while an ensemble classifier is formed from a combination of base classifiers [33]. Combined weak LVQ classifiers are integrated to enhance the overall accuracy of the system, as in [34]. In [35], we used face, iris, and fingerprint multimodal biometric traits to identify subjects purportedly claiming to be in the system. The system uses not only multimodal biometric traits but also multi-sample face images, and multi-instances iris and fingerprint images. The evaluation results were based on a validated SDUMLA-HMT dataset with only 106 individuals and a fusion process based on majority voting algorithm results from Combined LVQ classifiers. We expected the genuine acceptance rate (GAR) to decrease when the iris traits were removed and the multimodal biometric database expanded, as presented in this work. Marcialis et al. [36] proposed serial fusion of face and fingerprint to tackle the drawback of parallel fusion. Biggio et al. [37] proposed a system based on fusing face and fingerprint that protects against spoof attacks by using confidence level to distinguish genuine users. Poh et al. [38] proposed a criterion for ranking users based on the training scores of the ranked subjects. Shekhar et al. [39] proposed a multimodal biometrics based on joint sparse feature level fusion of face, iris, and fingerprint. Bharadwaj et al. [40] proposed an adaptive biometric system based on selection of the best biometric matcher to verify the identity of individuals. Nguyen et al. [41] presented a Dempster-Shafer (D-S) theory-based model for multimodal biometrics in uncertainty factors. Their evaluation results are related to the combination of the quality measures for the input data and classifier based on D-S theory. D-S can be used to enhance the overall performance of a system based on multiple rejection strategies, as in [42]. Wild et al. [43] demonstrated fusion of face and fingerprint scores with liveness scores based on 1-median filtering as spoofing-resistant against sum-rule. Chen et al. [44] also proposed a multimodal biometrics recognition system based on local fusion of visual features and an extreme learning machine (ELM). They use an algorithm to delete the undesired duplicated features and keep only the required features in local visual features by assigning the input weights randomly. In general, there are two strategies for multimodal biometrics fusion: parallel fusion, which depends on matching scores, and serial fusion. In this paper, we present both serial and parallel fusion of face and fingerprint. Furthermore, we propose an ADLVQ system that quantises the observed features and produces codewords using the K-means algorithm. Evaluation scores are obtained based on the K-NN algorithm and the EM algorithm for the unobserved data.

3. **Methodology.** In this section, we present the structure of the proposed system and describe the main objective of each stage in the overall system. The general framework of the proposed system is shown in Figure 3. The main parts determine the overall performance of the proposed system in both training and testing phases. They are enrollment, preprocessing, feature extraction, and ADLVQ classifier.

3.1. **Enrollment phase.** The acquisition process is typically the first step in any generic biometric system. In this step, each subject starts to deal with the biometric system by acquiring biometric evidence. The enrollment involves the generated templates and storing of the resulting templates in the database after following the preprocessing and feature extraction stages. Failure to enrol (FTE) is one of the important benchmarks for measuring the performance of biometric systems. Designing ergonomic and convenient graphical user interfaces (GUIs) can efficiently lead to lower FTE [4]. In this paper,
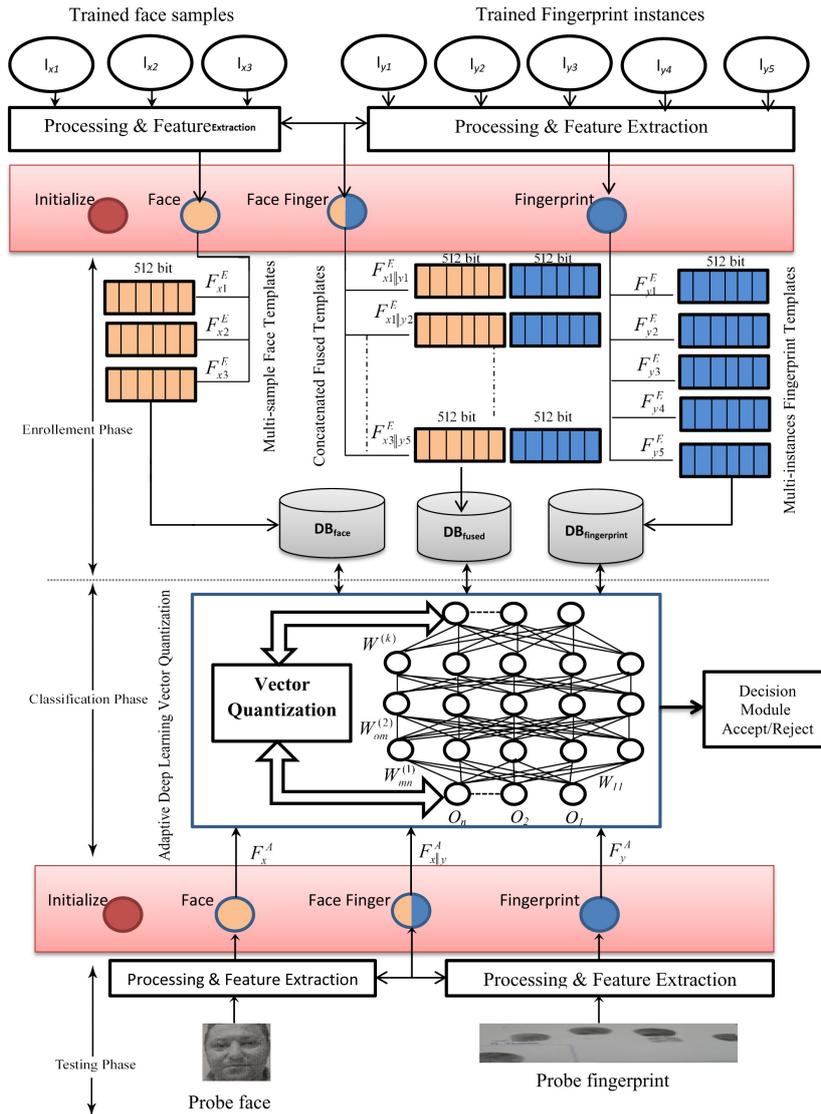
FIGURE 3. General schematic diagram for the proposed multimodal biometric system based on ADLVQ..

we implemented a GUI that acquires subjects' modalities in both parallel and series fashion. In the training phase, we have four cases produced by enrolling the acquired face and fingerprint modalities, as shown in Table 1. Case 0 is the initial case, in which the evidence are prepared to enrol the biometric traits to the system and the previously trained modalities cleared. In Case 1, each person entered into the system submits their fingerprints, and the system enrols five instances for one hand simultaneously. Case 2 enables users to enrol three face modality samples in parallel. Finally, in Case 3, 15 concatenated templates representing multimodal face and fingerprint biometric traits are gathered and stored in the database. Three distinct databases are created and indexed based on Cases 1, 2, and 3, respectively, for each enrolled subject, as shown in Figure 3.

Invariably, the enrollment process in the training phase consumes more time to ensure non-repudiation authentication. Thus, the system acquires and produces the templates in parallel (i.e., parallel acquisition), which helps to reduce the time consumed. In order to retrieve the templates from the database, each template is indexed using the main information associated with the generated templates and the class number of each person.

TABLE 1. Possible four cases for acquiring face and fingerprint modalities

Case	Face modality	Fingerprint modality	Function
0	N	N	Initialise the training network
1	N	Y	Input five fingerprint instances for training
2	Y	N	Input three face samples for training
3	Y	Y	Input $5_f ingerprint \times 3_f ace$ concatenated fingerprint,and face biometric traits, respectively

The problem of spoof attacks is tackled in accordance with two possible scenarios, for which the procedure in the system is adapted. In the first scenario, the user enrols his/her face first and needs one or more instances of fingerprint modalities; in the second, the user first enrols one or more fingerprint instances and needs one or more face modality samples, as shown in Figure 4. The proposed system ensures that the probability of the system suffering from spoof attacks will be low as long as it randomly chooses the second of one or more traits given the first modality. For example, if the user enrols his/her face first and then wishes to enter the system, the system will respond dynamically by choosing one or more fingerprint modalities to be concatenated with the face image generated by one of the fifteen concatenated pre-stored templates in the database. Any hacker must have fifteen concatenated templates in order to break the security of the system against spoof attacks. Thus, such an attack is unlikely to succeed as the liveness inspired by Wild et al. [43] is required.

3.2. **Preprocessing and Feature extraction.** : The preprocessing is based on our prior work in [35]. We tested our system using a well-known local binary pattern (LBP) variant to extract the gradient information of the input image, called local gradient pattern (LGP), The gradient information of the preprocessed images was successfully extracted using the LGP operator, which is insensitive to global intensity variations of the input images [45]. Figure 5 shows the calculation of the LGP operator for a 3×3 image [46]. Recently, research on gradient information, such as local gradient probabilistic pattern (LGPP) [47] and gradient weighted histogram of local binary pattern (GWH-GLBP) [48], has attracted increased attention. The approaches seek to overcome the drawbacks associated with LBP. In this work, we modified the variance histogram of the LGB to obtain local gradient pattern with variance (LGBV), in order to integrate both gradient features and the variance histogram of input images to address the problem of training and quantisation presented by Guo et al. [49]. Following application of the LBPV operator to the LGB, the resulting LGPV code is determined by the following equations:

$$LGP_{P,R} = \sum_{p=0}^{p-1} s(g_p - g_c)2^p \tag{1}$$

$$S(x) = \begin{cases} 1, x \geq 0 \\ 0, x < 0 \end{cases} \tag{2}$$

$$u = \frac{1}{P} \sum_{p=0}^{p-1} g_p \tag{3}$$

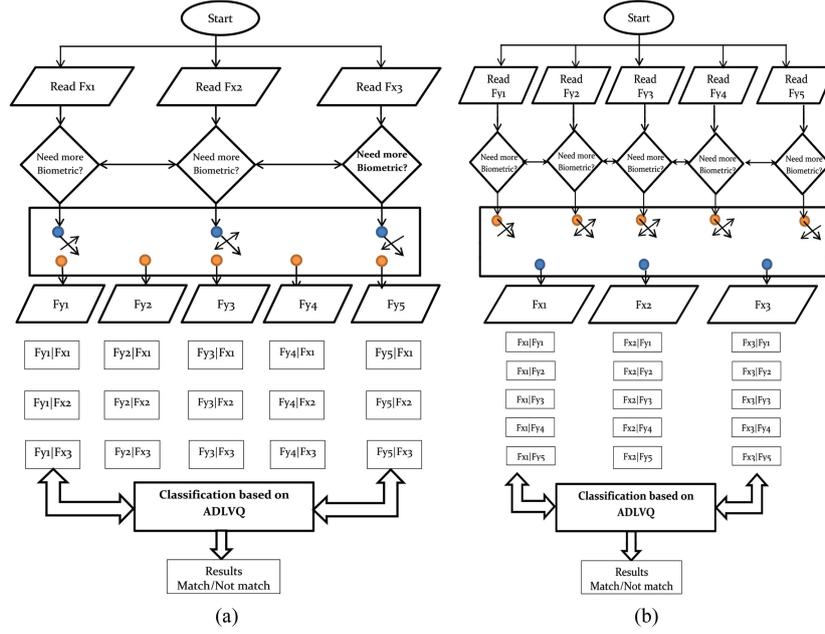$$VAR_{P,R} = \frac{1}{P} \sum_{p=0}^{p-1} (g_p - u)^2 \tag{4}$$

FIGURE 4. Testing phase of the proposed ADLVQ system with two scenarios: a) user enrols face samples first. b) user enrols fingerprint instances first.
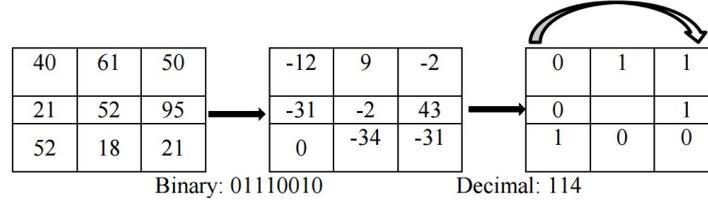

FIGURE 5. Example of 3×3 LGP operator computation.

$$LGPV_{P,R}(k) = \sum_{i=1}^{N} \sum_{j=1}^{M} w(LGP_{P,R}(i,j),k), k \in [0,1] \tag{5}$$

$$w(LGP_{P,R}(i,j),k) = \begin{cases} VAR_{P,R}(i,j), LGP_{P,R}(i,j) = k \\ 0, otherwise. \end{cases} \tag{6}$$

where P is the sampling pixel with radius R, and g_c and g_p are the grey value of the central pixel and the grey value of the pth neighbour, respectively. In this work, there are a number of possible inference forms. They are as follows: (1) Consider that the input face training images for one subject are I_{x1}, I_{x2}, and I_{x3}, and the input fingerprint instances for one subject are $I_{y1}, I_{y2}, I_{y3}, I_{y4}$ and I_{y5}. (2) The extracted features resulting from the face and fingerprint are normalised using Z-score normalisation, and the concatenated features resulting from fusing of face and fingerprint are selected (see Figure 6). The summarised indexing and templates storage steps are as follows:

1. Each extracted template from LGPV is normalised using z-score normalisation.
2. Each template is indexed based on its class number.
3. Template = Class no. + LGPV Code.
4. The templates are stored in [DBface, DBfingerprint, and DBfused]

$$F_x^1 = \begin{bmatrix} F_{x1}^1 \\ F_{x2}^1 \\ F_{x3}^1 \end{bmatrix}$$

Extracted face features
for subject # 1

Extracted face features
for subject # N

$$F_x^N = \begin{bmatrix} F_{x1}^N \\ F_{x2}^N \\ F_{x3}^N \end{bmatrix}$$

$$F_y^1 = \begin{bmatrix} F_{y1}^1 \\ F_{y2}^1 \\ F_{y3}^1 \\ F_{y4}^1 \\ F_{y5}^1 \end{bmatrix}$$

Extracted fingerprint features
for subject # 1 to N

$$F_y^N = \begin{bmatrix} F_{y1}^N \\ F_{y2}^N \\ F_{y3}^N \\ F_{y4}^N \\ F_{y5}^N \end{bmatrix}$$

$$F_{x\|y}^1 = \begin{bmatrix} F_{x1\|y1}^1 \\ F_{x2\|y1}^1 \\ F_{x3\|y1}^1 \\ ..... \\ F_{x3\|y5}^1 \end{bmatrix}$$

Concatenated features for
subject # 1 to N

$$F_{x\|y}^N = \begin{bmatrix} F_{x1\|y1}^N \\ F_{x2\|y1}^N \\ F_{x3\|y1}^N \\ ..... \\ F_{x3\|y5}^N \end{bmatrix}$$

FIGURE 6. Extracted features for face, fingerprint, and the concatenated templates for one to N subjects.

3.3. **Adaptive Deep Learning Vector Quantisation (ADLVQ).** DNNs are one of the most popular approaches for handling the problem of feature extraction and classification. Sparse representation of the input images and prediction of the absent features of an object are the main challenges in computer vision and pattern recognition. Dealing with large-scale databases and consuming minimum processing time to retrieve image information are also very important challenges faced by researchers in the image-processing field. In this paper, we propose ADLVQ for classifying the input extracted templates. Vector quantisation is used to handle the overfitting and time consumption problems. The extracted features from LGPV are quantised and clustered using the K-means algorithm presented by Huang et al. [50]. Practically, dimensional reduction of the features is desired without significant loss of template information. Principal component analysis (PCA), Dual PCA, and K-means are very common DR approaches. In this study, we adopted K-means as an unsupervised DR algorithm to cluster a set of features into K sets (clusters) while ensuring that the points in each cluster are close to each other. The approximated data are given by $o_i \approx \mu_{\zeta_i}$. $\zeta_i \in \{1, 2, ..., K\}$ is an index that identifies which of the K prototype vectors $\{\mu_k\}_{k=1}^K$ approximates the ith example. Equation (7) defines the assignment of the prototype vector optimisation. Equation (8) minimises the cost function based on an alternative strategy to the to the nearest prototype. Thus, the prototypes are updated as in Equation (9), where $\delta(0)$ returns one when its argument is zero, and returns zero otherwise [51] [52].

$$\hat{\mu}_{1...K}, \hat{\zeta}_{1...I} = \arg\min_{\mu,\zeta} \left[\sum_{i=1}^I (o_i - \mu_{\zeta_i})^T (x_i - \mu_{\zeta_i}) \right] \tag{7}$$

$$\hat{\zeta}_i = \arg\min_{\zeta_i} \left[(o_i - \mu_{h_i})^T (x_i - \mu_{\zeta_i}) \right] \tag{8}$$

$$\hat{\mu}_K = \arg\min_{\mu_K} \left[\sum_{i=1}^I (o_i - \mu_{h_i})^T (x_i - \mu_{\zeta_i}) \right] \tag{9}$$

$$\hat{\mu}_K = \frac{\sum_{i=1}^{I} o_i \delta \left[\zeta_i - k \right]}{\sum_{i=1}^{I} \delta \left[\zeta_i - k \right]} \tag{10}$$

The resulting clustered feature vector $o_1, o_2, ...o_n$ are the observed features of the visible layer in the DNN, which is constructed by stacking more than one RBM layer. Maximum likelihood is used to estimate the RBM given the observed training data. The inference of the RBM is theoretically similar to the inference of a single-layer neural network [9] [53]. Equation (11) shows the vector representation of a single-layer neural network.

$$h = \sigma(w^T o + b) \tag{11}$$

where o is an observed visible feature, w is the weight, and b is a constant. Dimensionality reduction is achieved by minimising the difference between the output data and the teaching data (t) that occur when estimating the weights w and the bias b given the objective function I(θ), as in Equation (12). The gradient is computed via partial differentiation of the object function, as in Equation (13). It is calculated using back-propagation, which is difficult to implement but is computationally more efficient, in order to simplify the implementation of the differential approximation in Equation (15).

$$I(\theta) = \frac{1}{2} \sum_k h_k^{(L)}(o; \theta) - t_k)^2 \tag{12}$$

$$\frac{\partial I}{\partial \theta_i} = \lim_{\varepsilon \to 0} \frac{I(\theta + \varepsilon \psi_i) - I(\theta)}{\varepsilon} \tag{13}$$

$$\psi = \begin{cases} 1, & i^{th} element \\ 0, & otherwise \end{cases} \tag{14}$$

$$\Delta_i I = \frac{I(\theta + \varepsilon \psi_i) - I(\theta)}{\varepsilon} \tag{15}$$

Note that $\frac{\partial I}{\partial \theta_i} = \Delta_i I$ for small ε. The practical parameters are updated and the differential approximation depends on the gradient, weight decay, and momentum, as shown in Equations (16) and (17).

$$\theta^{(t+1)} = \theta^{(t)} + \Delta \theta^{(t)} \tag{16}$$

$$\Delta \theta^{(t)} = -\eta \frac{\partial I}{\partial \theta} - \lambda \theta^{(t)} + \nu \Delta \theta^{(t-1)} \tag{17}$$

where η, λ, and ν are the learning rate, weighted decay rate, and the momentum rate. Note that these parameters are empirically determined [54]. The probabilistic and energy models of the RBM are given by Equations (18), (19), (20), and (21).

$$P(o, h; \theta) = \frac{1}{Z(\theta)} e^{-(E(o,h;\theta))} \tag{18}$$

$$E(o, h; \theta) = -\sum_{i,j} o_i w_{i,j} h_{j-} \sum_j b_j h_j - \sum_i c_i o_i \tag{19}$$

$$E(o, h; \theta) = -o^T W h - b^T h - c^T o \tag{20}$$

$$Z(\theta) = \sum_{o,h \in \{0,1\}} e^{-(E(o,h;\theta))} \tag{21}$$

Equations (22) and (23) computes the posterior probability over the hidden variable h_j using conditional probability $P(h_j|o;\theta)$ (inference model):

$$P(h_j|o;\theta) = \frac{P(o,h;\theta)}{\sum\limits_{h\in\{0,1\}} P(o,h;\theta)} = \frac{\prod_i e^{c_i o_i}\left(\prod_j e^{\sum_i o_i w_{i,j} h_j + b_j h_j}\right)}{\prod_i e^{c_i o_i}\left(\prod_j \sum_{h_j\in\{0,1\}} e^{\sum_i o_i w_{i,j} h_j + b_j h_j}\right)} \qquad (22)$$

$$P(h_j|o;\theta) = \frac{e^{\sum_i o_i w_{i,j} h_j + b_j h_j}}{\sum_{h_j\in\{0,1\}} e^{\sum_i o_i w_{i,j} h_j + b_j h_j}} \qquad (23)$$

The EM algorithm provides sufficient information to use the RBM to fit parameter θ in the model, as in Equation (24), and is applied to handle the unobserved hidden data [51].

$$\overset{\wedge}{\theta} = \arg\max_\theta \sum_{j=1}^{J} \log P(o,h;\theta) \qquad (24)$$

In this paper, we use two modalities: face samples and multi-instances fingerprints. The extracted features based on LGPV are first clustered and quantised using the K-means algorithm. The clustered features are then the visible layer of the DNN. The DNN is formulated by stacking RBM layers. The construction of hidden layers hj is based on the Sigmoid function, whereas the output layer represents a softmax layer for the DNN. Consider N activations for each class such that $a_n = \varphi_n^T o$, where $\{\varphi_n\}_{n=1}^{N}$ are the parameter vectors. The softmax of N activation parameter vectors is found using Equation (25) [51]:

$$softmax\,[a_1, a_2, ...a_N] = \frac{e^{a_n}}{\sum_{j=1}^{N} e^{a_j}} \qquad (25)$$

The codewords learned by DNN are generated based on production of bag of words (BoW) by the K-means algorithm. Recognition is realised in the decision step by which the classes are produced by clustering the entered feature vectors of the observed template using the K-means algorithm. For supervised data, we obtain the matching scores of the BoW using a fast K-nearest neighbour (KNN) algorithm [55]. In this paper, we propose an algorithm that switches between two schemes: a parallel scheme in which subjects are enrolled simultaneously, see Figure 3, and a sequential scheme in which the system dynamically chooses one or more biometric traits given the acquisition template of each subject, as shown in Figure 4. Vector quantisation (VQ) is utilised to compress the input data and then the resulting codewords are learned by deep learning (DL). The proposed system is adaptive as the output of the system depends fundamentally on the input features. The procedural parameter adaptively receives any data that will be updated and affected by the input features. Any missing features or unobserved data undergo feedback to be adapted using vector quantisation and DNN classifier. The overall steps in the proposed system are summarised in Algorithms 1.

Algorithm 1 Adaptive Deep Learning Vector Quantisation (ADLVQ)

Input: Set of feature vector results from LGPV. Let N be the total number of enrolled subjects:

- $o_x = Fx = \{F_{x1}, F_{x2}, F_{x3}\}_{n=1}^{N}$ Extracted face samples.
- $o_y = Fy = \{F_{y1}, F_{y2}, F_{y3}, F_{y4}, F_{y5}\}_{n=1}^{N}$ Extracted fingerprint multi-instances.
- $o_z = Fxy = \{F_{x1y1}, \ldots\ldots F_{x3y5}\}_{n=1}^{N}$ Concatenated face and fingerprint templates.

Initialisation: Pre-train the network parameters o_x, o_y, and o_z through the DNN layer by layer; Set the weights, W, to zero; Quantise the observed vectors, o using the K-means algorithm, given $o_i \approx \mu_{\zeta_i}$.

Procedure:

1. Set the vector representation of a single-layer RBM network $h = \sigma(w^T o + b)$
2. Calculate the gradient information of the observed FV templates $\Delta_i I = \frac{I(\theta + \varepsilon \psi_i) - I(\theta)}{\varepsilon}$
3. Update approximation FV based on the gradient, weight decay, and momentum $\theta^{(t+1)} = \theta^{(t)} + \Delta\theta^{(t)}$, $\Delta\theta^{(t)} = -\eta\frac{\partial I}{\partial \theta} - \lambda\theta^{(t)} + \nu\Delta\theta^{(t-1)}$
4. Determine the inference model based on posterior probability over the hidden variable, hj: $P(h_j | o; \theta) = \frac{e^{\sum_i o_i w_{i,j} h_j + b_j h_j}}{\sum_{h_j \in \{0,1\}} e^{\sum_i o_i w_{i,j} h_j + b_j h_j}}$
5. Use expectation maximisation (EM) to provide sufficient information to handle the unobserved hidden data: $\overset{\wedge}{\theta} = \arg\max\limits_{\theta} \sum\limits_{j=1}^{J} \log P(o, h; \theta)$
6. Apply the softmax of N activation parameter vectors: $softmax\,[a_1, a_2, ...a_N] = \frac{e^{a_n}}{\sum_{j=1}^{N} e^{a_j}}$
7. Generate the codewords using the K-means algorithm to produce BoW vector representation.
8. Use K-NN to classify the input BoW.
9. Determine the similarity metric between the query qi and stored oi templates using Euclidean distance with dynamic metric adaptation, called the asymmetric quantiser distance (AQD): $AQD = \sum\limits_{n=1}^{N} \|(q_i - o_i)\|_2^2$

Output:

1. Matching scores for each class S = $(S_1, ... S_n)$.
2. A decision (Accept/Reject) based on confidence level.

---

4. **Experimental Evaluations.** To investigate the efficacy of the proposed ADLVQ-based system for personal verification, we evaluated it on two publicly available face and fingerprint databases: SDUMLA-HMT [56] and CASIA-V5 [57]. SDUMLA-HMT is a multimodal biometric database containing five biometric modalities, face, fingerprint, finger vein, iris, and gait, from 106 individuals. CASIA-V5 contains face and fingerprint and is used as a chimeric multimodal dataset with 500 subjects. We utilised it to evaluate the performance of the proposed system with a large-scale database, see Table 2. The total number of face images contained in the SDUMLA-HMT dataset is 8904. We utilised only 424 face images: 318 (= 106 × 3) face images for training and 106 for testing, as we took three poses for training and randomly chose another different pose image to test the overall collected number of face images.

We collected 530 fingerprint images (= 106 × 5) for training and another 530 for testing, with five fingerprint instances per subject. Ideally, in order to prove the superiority of the proposed system over the state of the art, a large-scale public multimodal biometric

TABLE 2. Number of images collected for reference databases SDUMLA-HMT and CASIA-V5.

Public database	Modality	Overall DB	No. of trained images	No. of tested images	No. of collected images	Total No. of subjects
SDUMLA-HMT	Face	8904	318	106	424	106
	Fingerprint	25440	530	530	1060	
CASIA-V5	Face	2500	1500	1000	2500	500
	Fingerprint	20000	2500	2500	5000	

TABLE 3. Hyperparameter values for the ADLVQ network

Hyperparameters	Values
Input data number for one modality (Face or Fingerprint)	512
Input data number for concatenated Face and Fingerprint	1024
Output number	16
Hidden number	8
Input number	4
Maximum iteration number (no. of epochs)	200
Initial momentum	0.5
Final momentum	0.9
Weight cost	0.002
Learning rate	0.01

database is required. Unfortunately, however, to the best of our knowledge, SDUMLA-HMT is the only free multimodal biometric database that collects face and fingerprints from the same subject. Therefore, we empirically evaluated our results using CASIA-V5 as a heterogeneous database. We utilised 2500 (= 5 × 500) face images from CASIA-V5-1500 (= 3 × 500) for training and 1000 (= 2 × 500) for testing. Further, we utilised 5000 fingerprints-2500 (= 5 × 500) for training and another 2500 for testing (note that the number of fingerprint instances was five). In the training phase, following extraction of the features using LGPV we obtained the feature vectors, which are the input patterns to the ADLVQ classifier. As shown in Figure 3, the templates were produced and normalised to 512 FV length, after which the features were encoded, indexed, and stored in the template database. The construction of ADLVQ is based on the following hyperparameters that are empirically fixed through the experimental evaluation results, as shown in Table 3.

In the testing phase, to compare the probe features against the gallery templates stored previously in the database, we utilised Euclidean distance with dynamic metric adaptation, called asymmetric quantiser distance (AQD), to measure the distance scores. By using matching scores, we determine the similarity for smaller values of the Euclidean distance scores in the case of concatenated templates that fuse both face and fingerprints. In general, the experimental results are based on determination of the factors FAR, FRR, and error rate (ERR) for different threshold values. FAR and FRR are calculated based on the generated matching scores for all possible genuine and imposter subjects as follows:

1. FAR: Probability that the system will accept an imposter. This value is obtained by determining the ratio of the match count (MC) to the total number of imposter persons (IP) based on matching scores.
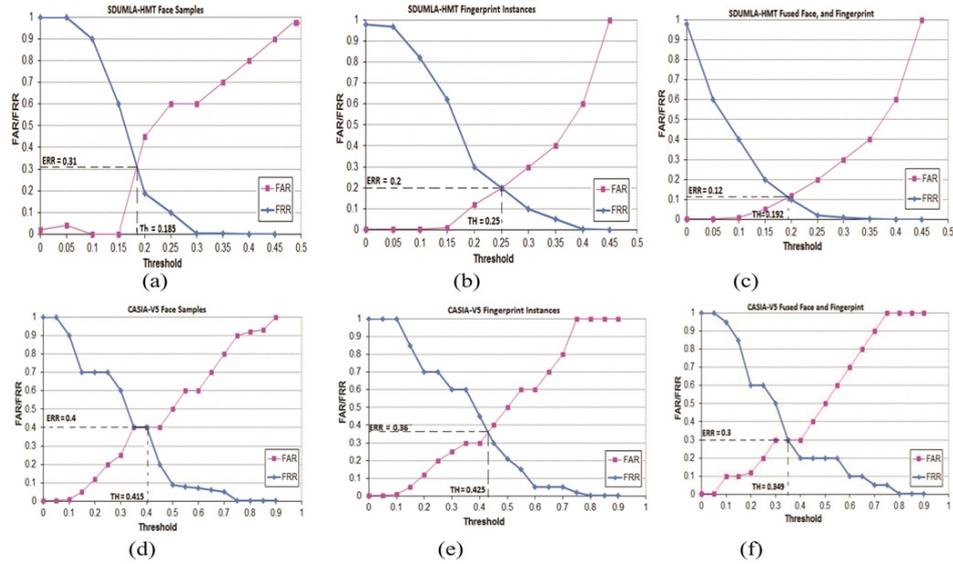
FIGURE 7. FAR/FRR versus various threshold values for average face samples, (a) and (d); fingerprint instances, (b) and (e); and concatenated face and fingerprint, (c) and (f), for SDUMLA-HMT and CASIA-V5, respectively.

2. FRR: Probability that the system will reject a genuine subject using mismatch count (MMC) to the total number of genuine persons (GP).
3. Determination of the threshold value, which is a measure of the acceptance or rejection of biometric data (face and/or fingerprint), according to the matching score falling above or below the threshold.
4. Equal error rate (ERR) is defined as the intersection between FAR and FRR, where FAR is equal to FRR. A smaller ERR ensures a more precise biometric system.

Figure 7 shows the variations of both FAR and FRR against the threshold values using the average values of face samples and fingerprint instances obtained for the SDUMLA-HMT and CASIA-V5 databases. The results are based on parallel acquisition of the biometric traits for face, fingerprint, and the concatenated features. It can be seen that the concatenated face and fingerprint have a very small ERR of 0.12 in the case of SDUMLA-HMT and 0.3 for CASIA-V5. Further, Figure 8 shows a plot of the receiver operating characteristic (ROC), which measures the classification performance of the proposed system, for the GAR against the FAR in the presence of multi-sample face images and multi-instance fingerprints for SDUMLA-HMT and CASIA-V5. Note that the average values of the fifteen fused concatenated features of both the face and fingerprint are used. The experimental results demonstrate the efficiency of the proposed system in terms of retrieval information based on the probe features in the testing phase. As mentioned in Section 3.1, two possible scenarios are covered. The first scenario covers a person trying to enter the system with one or more of his/her biometric traits and the administrator wishing to protect the system against spoof attacks. In this scenario, it is assumed that the person has enrolled his/her face first and so the system randomly requests the other one or more traits of the multi-instance fingerprint, as shown in Figure 4-a. In the second scenario, the multi-instances are first enrolled and the system randomly requests one or more face image samples, see Figure 4-b. Figure 9 shows the results obtained for the stated scenarios on the SDUMLA-HMT and CASIA-V5 datasets.
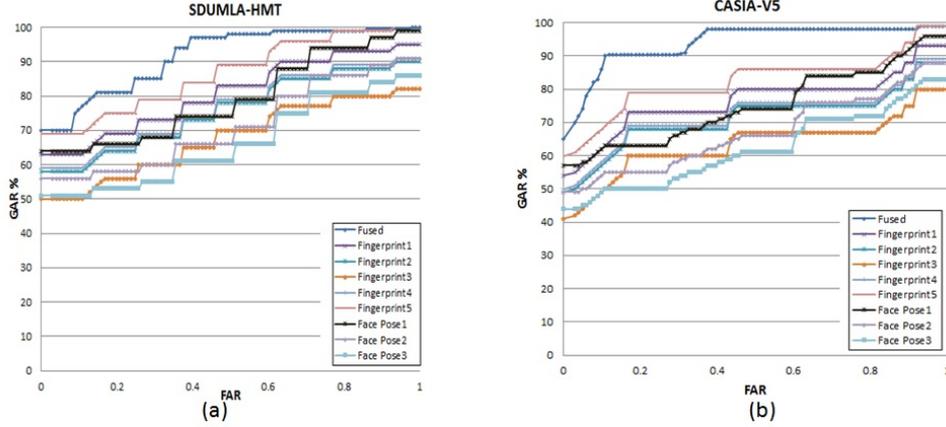
FIGURE 8. ROC curves showing the performance of multi-sample face images and multi-instances fingerprints for the (a) SDUMLA-HMT and (b) CASIA-V5 datasets.
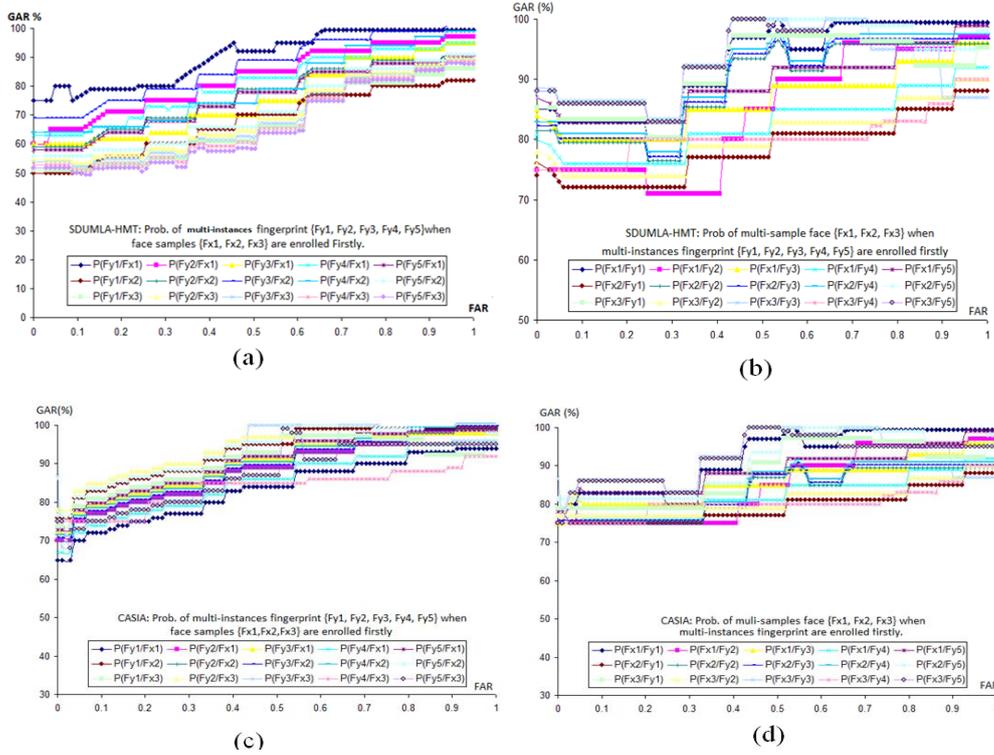

FIGURE 9. ROC curves showing the performance for fused face and fingerprint as follows: (a) and (c) Probability of multi-instances fingerprints given face samples being first enrolled for the SDUMLA-HMT and CASIA-V5 datasets, respectively; (b) and (d) Probability of multi-sample face images given multi-instance fingerprints being first enrolled for the SDUMLA-HMT and CASIA-V5 datasets, respectively.

Note that P(Fy1/Fx1) signifies the conditional probability of fingerprint number 1 given the probability of face sample number 1. It is calculated as in Equation 26:

$$P\left(\frac{F_{y1}}{F_{x1}}\right) = \frac{P\left(F_{y1}, F_{x1}\right)}{P\left(F_{x1}\right)} = \frac{P_{concatenated}\left(F_{x1}, F_{y1}\right)}{P\left(F_{x1}\right)} \qquad (26)$$
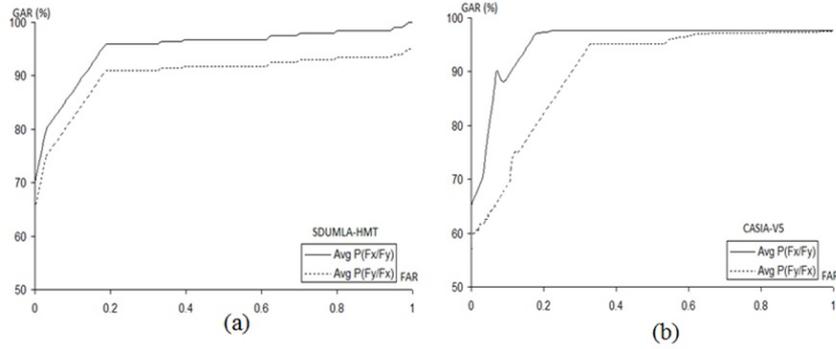
FIGURE 10. ROC curves showing the average values for face samples given fingerprint instances and for fingerprint instances given face samples on the (a) SDUMLA-HMT and (b) CASIA-V5 datasets, respectively.
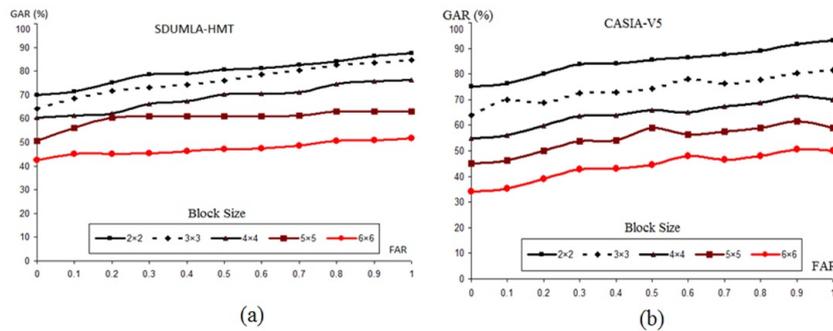

FIGURE 11. ROC curves demonstrating the performance of occluded face images with different block sizes (2×2, 3×3, 4×4, 5×5, 6×6) on the (a) SDUMLA-HMT and (b) CASIA-V5 datasets, respectively.

The GAR with respect to the FAR variations is obtained using Equation 27:

$$GAR(\%) = \left[1 - \left(\frac{P_{concatenated}(\ F_{x1}, F_{y1})}{P\ (F_{x1})} \right) \right] \times 100 \qquad (27)$$

where $P_{concatenated}(\ F_{x1}, F_{y1})$ can be determined as the probability of mismatching the concatenated face and fingerprint representing the authorised subjects, and $P\ (F_{x1})$ is the probability of mismatching the face sample for an authorised person. For simplicity, we compute the average of the GAR for the fifteen concatenated features based on the two scenarios presented, as shown in Figure 10. To evaluate the ability of our proposed system to predict missing and/or contaminating features resulting from occlusion of face images, we utilised 50 occluded face images for 50 subjects from SDUMLA-HMT and CASIA-V5, respectively. GAR is affected by the block size used to occlude the face image; it decreases as block size increases, as shown in Figure 11. It can be seen that our system achieved satisfactory results in its prediction and handling of the occlusion problem based on the EM algorithm for the tested occluded face images. What about fused face and fingerprint or, in other words, the probability of obtaining fingerprints given the occluded face image? In multimodal biometrics, the overall performance is enhanced. Thus, we attempt to prove that using the concatenated face and fingerprint. As shown in Figure 12, there is a slight improvement in GAR when obtaining the fingerprint instances given face-occluded samples.

Table 4 compares the proposed system with well-known statistical approaches: SVM, linear discriminant analysis (LDA), PCA, combined learning vector quantisation (CLVQ),
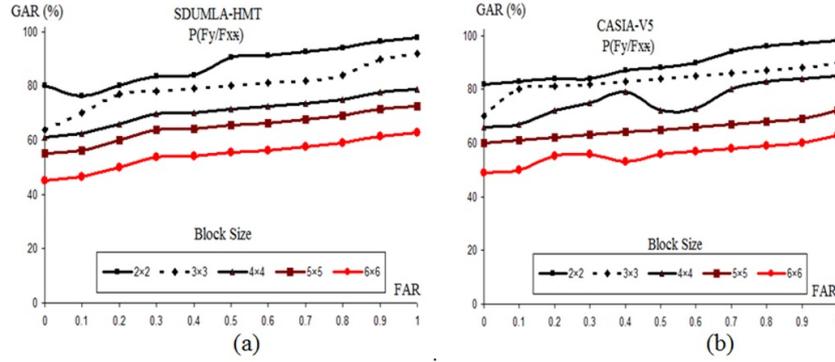
FIGURE 12. ROC curves showing the performance of concatenated face fingerprint given occluded face images with different block sizes (2×2, 3×3, 4×4, 5×5, 6×6) on the (a) SDUMLA-HMT and (b) CASIA-V5 datasets, respectively.

TABLE 4. Comparative evaluation of the proposed system based on face and fingerprint modalities on the SDUMLA-HMT and CASIA-V5 datasets with different fusion strategies.

Approaches	Fusion Strategy	SDUMLA-HMT			CASIA-V5		
		Face	Fingerprint	Fused	Face	Fingerprint	Fused
CLVQ	WS	87.58	89.86	94.72	90.34	88.64	94.01
	WP	86.25	88.65	93.02	89.54	87.21	93.54
	BC	87.35	87.78	92.54	88.00	87.06	93.40
CRBF	WS	86.24	87.69	90.25	88.65	88.21	90.24
	WP	84.30	86.56	89.32	85.01	84.35	87.12
	BC	85.99	85.07	89.64	87.32	83.99	88.87
SVM	WS	90.25	91.65	93.64	91.25	88.65	93.54
	WP	89.21	89.10	93.08	89.12	87.05	92.16
	BC	90.15	89.24	93.52	90.69	88.61	92.98
LDA	WS	88.54	89.34	90.45	84.49	86.54	89.94
	WP	87.06	88.24	90.25	84.36	86.17	89.25
	BC	88.00	88.23	90.00	84.09	86.27	89.57
PCA	WS	85.64	87.67	92.45	81.34	85.54	88.28
	WP	84.24	86.25	91.64	80.34	84.64	87.64
	BC	85.00	86.69	92.00	80.95	84.65	88.00
ADLVQ (Proposed)	WS	92.96	94.65	96.76	92.35	89.22	95.17
	WP	91.89	93.56	95.87	91.65	89.12	95.01
	BC	92.00	94.00	96.00	92.20	89.00	95.02

and combined radial basis function (CRBF)a classical neural network approach. We tested the proposed system using different fusion strategies on the confidence level by calculating the weighted sum (WS), weighted product (WP), and Borda count (BC) for each resulting matching scores. The comparison was performed using the SDUMLA-HMT and CASIA-V5 datasets and the same protocols presented in this paper. The efficiency of the different parts of the proposed system in both training and testing are shown in Table 5.

5. **Conclusions and Future work.** Classical neural networks are currently suffering from multiple problems, such as big data causing memory problems and information

TABLE 5. Efficiency of different parts of the proposed system in training and testing phases

Efficiency	Training	Testing
Enrollment	99.05	99.15
Preprocessing	98.24	98.38
Feature extraction	96.54	96.98
ADLVQ classification	95.15	95.17

retrieval problems. In this paper, we proposed an adaptive deep learning vector quantisation (ADLVQ)-based system to overcome these problems. The proposed system uses not only multimodal biometrics to fuse both face and fingerprints but also multi-sample face images and multi-instances fingerprints. With deep neural networks (DNNs) being considered as one of the most crucial steps to solving big data problems, we leveraged prior knowledge of the learning vector quantisation algorithm to handle the memory and overfitting problems. Thus, we applied the K-means algorithm for vector quantisation and clustered the input features of DNNs. We also utilised a stacked RBM to formulate a DNN by which the vectors are quantised using Kmeans to generate the codewords. The successive vectors are often not independent, thus we used adaptive vector quantiser in order to match the observed features of the input templates learned by ADLVQ classifier. We further utilised KNN to classify BoW, the output of the DNN. The experimental results obtained indicate that the proposed system selects the input features according to enrolled users based on serial and parallel structures. We also utilised sparse representation of the occluded face images to prove the ability of the proposed system to predict the missing features in the form of block images of various sizes. The empirical results obtained indicate that the system is able to handle the occlusion problem using the EM algorithm. In future work, we plan to use disguised face image datasets based on both thermal and video images. We will also address the problem of a significant decline in the results for occlude fingerprint images by using convolutional DNNs. The primary goal of our proposed system is to retrieve the information adopted by the enrolled user. The proposed system sequentially selects one biometric trait and randomly concatenates another trait to counter spoof attacks. As a result of the availability of raw materials to use in cases of counterfeiting, the user may have as many as fifteen concatenated features. Thus, we propose to add liveness parameters to the system in future work.

REFERENCES

[1] A. K. Jain, A. Ross, S. Prabhakar, An introduction to biometric recognition, *IEEE Trans Circuits Syst Video Technol*, vol.14.1, pp. 4-20, 2004.

[2] A. Gyaourova, A. Ross, A coding scheme for indexing multimodal biometric databases, I*n 2009 IEEE Conf Comput Vis Pattern Recognit Workshops*, pp. 93-98, 2009.

[3] D. Wang, A.K. Jain, Face retriever: Pre-filtering the gallery via deep neural net, *In 2015 International Conference on Biometrics (ICB)*, pp. 473-480, 2015.

[4] A. Jain, A.A. Ross, K. Nandakumar, Introduction to Biometrics, *Springer Science Business Media*, 2011.

[5] A. Jain, K. Nandakumar, A. Ross, Score normalization in multimodal biometric systems, *Pattern Recogn*, vol. 38, no .12, pp. 2270-2285,2005.

[6] Y. Bengio, Learning deep architectures for AI, *Foundations and Trends in Machine Learning*, vol. 2, no.1, pp. 1-127, 2009.

[7] D. Erhan, Y. Bengio, A. Courville, P. Manzagol, P. Vincent, S. Bengio, Why does unsupervised pre-training help deep learning? *J Mach Learn Res*, vol. 11, pp. 625-660,2010.

[8] Y. LeCun, M. Ranzato, Deep learning tutorial, *In Tutorials in International Conference on Machine Learning* (ICML13), 2013.

[9] M. Tanaka, M. Okutomi, A novel inference of a restricted Boltzmann machine, In ICPR, 2014, pp. 1526-1531.

[10] X. Y. Chen, X. Y. Peng, Y. Peng, and J. B. Li, The classification of synthetic aperture radar image target based on deep learning, *Journal of Information Hiding and Multimedia Signal Processing*, vol. 7, no .6 , pp. 1345-1353, 2016.

[11] H. Lee, Deep Learning Tutorial, people.tamu.edu, Deep Learning Tutorial Hungyi Lee.pdf (accessed 18/12/2016).

[12] J. Bruna, S. Mallat, Invariant scattering convolution networks, *IEEE Trans Pattern Anal Mach Intell*, vol. 35, no. 8 , pp. 1872-1886, 2013.

[13] M. Carvalho, M. Cord, S. Avila, N. Thome, E. Valle, Deep neural networks under stress, *arXiv preprint arXiv*, pp.1605.03498 (2016).

[14] Y. Cao, M. Long, J. Wang, H. Zhu, Q. Wen, Deep quantization network for efficient image retrieval, *In AAAI Conference on Artificial Intelligence*, 2016.

[15] D. Wu, L. Pigou, P.J. Kindermans, N. LE, L. Shao, J. Dambre, J.M. Odobez, Deep dynamic neural networks for multimodal gesture segmentation and recognition, *IEEE Trans Pattern Anal Mach Intell* (2016).

[16] J. Ngiam, A. Khosla, M. Kim, J. Nam, H. Lee, A.Y. Ng, Multimodal deep learning, *In Proceedings of the 28th International Conference on Machine Learning* (ICML-11), pp. 689-696, 2011.

[17] C. Ding, D. Tao, Robust face recognition via multimodal deep face representation, *IEEE T Multi-media*, vol. 17, no.11 , pp. 2049-2058, 2015.

[18] D. Menotti, G. Chiachia, A. Pinto, W.R. Schwartz, H. Pedrini, A.X. Falcao, A. Rocha, Deep representations for iris, face, and fingerprint spoofing detection, IEEE T Inf Foren Sec, vol. 10, no.4 , pp. 864-879, 2015.

[19] A. Lumini, L. Nanni, Overview of the combination of biometric matchers, *Inform Fusion*, vol. 33 , pp. 71-85, 2017.

[20] C. Lin, B. Wang, X. Fan, Y. Ma, H. Liu, Orthogonal enhanced linear discriminant analysis for face recognition, *IET Biometrics*, vol. 5, no.2 , pp. 100-110, 2016.

[21] H. Li, C.Y. Suen, Robust face recognition based on dynamic rank representation, *Pattern Recogn*, vol. 60 , pp. 13-24,2016.

[22] E. Learned-Miller, G.B. Huang, A. RoyChowdhury, H. Li, G. Hua, Labeled faces in the wild, pp. A survey, *In Advances in Face Detection and Facial Image Analysis, Springer International Publishing*, pp. 189-248, 2016.

[23] M. Y. Shams, A. S. Tolba, S. H. Sarhan, A Vision System for Multi-View Face Recognition, Int. *J. of Circuits Systems and Signal Processing*, vol. 10.1 pp. 455-461, 2016.

[24] A. Gyaourova, A. Ross, A novel coding scheme for indexing fingerprint patterns, In Joint IAPR *International Workshops on Statistical Techniques in Pattern Recognition (SPR) and Structural and Syntactic Pattern Recognition (SSPR)*, Springer Berlin Heidelberg, pp. 755-764, 2008.

[25] L. Nanni, A. Lumini, M. Ferrara, R. Cappelli, Combining biometric matchers by means of machine learning and statistical approaches, *Neurocomputing*, vol. 149 , pp. 526-535, 2015.

[26] D. Peralta, M. Galar, I. Triguero, D. Paternain, S. Garca, E. Barrenechea, J.M. Bentez, H. Bustince, F. Herrera, A survey on fingerprint minutiae-based local matching for verification and identification: Taxonomy and experimental evaluation, *Inform Sciences*, vol. 315 , pp. 67-87, 2015.

[27] P.A. Kumari, G. JayaSuma, A comparative study of various minutiae extraction methods for fingerprint recognition based on score level fusion, *In Application of Computational Intelligence to Biology, Springer Singapore*, pp. 29-41, 2016.

[28] R. Kumar, P. Chandra, M. Hanmandlu, A robust fingerprint matching system using orientation features, *J Inf Proc Syst*, vol. 12, no.1 , pp. 83-99,2016.

[29] A. Jain, MVNK Prasad, A novel fingerprint indexing scheme using dynamic clustering, *J Reliab Intell Environ*, , pp. 1-13,2016.

[30] Y. Su, J. Feng, J. Zhou, Fingerprint indexing with pose constraint, *Pattern Recogn*, vol. 54 , pp. 1-13, 2016.

[31] A. Sankaran, A. Jain, T. Vashisth, M. Vatsa, R. Singh, Adaptive latent fingerprint segmentation using feature selection and random decision forest classification, I*nform Fusion*, vol.34 , pp. 1-15, 2017.

[32] S. H. Sarhan, S. Alhassan, S. Elmougy, Multimodal biometric systems: A comparative study, Arab J Sci Eng, , pp. 1-15, 2016.

[33] A.H. El-Baz, A.S. Tolba, S.K. Pal, Robust boosted parameter based combined classifier for rotation invariant texture classification, *Appl Artif Intell*,vol. 30, no.2 , pp. 77-96, 2016.

[34] H.M. Raafat, A.S. Tolba, A.M. Aly, A novel training weighted ensemble (TWE) with application to face recognition, *Appl Soft Comput*, vol. 11, no.4 , pp. 3608-3617,2011.

[35] M.Y. Shams, A.S. Tolba, H. Shahenda, Face, iris, and fingerprint multimodal identification system based on local binary pattern with variance histogram and combined learning vector quantization, *J Theor Appl Inf Technol*, vol. 89, no.1 , pp. 53-70,2016.

[36] G.L. Marcialis, F. Roli, L. Didaci, Personal identity verification by serial fusion of fingerprint and face matchers, *Pattern Recogn*, vol. 42, no.11 , pp. 2807-2817,2009.

[37] B. Biggio, Z. Akhtar, G. Fumera, G.L. Marcialis, F. Roli, Security evaluation of biometric authentication systems under real spoofing attacks, *IET Biometrics*, vol. 1.1 , pp. 11-24,2012.

[38] N. Poh, A. Ross, W. Lee, J. Kittler, A user-specific and selective multimodal biometric fusion strategy by ranking subjects, *Pattern Recogn*, vol. 46, no.12 , pp. 3341-3357,2013.

[39] S. Shekhar, V.M. Patel, N.M. Nasrabadi, R. Chellappa, Joint sparse representation for robust multimodal biometrics recognition, *IEEE Trans Pattern Anal Mach Intell*, vol. 36.1 , pp. 113-126,2014.

[40] S. Bharadwaj, H.S. Bhatt, R. Singh, M. Vatsa, A. Noore, QFuse: Online learning framework for adaptive biometric system, *Pattern Recogn*, vol. 48, no.11 , pp. 3428-3439, 2015.

[41] K. Nguyen, S. Denman, S. Sridharan, C. Fookes, Score-level multibiometric fusion based on DempsterShafer theory incorporating uncertainty factors, *IEEE Trans Hum Mach Syst*, vol. 45, no.1 , pp. 132-140,2015.

[42] Y. Kessentini, T. Burger, T. Paquet, A DempsterShafer Theory based combination of handwriting recognition systems with multiple rejection strategies, Pattern Recogn, vol. 48, no.2 , pp. 534-544,2015.

[43] P. Wild, P. Radu, L. Chen, J. Ferryman, Robust multimodal face and fingerprint fusion in the presence of spoofing attacks, *Pattern Recogn*, vol. 50 , pp. 17-25,2016.

[44] Y. Chen, J. Yang, C. Wang, N. Liu, Multimodal biometrics recognition based on local fusion visual features and variational Bayesian extreme learning machine, *Expert Syst Appl*, 64 , pp. 93-103, 2016.

[45] B. Jun, D. Kim, Robust face detection using local gradient patterns and evidence accumulation, Pattern Recogn, vol. 45, no.9 , pp. 3304-3316, 2012.

[46] Y. Xie, J. Yang, X. Zhao, X. Zhang, Finger vein recognition based on local opposite directional pattern, *In Chinese Conference on Biometric Recognition,* Springer International Publishing, pp. 297-304, 2015.

[47] A. Dahmouni, K. Moutaouakil, K. Satori, Robust face recognition using local gradient probabilistic pattern (LGPP), *In Proceedings of the Mediterranean Conference on Information Communication Technologies*, 2015, Springer International Publishing, pp. 277-286, 2016.

[48] Q. Li, W. Lin, Y. Fang, No-reference quality assessment for multiply-distorted images in gradient domain, *IEEE Signal Process Lett*, vol. 23, no.4 , pp. 541-545(2016).

[49] Z. Guo, L. Zhang, D. Zhang, Rotation invariant texture classification using LBP variance (LBPV) with global matching, Pattern Recogn, vol. 43, no.3 , pp. 706-719(2010).

[50] Z. Huang, C. Weng, K. Li, Y. Cheng, C. Lee, Deep learning vector quantization for acoustic information retrieval, In 2014 IEEE International Conference on Acoustics, *Speech and Signal Processing (ICASSP)*, pp. 1350-1354, 2014.

[51] S.J.D. Prince, Computer Vision: Models, Learning, and Inference, Cambridge University Press, 2012.

[52] A.K. Jain, Data clustering: 50 years beyond K-means, *Pattern recognition letters*, vol. 31, no.8, pp. 651-666,2010.

[53] S. Tran, Representation decomposition for knowledge extraction and sharing using restricted Boltzmann machines, *PhD diss., City University London*, 2016.

[54] G.E. Hinton, A practical guide to training restricted Boltzmann machines, *In Neural Networks: Tricks of the Trade,* Springer Berlin Heidelberg, pp. 599-619, 2012.

[55] B. Zhang, S.N. Srihari, Fast k-nearest neighbor classification using cluster-based trees, *IEEE Trans Pattern Anal Mach Intell*, vol. 26, no.4 (2004), pp. 525-528.

[56] Y. Yin, L. Liu, X. Sun, SDUMLA-HMT: A multimodal biometric database, *In Chinese Conference on Biometric Recognition, Springer Berlin Heidelberg*, pp. 260-268, 2011.

[57] CASIA-V5 face, and fingerprint images, http://www.idealtest.org, 2015 (accessed 16.09.06).