

## Research Article

# Multipose Face Recognition-Based Combined Adaptive Deep Learning Vector Quantization

Shahenda Sarhan <sup>1,2</sup>, Aida A. Nasr,<sup>3</sup> and Mahmoud Y. Shams <sup>3</sup>

<sup>1</sup>Faculty of Computing and Information Technology, King Abdulaziz University, Jeddah, Saudi Arabia

<sup>2</sup>Faculty of Computers and Information Sciences, Mansoura University, Mansoura, Egypt

<sup>3</sup>Faculty of Artificial Intelligence, Kafrelsheikh University, Kafr El-Sheikh, Egypt

Correspondence should be addressed to Shahenda Sarhan; shahenda\_sarhan@yahoo.com

Received 17 March 2020; Revised 9 September 2020; Accepted 15 September 2020; Published 24 September 2020

Academic Editor: Anastasios D. Doulamis

Copyright © 2020 Shahenda Sarhan et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Multipose face recognition system is one of the recent challenges faced by the researchers interested in security applications. Different researches have been introduced discussing the accuracy improvement of multipose face recognition through enhancing the face detector as Viola-Jones, Real Adaboost, and Cascade Object Detector while others concentrated on the recognition systems as support vector machine and deep convolution neural networks. In this paper, a combined adaptive deep learning vector quantization (CADLVQ) classifier is proposed. The proposed classifier has boosted the weakness of the adaptive deep learning vector quantization classifiers through using the majority voting algorithm with the speeded up robust feature extractor. Experimental results indicate that, the proposed classifier provided promising results in terms of sensitivity, specificity, precision, and accuracy compared to recent approaches in deep learning, statistical, and classical neural networks. Finally, the comparison is empirically performed using confusion matrix to ensure the reliability and robustness of the proposed system compared to the state-of-art.

## 1. Introduction

Recently, face recognition has become one of the most discussed topics in the computer vision field, although it is not new but with the revolution in electronic devices as mobile phones researchers have realized that it can be a solution to different issues. Many studies have claimed that they achieved a very high recognition rate of facial images approaching the human recognition rate. This may be true in clear frontal facial images, not noisy or occluded [1, 2]. On the other hand, multipose face recognition with various poses and different head angles will be difficult where multiple snapshots at different time instances should be captured. Multipose face recognition has been discussed lately in different studies where some concentrated on face reconstruction from different head poses using 3D Models [3–5]. Others discussed misalignment, varying illumination, and noise [6–9], while the rest concentrated on achieving

higher recognition rate using different techniques as support vector machine [10, 11], Local binary patterns [12], and deep neural networks [13–16].

Restricted Boltzmann Machine (RBM) is a generative stochastic unsupervised artificial neural network that can learn a probability distribution over its set of inputs [17–19]. RBM models use stochastic approach instead of deterministic by which the RBM model possesses inherent randomness of a set of parameters values, which leads an ensemble of different outputs. Therefore, in order to make the learning easier, we have restricted the randomness of the parameters values through using RBM as a deep learner and quantizing the vectors based on  $k$ -means algorithm. In this paper, we introduce a combined adaptive deep learning vector quantization classifier (CADLVQ) with different parameters sets, in which we seek for changing the entire network parameters of adaptive deep learning vector quantization (ADLVQ) to increase the recognition accuracy

of multipose face images [19, 20] and to boost the instability and reliability issues of ADLVQ suffering from varied results obtained with each run time. The general schematic diagram of the CADLVQ classifier with different parameters sets is shown in Figure 1.

As clarified in Figure 1, the number of CADLVQ classifiers depends on different parameters sets producing the output classes. And to manage that, we have used the majority voting algorithm to determine the winner class by which the recognition of the system is realized. The contribution of this paper can be summarized as follows:

- (1) Proposing a combined adaptive deep learning vector quantization classifier to boost the weakness of the ADLVQ classifier
- (2) Enhancing the stability and reliability of the CADLVQ based multipose face images using different parameters sets
- (3) Comparing the proposed CADLVQ with the most recent deep learning approaches using matching scores, weighted sum, weighted product, and majority voting
- (4) Handling colluded face images with different block sizes by predicting the missing features using expectation maximization (EM) algorithm

This paper is organized as follows. Section 2 covers a brief review of face recognition techniques. In Section 3, the proposed combined adaptive deep learning vector quantization is described in details. Section 4 introduces simulation and experiments while section 5 comprises results and discussion. Finally Section 6 outlines main conclusions and presents some future work.

## 2. Related Work

*2.1. Adaptive Deep Learning Approaches.* Classical neural networks (NNs) are commonly used to classify the input patterns and objects based on activation function and the weighted summation of the inputs to produce the estimated output values [21]. Online retainable neural network algorithm to retrain the procedure parameter of the network for estimating the optimal selection of the training input and determining the maximum a posteriori as Markov random field presented by Doulamis et al. [22]. To boost weak classifiers in two-classes classification problem, AdaBoost learning algorithm presented by Hastie et. al., [23] has been successfully producing accurate classifiers based on multi-class exponential loss function and forward stage-wise additive modeling.

On the other hand, many researchers investigated using deep learning in face detection as in [24] a detailed survey of deep learning techniques is illustrated which is commonly classified to deep Boltzmann machine (DBM), deep belief neural network (DBN), auto encoder (AE), and Deep convolutional neural networks (DCNN). In [25], the authors studied the effect of sharing of convolutional layers between the region proposal network (RPN) and Fast region-based convolutional neural network (R-CNN) detector module

using two datasets. Hao et. al., in [26] used CNN as a base to handle faces of diverse scales through proposing a scale-aware face detector (SAFD) which automatically normalizes face sizes prior to detection. And in [27] Zhu et al. introduced contextual multiscale region-based convolution neural network (CMS-RCNN) to robustly detect human faces and solve occlusions, low resolutions, facial expressions, and illumination variations problems. They tested their proposed system using two datasets containing images collected under various challenging conditions.

Adaptive deep learning fusion paradigm presented by Doulamis and Doulamis [28] can be utilized for tracking objects in stable manner having few labeled data to be trained based on multilayered deep structures and the paradigm dynamically updates the new unlabeled data to adjust the performance of the deep structures.

One of the major problems used in deep learning especially deep convolution neural network is the data compressing for handling the storage of the model. Therefore Gong et. al. [29] utilized k-means clustering approach for vector quantization, which produces a good balance between the model size and the recognition accuracy.

In [30], Doulamis and Voulodimos proposed a fast adaptive learning algorithm, named as FAST-MDL, for detecting humans in complex scenes through dynamically updating the parameters of a multilayered deep learning structure. The results indicated that the adaptation of the FAST-MDL was 5 times faster.

For visual domain localization, Angeletti et al. [31] presented local adaptive (LoAd) deep visual learning by which discriminative information about the training deep models are differentiated and stored in the inner layer of the network. Moreover, an adaptive deep learning model to extract both facial expressions and human actions is presented by Kim and Rhee in [32] to produce human synthetic emotion. Therefore, for autonomous vehicles and surveillance systems, an adaptive deep learning paradigm is presented by Tahboub et. al. [33] to detect the pedestrian of the input video sequences to estimate the detections in the presence of video compression to handle data-rate problem. Doulamis and Doulamis [34] presented an adaptive deep learning architecture based on stacked auto-encoder for fall detection by estimating the trained samples including the humans and background objects and the weights are updated in case of large changes.

Mughees et. al. [35] proposed an algorithm based on spatial updated deep belief network (SDBN) for hyperspectral image (HSI) classification by which weighted hyper-segmentation for regions is obtained to produce spatial features to be adapted to the actual HSI spatial structures.

*2.2. Multiview Face Recognition.* The improvements of face recognition algorithms are still under research and development. The researchers' efforts are speeded and sustained recently according to the different applications that required more reliable and precise face recognition system.

In [36], both heterogeneous and homogeneous face recognition approaches were investigated, through proposing two

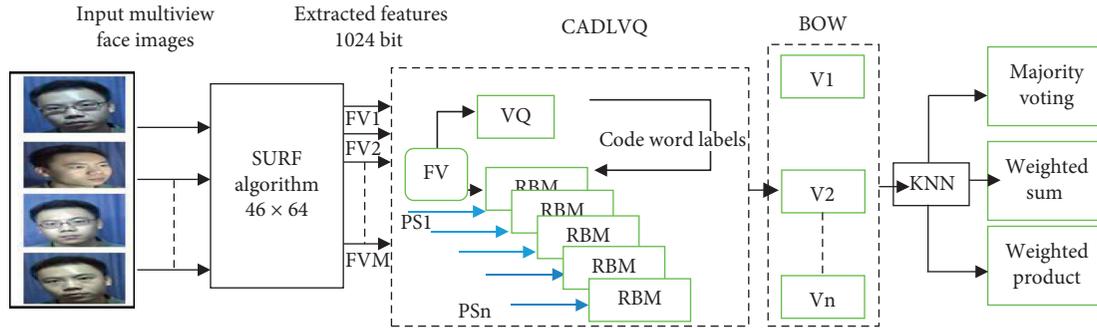


FIGURE 1: Schematic diagram of the CADLVQ classifier with  $N$  parameter sets.

methods, a context-aware local binary feature learning (CALBFL) and a context-aware local binary multi-scale feature learning (CALBMFL). The CALBFL is used for exploiting the contextual information of adjacent bits, by constraining the number of shifts from different binary bits, while the CALBMFL method is used to jointly learn multiple projection matrices for face representation. Hessian multiset canonical correlations analysis was presented in [37] to tackle this disadvantage of traditional canonical correlations analysis by reducing the extracted multipose features of face images. In [38], a hybrid approach of the kernelized support vector data description (SVDD) and a binary hierarchical decision tree was proposed. The SVDD was used to customize the needed space in large scale face recognition systems through using only the samples that formalize the image set boundaries, while the decision tree was used for enhancing the classification speed. Experiments were conducted using a new database of 285 person prepared by the authors, indicating that the proposed approach has reduced the needed disk space with better classification testing time and without affecting the accuracy.

Many attempts to detect, recognize, identify, and solve different multipose face recognition problems using deep neural networks were presented as in [39], the authors tried to solve the imposter authentication, by using multiple views of a person's face images at different angles achieving an error rate equal to 0.022. While in [19] authentication through proposing a technique based on local gradient pattern with variance (LGPV) and a deep neural network classifier called adaptive deep learning vector quantization was discussed. The LGPV was used to extract the features of the input modalities that are dynamically enrolled in the system, to be classified using deep neural networks (DNN), after quantization using the K-means algorithm based on prior learning vector quantization (LVQ) knowledge.

In [20], we tried to reduce the additive white Gaussian noise effect by proposing a system to detect multipose face images based on Speeded Up Robust Features (SURF), Multi-Layer Perceptron (MLP), and a combined classifier of Learning Vector Quantization (LVQ) and Radial Basis Function (RBF) classifiers. Also in [40], the authors tried to reduce noise through proposing a two stages method, consisting of a data cleaning from noise model and multipose deep convolution neural networks (DCNN) classifier. In [8], convolutional neural networks were also used to improve the region of interest ROI through pooling multiple

convolution face images. Two deep learning models, Lightened CNN and VGG-Face, have been used in [41] to solve different face recognition problems as lower and upper face occlusions and misalignment. Experiments results indicated that deep learning can tolerate localizations error of the intraocular distance.

Finally, a couple-agent pose guided generative adversarial network (CAPG-GAN) is proposed in [42] to synthesize face images with different poses, including extreme profile views. Also in [43], extreme pose variations were investigated through proposing a pose-aware convolutional neural network model to handle extreme out-of-plane pose variations of unconstrained face recognition. 3D rendering was used to adaptively model the appearance of face in different poses. Evaluation was conducted using IARPA Janus Benchmark A (IJB-A) and People-In-Photo-Albums (PIPA) datasets, indicating that the proposed model has emulated the accuracy of popular methods that are fine-tuned to the target dataset. In [44], a joint multipose convolutional network to handle landmarks for semi-frontal and profile faces pose variations in-the-wild is proposed. Experiments were conducted using standard static image datasets as IBUG, 300W, COFW, and the latest Menpo Benchmark indicating that the proposed technique achieved superiority over the state-of-the-art. Other papers addressing the face images with different poses could be found in [45–49].

### 3. Combined Adaptive Deep Learning Vector Quantization

As mentioned in section 1, here we introduce a system for multipose face recognition based on combined adaptive deep learning vector quantization. The proposed system is used to handle the problem of fixed parameters investigated in [19]. We are seeking to use variable parameter sets by which enhancement of the whole system will be achieved. During the running process, the results with the fixed parameters are slightly varied. By using variable parameter sets, we ensure the stability of the system results and enhancement of system accuracy. Algorithm 1 presents the steps of the proposed CADLVQ classifier.

The enrollment stage contains  $Y$  multipose facial images with different poses for  $M$  users entered to the ADLVQ classifiers. Instead of using fixed parameters value, a

CADLVQ is performed based on  $N$  parameters sets to boost the weakness of ADLVQ classifier. For features extraction of facial images, we used Speeded Up Robust Feature (SURF) extractor, as it is considered one of the most accurate algorithms for facial detection. SURF is a powerful alternative of the Scale Invariant Feature Transform (SIFT) algorithm although it is based on similar properties as SIFT, with a complexity stripped down even further. SURF uses an integer approximation of the determinant of Hessian blob detector, which can be computed with 3 integer operations using a precomputed integral image [50].

ADLVQ, RBM, and vector quantization are used to handle the overfitting problem of the enrolled features while the classification is performed as illustrated in steps from 1 to 7, clarified in Algorithm 1. Five RBMs with five parameter sets are used, where every RBM is tuned by one parameter set in parallel fashion as shown in Figure 1. Each RBM has stacked 20 hidden layers with the Sigmoid function as the activation function. The softmax layer is the final output layer for each ADLVQ classifier.

The initialization step starts with each feature vector of size 1024 from SURF, to be quantized by  $K$ -means algorithm in order not to consume memory and to prepare the vectors to pretrain using RBM. The codeword for each extracted feature is obtained by the standard vector quantization (VQ) which is used to compress the DNN parameters to improve the system efficiency as it creates a balance between model size and recognition rate, where any missing features or unobserved data undergo feedback to be adapted using vector quantization and DNN classifier.

The obtained codeword from the VQ is then utilized as the class label pretrain for each RBM with cross-entropy as the optimization objective. In order to determine the  $k$ -clustering of the unsupervised codebook from the extracted SURF features, we started with one cluster and then kept splitting clusters until the points are assigned to each cluster that have a Gaussian distribution.

#### 4. Simulation and Experiments

The CADLVQ was implemented and tested using MATLAB R2019b image processing and computer vision libraries. Experiments were conducted with 3 different views of variant size facial images from SDUMLA-HMT [51] and CASIA-V5 [52], transformed into gray scale with a fixed  $[64 \times 64]$  image size as input to SURF. SDUMLA-HMT as shown in Figure 2 is a standard database with the images of 106 subjects, 61 males, and 45 females with age between 17 and 31, where each subject is represented by 7 different views of facial images. CASIA-FaceV5 is also a standard database with 500 subjects, represented by 2500 16 bit color BMP facial images as shown in Figure 3, with images of  $640 \times 480$  resolution [53].

#### 5. Results and Discussion

As clarified from Table 1, five parameter sets (PS) values for the combined ADLVQ classifier are elected. An example of the voting strength for 10 classes entered the system is illustrated in Table 2.

According to the voting strength, the winner class resulting from each classifier is selected. The Genuine



FIGURE 2: SDUMLA-HMT database [51].



FIGURE 3: CASIA-FaceV5 database [52].

Acceptance Rate (GAR) of the tested multipose face images based on CADLVQ classifier is shown in Table 3 from which we can see that five parameter sets were used; V1 represents the frontal view of the face image and V2 and V3 represents different views of the face images. The influence of multipose face images (V1, V2, and V3) on the recognition process with SDUMLA-HMT and CASIA-V5 datasets for CADLVQ with the five parameter sets is shown in Figure 4. It is noticed that an improvement of the GAR (%) is obtained, due to the combined ADLVQ parameter sets variations.

For evaluation the sensitivity, specificity, precision, accuracy, and F1score from the confusion matrix for the proposed system have been calculated based on equations (1)–(5):

$$\text{sensitivity} = \frac{TP}{(TP + FN)}, \quad (1)$$

$$\text{specificity} = \frac{TN}{(TN + FP)}, \quad (2)$$

$$\text{precision} = \frac{TP}{(TP + FP)}, \quad (3)$$

$$\text{accuracy} = \frac{(TP + TN)}{(TP + TN + FP + FN)}, \quad (4)$$

$$F_1 \text{ score} = \frac{2TP}{(2TP + FP + FN)}, \quad (5)$$

where TP is true positive, TN is true negative, FP is false positive, and FN is false negative.

Input: The SURF extracted feature vectors (FV).

For  $M$  Users,  $Y$  view face images and  $N$  Parameter Sets (PS):

$$FV_y = \{FV_{y1}, FV_{y2}, \dots, FV_{ym}\}_{m=1}^M$$

Initialization: Initialization of PS by pretraining the DNN layer by layer with different PS, setting the weights,  $W$ , to zero; using the  $K$ -means algorithm to quantise the FV vectors

Output: The confusion matrix including Sensitivity, Specificity, Precision, Accuracy, and F1score of the matching scores produced for each class with different parameter sets (PS). A decision (Accept/Reject) based on confidence level.

Procedure:

- (1) Setting the FV of a single-layer RBM network.
- (2) FV updating based on the gradient, weight decay, and momentum.
- (3) Inference determination based on posterior probability over the hidden variable.
- (4) Determination of expectation maximization (EM) to provide sufficient information for unobserved data.
- (5) Apply the softmax of PS activation parameter vectors:
- (6) Generate the codewords by using the  $K$ -means algorithm.
- (7) Classification of Bag of Words (BoW) Using  $K$ -NN
- (8) Determination of the similarity metric between the query and stored templates using Euclidean distance with dynamic metric adaptation.
- (9) Repeat for each parameter set (PS).

ALGORITHM 1: CADLVQ.

TABLE 1: Five parameter sets (PS) for the combined ADLVQ classifier.

	Parameter sets (PS)				
	PS1	PS2	PS3	PS4	PS5
No. of iteration	100	200	300	400	500
Initial momentum	0.2	0.3	0.4	0.5	0.6
Final momentum	0.7	0.8	0.9	0.8	0.9
Weighted cost	0.001	0.002	0.003	0.004	0.005
Learning rate	0.01	0.02	0.03	0.1	0.2

One of the well-known and reliable benchmarks that summarize the classification performance is the confusion matrix, for an unequal number of observations. Since we have more than two classes, the classification accuracy alone is not sufficient for measuring the system performance. By determining the confusion matrix, a better idea of our CADLVQ classification model will be achieved.

SDUMLA-HMT and CASIA-V5 are used to calculate a confusion matrix with expected outcome values. A comparative evaluation results between the combined learning vector quantization (CLVQ) [54], ADLVQ [19], and CADLVQ are shown in Table 4 in both training and testing phases for SDUMLA-HMT and CASIA-V5. We utilized 318 multipose face images for training and 106 for testing out of SDUMLA-HMT dataset. For CASIA-V5 dataset, we used 2500 multipose face images, 1500 for training and 1000 for testing. The results show the superiority of the proposed system when compared with CLVQ and ADLVQ.

An example for calculating the confusion matrix factors is shown in Figure 5. In this example, we used SDUMLA-HMT dataset applied to CLVQ in the training phase by using 318 face images and found that  $TP = 285$ ,  $TN = 12$ ,  $FP = 1$ , and  $FN = 20$ . The training and testing calculations of the confusion matrix of the CLVQ, ADLVQ, and CADLVQ systems based on SDUMLA-HMT and CASIA-V5 datasets are shown in Figure 6.

A comparative evaluation based on the accuracy of CLVQ [54], ADLVQ [19], and the proposed CADLVQ, compared to Support Vector Machine (SVM), Linear Discriminant Analysis (LDA), Principal Component Analysis (PCA), as statistical approach, Multi-Layer Perceptron (MLP), Combined Radial Basis Function (CRBF), as neural network approach, Deep Restricted Boltzmann Machine (DRBM), Deep Belief Neural Nets (DBNN), and Deep Convolutional Neural Network (DCNN) is investigated in Figure 7. The results show that the proposed CADLVQ achieves higher accuracy compared to other approaches.

Figures 8 and 9, present the resulting 100 samples of multipose facial images obtained respectively from both SDUMLA-HMT and CASIA-V5. While in Table 5, the weighted sum, weight product, and the majority voting results of the CADLVQ are presented, clarifying that a higher accuracy is obtained in case of majority voting than the obtained with the weighted sum, and weighted product, respectively.

A final experiment has been conducted for proving the proposed CADLVQ classifier ability to predict and handle the missing features results from occluded facial images. We empirically tested the proposed system using 40 occluded facial images out of SDUMLA-HMT and CASIA-V5 datasets, respectively. These 40 tested facial images are occluded using different block size images as shown in Figure 10. The recognition accuracy of the 40 occluded facial images with different block sizes is shown in Table 6.

From the above results, we find that restricting the randomness of RBM parameters through using RBM as a deep learner while quantizing vectors based on  $k$ -means algorithm has enhanced the recognition accuracy of the RBM. The obtained results were evaluated using large number of facial images with three different views obtained from two different datasets. As is known, the RBM is a generative unsupervised approach that can learn a probability distribution over its set of inputs using stochastic

TABLE 2: Voting strength for 10 classes of CADLVQ classifier.

CADLVQ classifier no. [PS]	Voting strength for 10 classes									
	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10
ADLVQ-1[PS1]	1.0	0.9	0.6	0.9	1.0	0.9	0.9	1.0	0.8	0.6
ADLVQ-2[PS2]	0.7	0.6	1.0	1.0	0.7	1.0	0.7	0.6	0.7	0.7
ADLVQ-3[PS3]	0.7	0.8	0.8	1.0	0.8	0.8	0.9	0.9	0.8	0.7
ADLVQ-4[PS4]	0.9	0.6	0.7	0.9	0.6	0.9	0.7	0.7	0.7	0.7
ADLVQ-5[PS5]	1.0	0.8	0.7	0.8	0.6	0.9	0.9	0.6	1.0	0.8

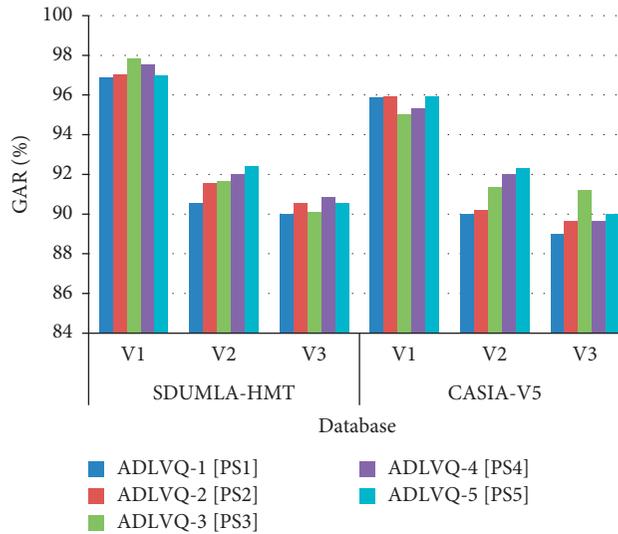


FIGURE 4: Multipose face recognition based on CADLVQ based on SDUMLA-HMT and CASIA-V5 datasets.

TABLE 3: The GAR (%) of the frontal view (V1), and multipose (V2, V3) for combined ADLVQ classifier.

Multipose face images (V1, V2, V3)	SDUMLA-HMT			CASIA-V5		
	V1	V2	V3	V1	V2	V3
ADLVQ-1[PS1]	96.87	90.56	90.00	95.87	90.00	89.00
ADLVQ-2[PS2]	97.05	91.54	90.56	95.92	90.20	89.64
ADLVQ-3[PS3]	<b>97.85</b>	91.65	90.12	95.00	91.38	<b>91.21</b>
ADLVQ-4[PS4]	97.56	92.00	<b>90.87</b>	95.35	92.00	89.66
ADLVQ-5[PS5]	97.00	<b>92.40</b>	90.54	95.95	<b>92.30</b>	90.01

V1 (Pose1): frontal view of face image, V2 and V3 (Pose2 and 3) multipose face images.

approach instead of deterministic. This stochastic base results in an ensemble of different outputs, so the restriction of RBM parameters randomness results in more efficient training and testing which is clear in enhancing the recognition accuracy of the proposed classifier compared to the state of art using both SDUMLA-HMT and CASIA-V5 datasets as in Figures 6 and 7.

The proposed algorithm clarified in Algorithm 1 complexity is calculated as  $O(M \times Y \times N)$ , based on three factors the number of enrolled users  $M$ , number of the multipose facial images of each user  $Y$  and finally the number of the Parameter Sets  $N$ . Figure 11 illustrates the variation in consumed memory in GBs versus the time in seconds, with respect to the trained multipose images in SDUMLA-HMT, and CASIA-V5 datasets, respectively.

The proposed system accommodated 2 GB in about 30 minutes.

Finally, Table 7 is describing the main advantages/limitations of the proposed method, called CADLVQ and the ADLVQ approach for again clarity of presentation.

As investigated in the table, the main advantage is the obtained accuracy which shows the ability of the proposed CADLVQ to recognize multiview face images. Furthermore, the ability of the proposed to be adapted with different input hyperparameter values as stochastic process to boost weak classifiers in procedure stage. Otherwise, the main limitation is the complexity of the CADLVQ as different training should be executed and we do not think that is not a great problem in the existence of GPU environment and available materials.

	Prediction positive	Prediction negative	
Positive	285	20	Sensitivity = $\frac{285}{285 + 20} \times 100 = 93.44\%$
Negative	1	12	Specificity = $\frac{12}{12 + 1} \times 100 = 92.31\%$
	Precision = $\frac{285}{285 + 1} \times 100 = 99.65\%$	Fscore = $F_1 = \frac{2 \times 285}{2 \times 285 + 1 + 20} = 96.45\%$	ACC = $\frac{285 + 12}{285 + 12 + 20 + 1} \times 100 = 93.40\%$

FIGURE 5: An example for calculating the confusion matrix for the CLVQ based on SDUMLA-HMT dataset in the training phase.

Datasets	Proposed approach	Training		Testing	
SDUMLA-HMT	CLVQ	285	20	95	5
	ADLVQ	295	15	101	4
	CADLVQ	300	7	101	1
CASIA-V5	CLVQ	1200	170	920	50
	ADLVQ	1400	68	950	25
	CADLVQ	1420	58	972	20

FIGURE 6: The confusion matrix of CLVQ, ADLVQ, and CADLVQ in training and testing.

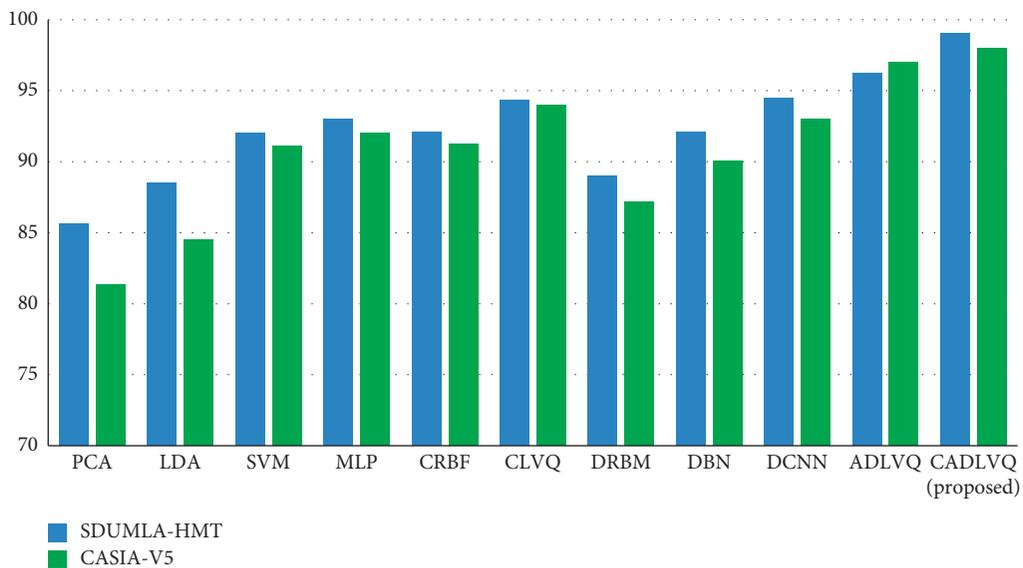


FIGURE 7: Recognition accuracy.

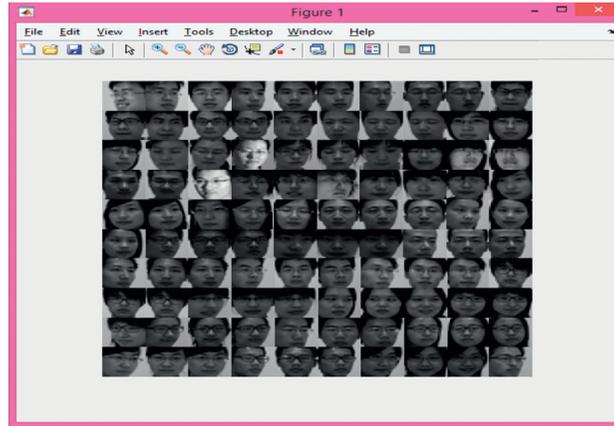


FIGURE 8: SDUMLA-HMT 100 samples class out of 106 subjects.

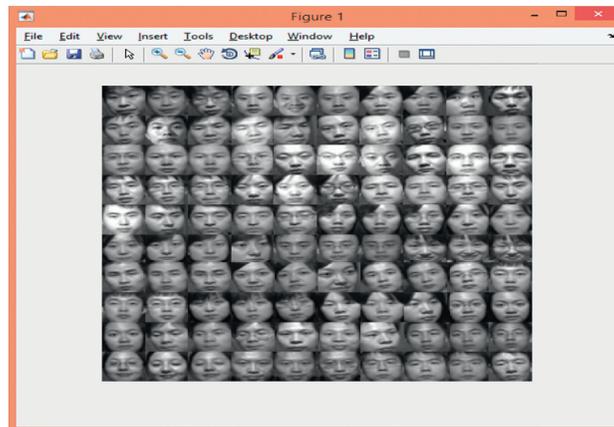


FIGURE 9: CASIA-V5 100 samples class out of 500 subjects.

TABLE 4: The confusion matrix for the proposed CLVQ, ADLVQ, and CADLVQ based on SDUMLA-HMT and CASIA-V5 datasets in the training and testing phases.

DB = SDUMLA-HMT						
	Training (318)			Testing (106)		
	CLVQ	ADLVQ	CADLVQ	CLVQ	ADLVQ	CADLVQ
Sensitivity	93.44	95.16	97.72	95.00	96.19	99.02
Specificity	92.31	87.50	90.91	83.33	100.00	100.00
Precision	99.65	99.66	99.67	98.96	100.00	100.00
Accuracy	93.40	94.97	97.48	94.34	96.23	99.06
F <sub>1</sub> score	96.45	97.36	98.68	96.94	98.06	99.51
DB = CASIA-V5						
	Training (1500)			Testing (1000)		
	CLVQ	ADLVQ	CADLVQ	CLVQ	ADLVQ	CADLVQ
Sensitivity	87.59	95.37	96.08	94.85	97.44	97.98
Specificity	92.31	93.75	90.91	66.67	80.00	100.00
Precision	99.17	99.86	99.86	98.92	99.48	100.00
Accuracy	88.00	95.33	96.00	94.00	97.00	98.00
F <sub>1</sub> score	93.02	97.56	97.93	96.84	98.45	98.98

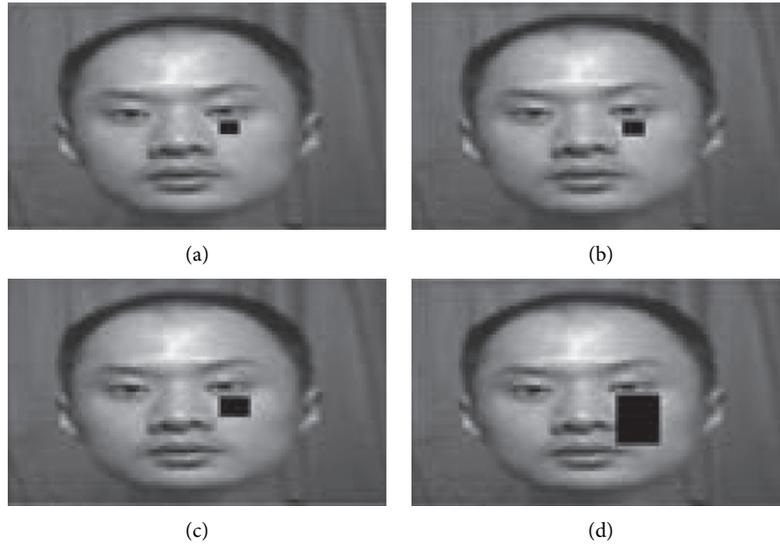


FIGURE 10: Occluded facial images with different block sizes. (a)  $5 \times 5$ . (b)  $6 \times 6$ . (c)  $8 \times 8$ . (d)  $10 \times 40$ .

TABLE 5: CADLVQ weighted sum, weighted product, and majority voting results.

	SDUMLA-HMT	CASIA-V5
Weighted sum	97.16	95.28
Weighted product	96.25	94.25
Majority voting	99.06	98.00

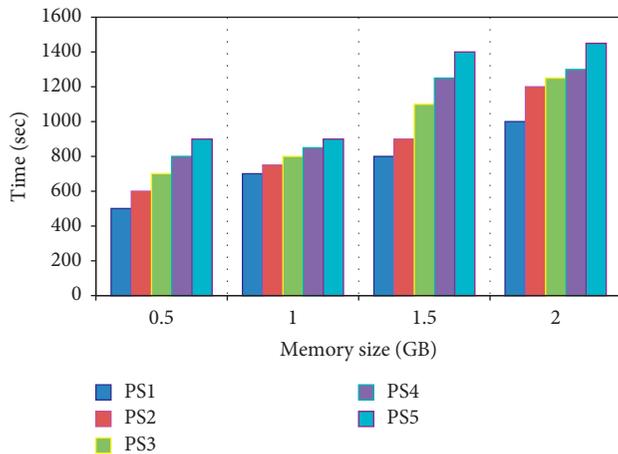


FIGURE 11: CADLVQ complexity.

TABLE 6: Recognition accuracy of the 40 occluded facial images with different block sizes.

Image size format	SDUMLA-HMT	CASIA-V5
$5 \times 5$	93.25	92.65
$6 \times 6$	90.36	90.05
$8 \times 8$	90.01	89.25
$10 \times 40$	85.54	85.08

TABLE 7: Comparison between ADLVQ and CADLVQ.

	ADLVQ	CADLVQ
Hyper-parameters	Fixed	Variable
Adaptation	Adaptation in procedure stage	Adaptation in both procedure stage and hyper-parameter inputs
Complexity	High	Very high
Maximum memory size	1.5 GB	2 GB
Average accuracy	96.62%	98.53%

## 6. Conclusion and Future Work

In this paper, we present a combined deep learning vector quantization classifier, utilizing different parameter sets to select the winner class using majority voting algorithm to classify and recognize multipose face images. Speeded Up Robust Feature (SURF) was used for extracting the features of facial images, which will be enrolled to the ADLVQ classifier. Experiments were conducted using SDUMLA-HMT and CASIA-V5 standard databases, considering three different views of the face including the frontal view. The proposed classifier performance evaluation was presented as a confusion matrix, in terms of sensitivity, specificity, precision, accuracy, and F1score. Results indicated that the proposed classifier has achieved higher recognition accuracy than ten other classifiers of the state of art. For the future, we will proceed to enhance the proposed classifier performance to be able to handle the spoof attacks problem that may be occurred by fake subjects.

## Data Availability

The data used in the study are available at <http://mla.sdu.edu.cn/sdumla-hmt.html> and <http://www.idealtest.org/dbDetailForUser.do?id=9>.

## Conflicts of Interest

The authors declare that they have no conflicts of interest.

## Acknowledgments

This project was funded by the Deanship of Scientific Research (DSR), King Abdulaziz University, Jeddah, under grant no. D-364-612-1441. The authors, therefore, gratefully acknowledge DSR technical and financial support.

## References

- [1] S. Yang, P. Luo, C. C. Loy, and X. Tang, "Wider face: A face detection benchmark," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 5525–5533, Las Vegas, NV, USA, July 2016.
- [2] N. Tsatsoulis, Y. Avrithis, and S. Kollias, "Efficient face detection for multimedia applications," in *Proceedings of the IEEE International Conference on Image Processing (ICIP)*, Vancouver, BC, Canada, September 2000.
- [3] P. Dou and I. A. Kakadiaris, "Multi-view 3D face reconstruction with deep recurrent neural networks," *Image and Vision Computing*, vol. 80, pp. 80–91, 2018.
- [4] X. Shao, J. Lyu, J. Xing et al., "3D face shape regression from 2D videos with multi-reconstruction and mesh retrieval," in *Proceedings of the IEEE International Conference on Computer Vision Workshops*, Seoul, Republic of Korea, October 2019.
- [5] F. Wu, L. Bao, Y. Chen et al., "MVF-Net: Multi-view 3d face morphable model regression," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 959–968, Long beach, CA, USA, June 2019.
- [6] H. Zhou, P. Chen, and W. Shen, "A multi-view face recognition system based on cascade face detector and improved Dlib," in *MIPPR 2017: Pattern Recognition and Computer Vision*, Xiangyang, China, March 2018.
- [7] B. Renuka, B. Sivaranjani, A. M. Lakshmi, and D. N. Muthukumar, "Automatic enemy detecting defense robot by using face detection technique," *Asian Journal of Applied Science and Technology*, vol. 2, no. 2, pp. 495–501, 2018.
- [8] X. Sun, P. Wu, and S. C. H. Hoi, "Face detection using deep learning: An improved faster RCNN approach," *Neuro-computing*, vol. 299, pp. 42–50, 2018.
- [9] E. Zhou, Z. Cao, and J. Sun, "Gridface: Face rectification via learning local homography transformations," in *Proceedings of the European Conference on Computer Vision (ECCV)*, pp. 3–20, Munich, Germany, September 2018.
- [10] Y. Li, S. Gong, and H. Liddell, "Support vector regression and classification based multi-view face detection and recognition," in *Proceedings of the Fourth IEEE International Conference on Automatic Face and Gesture Recognition*, pp. 300–305, Grenoble, France, March 2000.
- [11] Y. Li, S. Gong, J. Sherrah, and H. Liddell, "Support vector machine based multi-view face detection and recognition," *Image and Vision Computing*, vol. 22, no. 5, pp. 413–427, 2004.
- [12] S. Moore and R. Bowden, "Local binary patterns for multi-view facial expression recognition," *Computer Vision and Image Understanding*, vol. 115, no. 4, pp. 541–558, 2011.
- [13] K. Zhang, Z. Zhang, Z. Li, and Y. Qiao, "Joint face detection and alignment using multitask cascaded convolutional networks," *IEEE Signal Processing Letters*, vol. 23, no. 10, pp. 1499–1503, 2016.
- [14] T. Zhang, W. Zheng, Z. Cui, Y. Zong, J. Yan, and K. Yan, "A deep neural network-driven feature learning method for multi-view facial expression recognition," *IEEE Transactions on Multimedia*, vol. 18, no. 12, pp. 2528–2536, 2016.
- [15] S. S. Farfate, M. J. Saberian, and L.-J. Li, "Multi-view face detection using deep convolutional neural networks," in *Proceedings of the 5th ACM on International Conference on Multimedia Retrieval*, pp. 643–650, Shanghai, China, June 2015.
- [16] Y. Tao, H. Chen, and C. Qiu, "Wind power prediction and pattern feature based on deep learning method," in *Proceedings of the IEEE PES Asia-Pacific Power and Energy Engineering Conference (APPEEC)*, pp. 1–4, Hong Kong, China, December 2014.
- [17] G. Li, L. Deng, Y. Xu et al., "Temperature based restricted Boltzmann machines," *Scientific Reports*, vol. 6, p. 19133, 2016.
- [18] T. K. Reddy and L. Behera, "Online eye state recognition from EEG data using deep architectures," in *Proceedings of the IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, pp. 000712–000717, Budapest, Hungary, October, 2016.
- [19] M. Shams, S. Sarhan, and A. Tolba, "Adaptive deep learning vector quantisation for multimodal authentication," *Journal of Information Hiding and Multimedia Signal Processing*, vol. 8, no. 3, pp. 702–722, 2017.
- [20] M. Shams, A. Tolba, and S. Sarhan, "A vision system for multi-view face recognition," *International Journal of Circuits, Systems, and Signal Processing*, vol. 10, no. 1, pp. 455–461, 2017.
- [21] L. K. Hansen and P. Salamon, "Neural network ensembles," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 12, no. 10, pp. 993–1001, 1990.
- [22] A. D. Doulamis, N. D. Doulamis, and S. D. Kollias, "On-line retrainable neural networks: Improving the performance of neural networks in image analysis problems," *IEEE Transactions on Neural Networks*, vol. 11, no. 1, pp. 137–155, 2000.
- [23] T. Hastie, S. Rosset, J. Zhu, and H. Zou, "Multi-class adaboost," *Statistics and its Interface*, vol. 2, no. 3, pp. 349–360, 2009.
- [24] A. Vouloimos, N. Doulamis, A. Doulamis, and E. Protopapadakis, "Deep learning for computer vision: A brief review," *Computational intelligence and neuroscience*, vol. 2018, pp. 1–13, Article ID 7068349, 2018.
- [25] H. Jiang and E. Learned-Miller, "Face detection with the faster R-CNN," in *Proceedings of the 12th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2017)*, pp. 650–657, Washington, DC, USA, May 2017.
- [26] Z. Hao, Y. Liu, H. Qin, J. Yan, X. Li, and X. Hu, "Scale-aware face detection," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 6186–6195, Honolulu, HI, USA, July 2017.
- [27] C. Zhu, Y. Zheng, K. Luu, and M. Savvides, "CMS-RCNN: Contextual multi-scale region-based CNN for unconstrained face detection," in *Deep Learning for Biometrics*, pp. 57–79, Springer, Berlin, Germany, 2017.
- [28] N. Doulamis and A. Doulamis, "Fast and adaptive deep fusion learning for detecting visual objects," in *Proceedings of the European Conference on Computer Vision*, pp. 345–354, Florence, Italy, October, 2012.
- [29] Y. Gong, L. Liu, M. Yang, and L. Bourdev, "Compressing deep convolutional networks using vector quantization," 2014, <http://arxiv.org/abs/1412.6115>.
- [30] N. Doulamis and A. Vouloimos, "FAST-MDL: Fast adaptive supervised training of multi-layered deep learning models for

- consistent object tracking and classification,” in *Proceedings of the IEEE International Conference on Imaging Systems and Techniques (IST)*, pp. 318–323, Crete, Greece, October 2016.
- [31] G. Angeletti, B. Caputo, and T. Tommasi, “Adaptive deep learning through visual domain localization,” in *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA)*, pp. 7135–7142, Brisbane, Australia, May 2018.
- [32] J.-W. Kim and P.-K. Rhee, “Image recognition based on adaptive deep learning,” *The Journal of The Institute of Internet, Broadcasting and Communication*, vol. 18, no. 1, pp. 113–117, 2018.
- [33] K. Tahboub, D. Güera, A. R. Reibman, and E. J. Delp, “Quality-adaptive deep learning for pedestrian detection,” in *Proceedings of the IEEE International Conference on Image Processing (ICIP)*, pp. 4187–4191, Beijing, China, September 2017.
- [34] A. Doulamis and N. Doulamis, “Adaptive deep learning for a vision-based fall detection,” in *Proceedings of the 11th Pervasive Technologies Related to Assistive Environments Conference*, pp. 558–565, Corfu, Greece, June 2018.
- [35] A. Mughees, A. Ali, and L. Tao, “Hyperspectral image classification via shape-adaptive deep learning,” in *Proceedings of the IEEE International Conference on Image Processing (ICIP)*, pp. 375–379, Beijing, China, September 2017.
- [36] Y. Duan, J. Lu, J. Feng, and J. Zhou, “Context-aware local binary feature learning for face recognition,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 40, no. 5, pp. 1139–1153, 2018.
- [37] W. Liu, X. Yang, D. Tao, J. Cheng, and Y. Tang, “Multiview dimension reduction via Hessian multiset canonical correlations,” *Information Fusion*, vol. 41, pp. 119–128, 2018.
- [38] H. Cevikalp, H. S. Yavuz, and B. Triggs, “Face recognition based on videos by using convex hulls,” *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 1, no. 1, p. 99, 2019.
- [39] Y. Gao, S. C. Hui, and A. C. M. Fong, “A multiview facial analysis technique for identity authentication,” *IEEE Pervasive Computing*, vol. 2, no. 1, pp. 38–45, 2003.
- [40] J. Li, J. Zhao, F. Zhao et al., “Robust face recognition with deep multi-view representation learning,” in *Proceedings of the 24th ACM international conference on Multimedia*, pp. 1068–1072, Amsterdam, Netherlands, October 2016.
- [41] P. S. Prasad, R. Pathak, V. K. Gunjan, and H. V. R. Rao, “Deep learning based representation for face recognition,” in *Proceedings of the ICCCE 2019*, pp. 419–424, Longyan, China, 2020.
- [42] Y. Hu, X. Wu, B. Yu, R. He, and Z. Sun, “Pose-guided photorealistic face rotation,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 8398–8406, Salt Lake, UT, USA, June 2018.
- [43] I. Masi, F. J. Chang, J. Choi et al., “Learning pose-aware models for pose-invariant face recognition in the wild,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 41, no. 2, pp. 379–393, 2018.
- [44] J. Deng, G. Trigeorgis, Y. Zhou, and S. Zafeiriou, “Joint multi-view face alignment in the wild,” *IEEE Transactions on Image Processing*, vol. 28, no. 7, pp. 3636–3648, 2019.
- [45] F. M. Pop, M. Gordan, C. Florea, and A. Vlaicu, “Fusion based approach for thermal and visible face recognition under pose and expresivity variation,” in *Proceedings of the 9th RoEduNet IEEE International Conference*, pp. 61–66, Sibiu, Romania, June 2010.
- [46] G. Passalis, P. Perakis, T. Theoharis, and I. A. Kakadiaris, “Using facial symmetry to handle pose variations in real-world 3D face recognition,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 33, no. 10, pp. 1938–1951, 2011.
- [47] S. Tulyakov, R. L. Vieri, S. Semeniuta, and N. Sebe, “Robust real-time extreme head pose estimation,” in *Proceedings of the 22nd International Conference on Pattern Recognition*, pp. 2263–2268, Stockholm, Sweden, August 2014.
- [48] I. Masi, S. Rawls, G. Medioni, and P. Natarajan, “Pose-aware face recognition in the wild,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 4838–4846, Las Vegas, NV, USA, June 2016.
- [49] B. Lu, J. Zheng, J.-C. Chen, and R. Chellappa, “Pose-robust face verification by exploiting competing tasks,” in *Proceedings of the IEEE Winter Conference on Applications of Computer Vision (WACV)*, pp. 1124–1132, Santa Rosa, CA, USA, March 2017.
- [50] S. Sarhan, S. Hamad, and S. Elmougy, “Human injected by Botox age estimation based on active shape models, speed up robust features, and support vector machine,” *Pattern Recognition and Image Analysis*, vol. 26, no. 3, pp. 617–629, 2016.
- [51] <http://mla.sdu.edu.cn/sdumla-hmt.html>, 2019.
- [52] <http://www.idealtest.org/dbDetailForUser.do?id=9>. 2019.
- [53] S. Sarhan, S. Alhassan, and S. Elmougy, “Multimodal biometric systems: A comparative study,” *Arabian Journal for Science and Engineering*, vol. 42, no. 2, pp. 443–457, 2017.
- [54] M. Y. Shams, A. S. Tolba, and S. H. Sarhan, “Face, iris, and fingerprint multimodal identification system based on local binary pattern with variance histogram and combined learning vector quantization,” *Journal of Theoretical and Applied Information Technology*, vol. 89, no. 1, 2016.