Deeply Learned Invariant Features for Component-based Facial Recognition

Adam Hassan[1], Serestina Viriri[2]
Sudan University of Science and Technology, College of Computer Science & Information Technology, Khartoum, Sudan[1]
University of KwaZulu-Natal, School of Mathematics, Statistics & Computer Science, Durban, South Africa[2]

***Abstract*—Face recognition underage variation is a challenging problem. It is a difficult task because ageing is an intrinsic variation, not like pose and illumination, which can be controlled. We propose an approach to extract invariant features to improve facial recognition using facial components. Can facial recognition over age progression be improved by resizing independently each individual facial component? The individual facial components: eyes, mouth, and nose were extracted using the Viola-Jones algorithm. Then we utilize the eyes region rectangle with upper coordinates to detect the forehead and lower coordinates with the nose rectangle to detect the cheeks. The proposed work uses Convolutional Neural Network with an ideal input image size for each facial component according to many experiments. We sum up component scores by applying weighted fusion for a final decision. The experiments prove that the nose component provides the highest score contribution among other ones, and the cheeks are the lowest. The experiments were conducted on two different facial databases-MORPH, and FG-NET databases. The proposed work achieves a state-of-the-art accuracy that reaches 100% on the FG-NET dataset and the results obtained on the MORPH dataset outperform the accuracy results of the related works in the literature.**

***Keywords*—Invariant features; facial components; facial recognition; convolutional neural network; weighted fusion**

I. Introduction

Facial recognition basically is the activity of verifying or identifying a person's identity through facial characteristics. It captures facial features, analyses, and performs patterns comparison to know and determine the identity, or to decide whether the person is the correct one. However, the difficulties that emerge in designing an invariant face recognition system comprise variations in illumination, pose, and age.

One of the most distinguished biometric attributes is facial features, which are the most compatible with Machine Readable Travel Documents [1]. The advantage of facial features is that they can be captured at a distance without permission.

Facial ageing is not a controllable process throughout human life and cannot be avoidable, not like other variations which are flexible to handle during the image acquisition period [2]. Besides, facial ageing images are not free from extrinsic variations such as illumination, pose, and expression which add up to the challenge of finding an appropriate method for age-invariant facial recognition [3]. These approaches succeeded in finding distinctive features. The approaches are divided mainly into two groups. One of the groups is local appearance-based techniques which are used to extract local features. The whole face is portioned into small patches [4]. The other group is Key-points-based techniques specified to detect points of interest in the full-face image.

A local binary pattern (LBP) is a texture extractor used to extract distinctive features from objects [5]. It is applied in many applications like face recognition [6], texture segmentation, texture classification, and facial expression recognition. Khoi et al. [7] present a fast face recognition model using LBP and its variants based on the content approach. Karaaba et al. [8] proposed multi-HOG that combined different histograms of oriented gradients for robust face recognition.

Prince et al. in [9] used LDA based on probabilistic linear discriminant analysis (PLDA) to maximize discriminability, which is applicable for face recognition with pose variation. Perlibakas et al. [10] used Gabor features to eliminate redundancy to obtain the best descriptors for face recognition, then, used cosine similarity for evaluation.

This paper illustrates subject recognition through the facial components during ageing and relevant variations due to illumination, pose, expressions, resolutions and facial image occultation; thus, face recognition with both intrinsic and extrinsic variation is a challenging process. Moreover, extracting facial components such as the forehead and cheeks is additional work.

Several researchers performed double feature extractors for better representation as a single concatenated vector [2] or separated to vote the robust to increase the accuracy and performance simultaneously [11]. However, features extracted using hand-crafted descriptors do not have sufficient capacity to adequately represent the appearance of the face [12]. The effectiveness of deep learning has been proven to extract highly discriminative features for promising results in the field of facial ageing research [13]. A suitable alternative is to train a convolutional neural network model to identify and extract discriminative features.

The remainder of the paper is organized as follows. Section II gives a brief literature review and related work. Section III describes our methodology and techniques. Section IV discusses the experiments and analysis. Section V draws paper conclusions and future work.

II. Literature Review

Research in recognition based on facial components has seen less attention compared to the attention being given to developing approaches that use global face descriptors. For instance, holistic representations like densely sampled feature extractors such as Local Binary Patterns (LBP) [14] and Scale Invariant Feature Transforms (SIFT) descriptors [15] are heavily used in automated face recognition [16]. In contrast, research concerning facial component-based representations such as features that are extracted from specific facial components is rarely addressed in the literature.

Component-based facial recognition is an alternative to using the full-face approach. A distinguished comparison between the two approaches is found in [17]. The authors proposed a component-based and two others full-face methods. Experiments showed the superiority of the components approach on the whole face with the consideration of the robustness against pose variations.

Wang Lijia et al. [18] used Procrustes analysis, which is insensitive to rotation, translation and scale to align facial components. They used a random measurement matrix to extract components' features and applied the Gradient projection Algorithm for classification. The method proved that component-based approaches matched better than holistic approaches in terms of accuracy rate.

Early research [19] illustrated that the upper face components are more discriminant than other components. Authors in [18] solved the one training sample problem by using a component-based linear discriminant analysis (LDA) method through five facial parts to construct component bunches.

Multi-feature extraction is also used to extract textural and shape from facial components such as eyes, mouth, and nose. The algorithm is flexible enough to be satisfied with the non-occluded facial parts to perform the recognition [20].

Authors in [21] proposed decision-level fusion containing 34 region classifiers decision-level fusion is performed with majority voting.

Boussaad et al. [22] used a pre-trained CNN model on resized components to cope with Alex Net input layer size, then Discriminant Correlation Analysis for fusion and Support Vector Machine for classification.

Component-based methods have proven effective when used to handle age-invariant features. For instance, [22] applied a Discriminant Correlation Analysis (DCA) as a feature-level fusion on separated components features and performed facial classification using a Support Vector Machine (SVM) and obtained a 97.87% as recognition accuracy rate.

This work proposes a component-based approach for age-invariant face recognition using deeply learned features extracted from separated components (eyes, nose, mouth, forehead, and cheeks), then matching score level fusion is performed, and cosine similarity is used for classification.

III. Methodology and Techniques

We propose a methodology to study invariant features for component-based facial recognition during age progression. Fig. 1 illustrates a flow diagram of our proposed methodology that consists of sequential steps. Firstly, facial images are pre-processed through facial component detection, cropping, resizing, and data augmentation. Secondly, the convolutional neural Network CNN extracts discriminative features through the augmented facial components for training the CNN model for validation and gaining the accuracy of testing sets. Thirdly, we fuse the scores obtained by every component for a final overall decision.

A. Facial Datasets

In this work, we use the two freely available datasets MORPH (Album II) [23] and FGNET [24] to perform age-invariant facial recognition. The MORPH dataset consists of 55,134 facial images belonging to 13,617 classes with an age range that extends from 16 to 77 years. The FGNET dataset includes 1002 images of a total number of 82 classes, and its age range begins from 0 to 69 years. The two sets of data that include age variation are also exposed to lighting, expressions, and head position. The Morph dataset contains 9260 images that belong to 754 classes that differ in age. The images are categorized into classes; each contains images of different ages, not exceeding the 5 years age gap. Two separate datasets were created and randomly selected. 70% of images are considered to train the CNN network, while the rest 30% remain for testing.

B. Facial Components Detection

Face alignment is a technique for identifying the geometric structure of the human face. In previous work [11] the angle of inclination of the straight line between the outer corners of the two eyes is used to align the face horizontally. The facial component algorithm using Viola and Jones performs well with optimum execution complexity [25]. However, we use the locations of the eyes, nose, and mouth to detect additional components such as the forehead and cheeks. Fig. 2(a) shows whole face and component detection, and Fig. 2(b) the cropped components analyzed in this work. The three other components are supposed to be discriminating for the facial recognition model of our work.

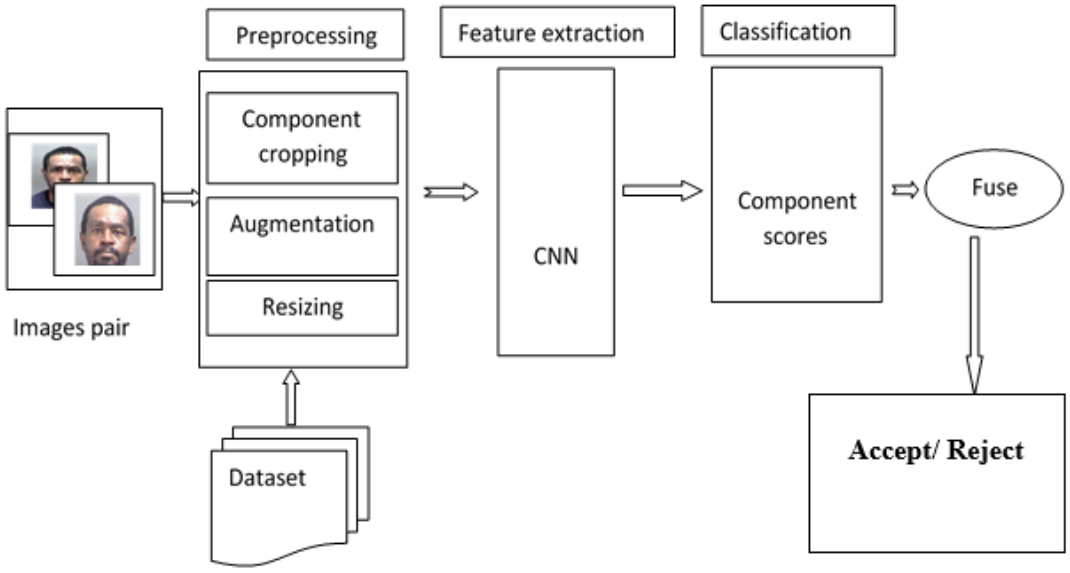

Fig. 1. Proposed Methodology Framework.

Algorithm 1 describes the other three components depending on Viola and Jones. Many images from the databases used in this work have different pose variations. Therefore, this work employs face alignment as in [11] for correcting horizontal facial images before applying even Viola and Jones detection. We get the benefit from the detected facial components and rectangles boundaries. Specifically, the coordinates of the top left, the width, and the height of the pre-detected components, are crucial to detecting the other components.

Algorithm 1**:** Components Extraction (forehead, left cheek, and right cheek) – from the Face dataset

Input**:** Training set *M*, with *m* classes

nj = number of images in a given class

Bbox: boundary box

for *i* = 1 *to m* **do**

for *j* = 1 *to nj* **do**

Bbox (forehead) → Bbox (eyes pair)

Bbox (left check) → Bbox (eyes pair) + Bbox (nose)

Bbox (left check) → Bbox (eyes pair) + Bbox (nose)

end for

end for

C. Image Pre-processing

Initially, images in the datasets were pre-processed to enhance performance and gain improved accuracy results. Facial components, specifically the eyes, nose, mouth, forehead, and cheeks are detected and cropped using the Viola-Jones algorithm for the first three components and the proposed algorithm illustrated above for the last two facial components. Then the images are translated in different directions, rotated and resided to increase the number of images in the training datasets. Finally, the images are capable enough to be fed to the convolutional neural network with RGB color channels.

D. Feature Extraction and Classification

The convolutional neural network is fed with images after completely being preprocessed. In this work, we propose age-invariant facial recognition using its components and utilizing a Convolutional neural network. CNN is the preferable and most powerful employed algorithm in the area of deep learning [26]. The clear advantage of CNN is that relevant features are automatically identified away from human supervision [27].

1) Convolutional Neural Network: CNN architecture is still an open problem. That, the size of training data determines the best number of layers and filters to avoid over fitting [28].

In this work, extract parts or components of the face like eyes, nose, mouth, forehead, and cheeks instead of a full face to recognize a person. We conduct the work using three convolution layers, two max pooling, and one fully connected layer. The multi-layer neural networks built to recognize discriminant features from the origin pixels of images preceded by suitable pre-processing for the intended purpose. In CNN architectures, the main layers for successful models are convolution layers followed by down-sampling in pooling layers and concluding with fully connected layers.

2) Cosine similarity: Cosine similarity is a suitable choice for metric learning due to its special property of providing similarity between the intervals -1 to +1 [29]. Cosine similarity (CS) between two vectors x and y is defined as the following:

$$CS(x, y) = \frac{X^T Y}{||x||\,||y||} \quad (1)$$

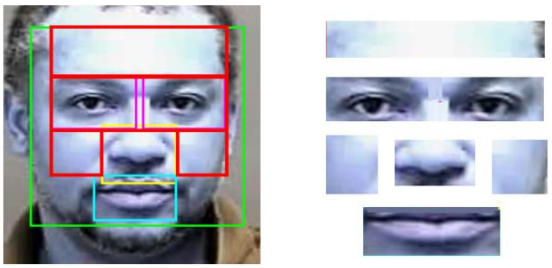

Fig. 2. (a) Facial Components Detection and (b) Facial Components Cropping.

3) Euclidean distance: Euclidean distance [30] is used to evaluate the performance between two feature vectors of pairs. Euclidian distance for two image pairs featuring vectors x and y can be calculated as follows:

$$d(x, y) = ||x - y|| \quad (2)$$

Given two image feature sets, $x = \{x_1, x_2 \ldots x_n\}$,

And $y = \{y_1, y_2 \ldots y_n\}$, the function defines similarity distance between the two sets adopting the minimum distance as the following.

$$h_{min}(X, Y) = \min d_{(x \epsilon X, y \epsilon Y)}(x,y) \quad (3)$$

Cosine similarity and Euclidean distance measures receive the feature vectors extracted by CNN, then calculate the similarity and distance between two image pair feature vectors as mentioned above. When the result yields less than the threshold, the two faces are regarded as the same identity, otherwise, regarded as different identities.

IV. Experimental Results and Analysis

MORPH (Album II) [23] is one of the data sets on which we conduct our experiments. The dataset is divided into two main experimental groups: 1) age gap 0-1 which is composed of 2590 images belonging to 494 classes. 2) Age gap 1-5, which includes 5335 images from 942 classes. Then we select randomly seventy per cent of the images in the dataset for the training process and the rest for testing. From both training and testing datasets, six different sub-datasets of eyes, nose, mouth, forehead, left cheek, and right cheek was created. Different CNN networks with the same number of sub-datasets are built. The matched image pairs with the highest similarity are considered as the same identity. To enhance the performance of the networks, we conducted various experiments for each component with different input sizes.

For experimental analysis, accuracy and error rates are calculated. As a result, the most appropriate learning rate and appropriate input image size for each component are adopted using the following equations:

$$\text{Accuracy} = \frac{\text{No. of testing samples classified correctly}}{\text{Total no.of testing samples}} \quad (4)$$

$$\text{Error rate} = \frac{\text{No. of misclassification}}{\text{Total no.of testing samples}} \quad (5)$$

To recognize the image ideally, the input image size plays a crucial role based on the extracted complex features. We notice that image expansion or compression degrades the accuracy. So, many input sizes are chosen to analyze and depending upon performance parameters, an applicable size for all components is recommended.

Learning rate is a hyper-parameter that specifies the adjustment in the weights of the network depending on the loss gradient descent function. It specifies how fast or slow the network will reach the optimal weights. If the learning rate is too high it will skip the optimal solution and if it is too low then, too much iteration will be spent to converge to reach the best values. Thus, utilizing a good learning rate is important.

The experiments involved various input sizes and learning rates and all results were recorded. Generally, six different networks are conducted, one for each component. Each network is trained using only one component and tested using the same component but from other classes not seen during the training phase. All relevant combinations of size and learning rate are carried out and the loss is computed.

For the eyes region component, throughout all various combinations, the size of 64x64 and 0.001 learning rate show the least error of 1.075 and accuracy of 98.925 at the training phase. Thus, it is recommended to use the mentioned size and learning rate for the eye region component for promising results. Fig. 3 illustrates the findings.

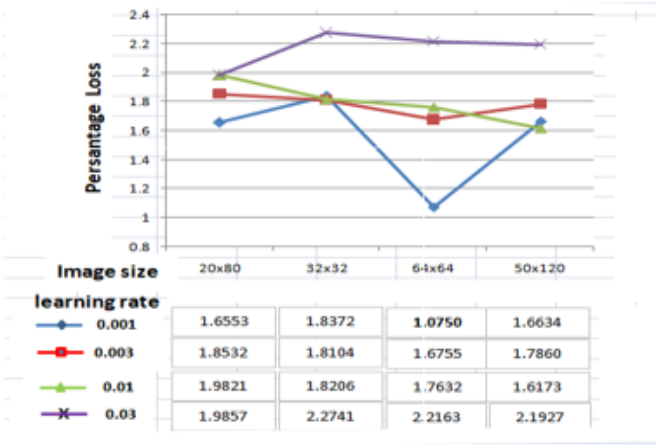

Fig. 3. Percentage Loss Graph for Various Input Image Sizes for Eyes Component.

For the nose component, as depicted in Fig. 4, the size of 32x32 and 0.003 learning rate show the least error of 0.5871% and accuracy of 99.0315. Therefore, the preferable size and learning rate for the component are 32x32 and 0.003 for the trained dataset. Similarly, Fig. 5 shows that the size of 60x80 and 0.001 learning rate for the mouth component produced the least error of 5.7469 and gives the accuracy of 93.9512%.

Two additional components are extracted- forehead and cheeks, using algorithm 1. Following a similar way to find the best image size and learning rate, the goal is to utilize various input sizes and combinations of learning rates. It is noticeable, as shown in Fig. 6 that the size of 40x64 and 0.001 learning rate are the best combination examined that gives the least error rate of 7.1472 and accuracy of 92.8528 for the forehead component.

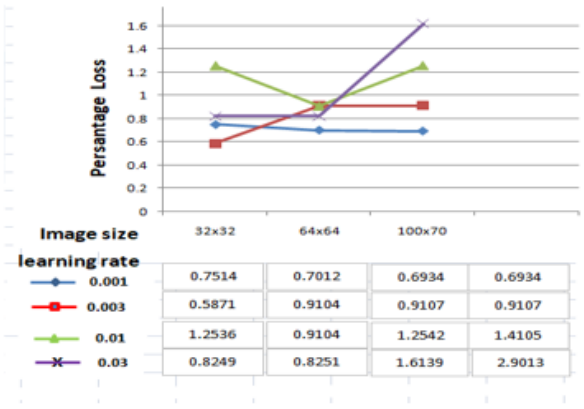

Fig. 4. Percentage Loss Graph for Various Input Image Sizes for Nose Component.

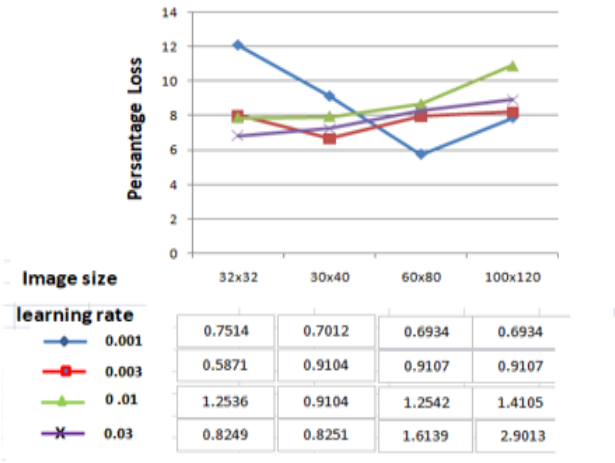

Fig. 5. Percentage Loss Graph for Various Input Image Sizes for Mouth Component.

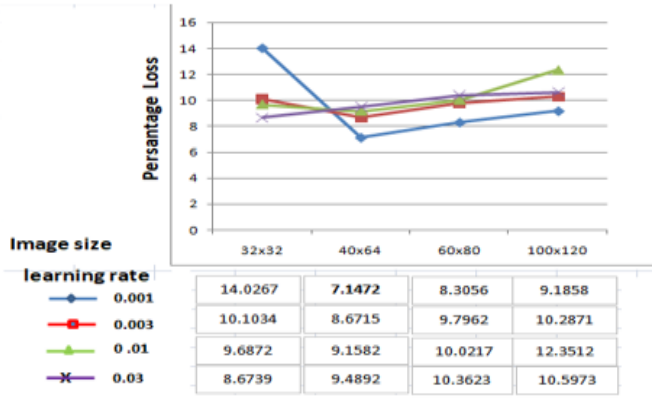

Fig. 6. Percentage Loss Graph for Various Input Image Sizes for Forehead Component.

The other two components are the left and right cheeks which give close results. Fig. 7 and Fig. 8 depicted the results and it's clear to deduce that the best size for both is 32x32 and the combined learning rate is 0.001. The best training accuracy is 90.5167% and 90.4512 for left and right cheek respectively.

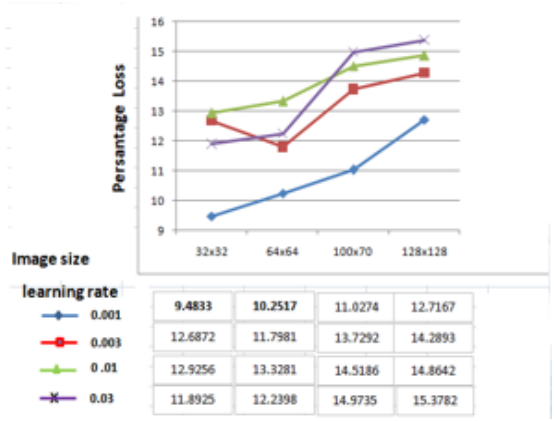

Fig. 7. Percentage Loss Graph for Various Input Image Sizes for Left Cheek Component.

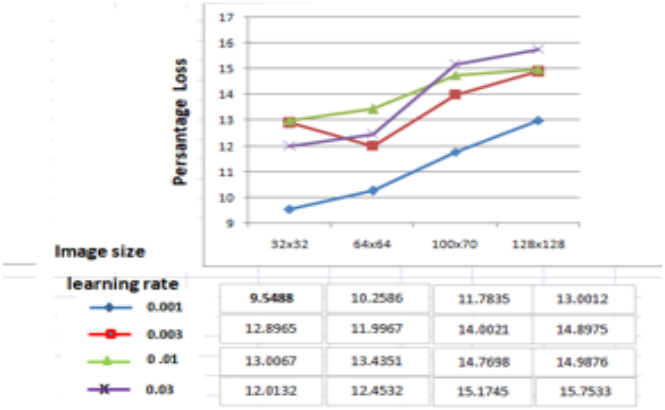

Fig. 8. Percentage Loss Graph for Various Input Image Sizes for Right Cheek Component.

A. Training Parameters

This work selects three convolutional layers, two max pooling, one fully connected layer, and a classification layer. All convolution layers are immediately followed by a ReLU (Rectified Linear Unit) activation function. The max-pooling size is 2×2 and stride 2 for down sampling. Images in datasets are RGB color and each component has its applicable resized input image and the best learning rate.

Images in each CNN of the mentioned components are divided into mini-batches. The mini-batch is set to size 100. Each batch is trained and convoluted with a fixed number of generated filters combined with a determined bias and a computed constant value.

In the phase forward pass for each hidden layer neuron, the activation is calculated as:

$$net_h = \sum_{i=1}^{n} w_i * x_i + b \quad (6)$$

Where net_h represents the total net input for the determinant hidden layer, n is the number of filter weights, w_i is i_{th} filter weight, x is the input neuron, b: bias. Perform the same process for successive layers. Repeat this process for the output layer neurons, using the output from the prior hidden layer neurons as input for the subsequent hidden layer. To get

the total error (E_{total}), squared error function is applied for each output neuron as follows.

$$\text{E total} = \frac{1}{2}\sum_{i=1}^{n} (T_i - Y_i)^2 \quad (7)$$

Where n is the number of output neurons, T_i is the target output for neuron i and, Yi is the output of the neuron i calculated by a forward convolution pass.

The Backwards Pass: the purpose is to update all weights in the network to provide the calculated output to be closer the target output, by minimizing the error for each output neuron and update the whole network**.** Now the crucial role is to know how much the change in each weight will affect the overall error. In other words, the role is to calculate the partial derivative of the total error with respect to each weight. The step describes the partial derivative of $E_{total}$with respect to w_i . That means to obtain the gradient with respect to w_i. Properly, the following chain rule is applied:

$$\frac{\partial \text{E total}}{\partial w_i} = \frac{\partial \text{E total}}{\partial out_{01}} * \frac{\partial out_{01}}{\partial net_{01}} * \frac{\partial net_{01}}{\partial w_i} \quad (8)$$

The final step is minimizing error by subtracting the value multiplied by some learning rate from the current weight. The following is actually what happens to update each weight:

$$w_i^+ = w_i - \eta * \frac{\partial e}{\partial w_i} \quad (9)$$

When all weights are updated after a repeated process, we roll into a forward pass using the updated weights. Final activations fed to the fully connected layer transform the learned neurons into a new embedding vector. The distance between the impostor pair is enlarged while the distance between the genuine pair is minimized.

B. Data Augmentation

Data augmentation balances the samples size of training set when some classes have abundant samples while the rest lack appropriate number of samples [26] . Data augmentation is a technique that prevents the network from over fitting through discriminant features of the training images [31]. The input facial component images are translated horizontally and vertically in the range [-30, 30]. Then, we rotate the images and resize them to the ideal input size for each component.

C. Integrating Facial Components

In this work, to verify components of the same type, metric learning to discriminate the similarity is used. Many research works applied Mahalanobis distance learning [32] and Euclidean distance [33] but the drawback of the Mahalanobis distance is the equal adding up of the variance normalized squared distances of the features. However, the key issue of Euclidian norm that it gives the same importance to any direction. So, metric learning that capable with angular distribution to calculate similarity is needed. The best choice is cosine distance to deal with metric learning that provide reliable classification [34].

D. Combining Components Scores

All scores of the facial components are fused at the matching score level. The obtained output matching scores of each of these components are weighted and combined. Fusion at the matching score level is usually preferred, as it is relatively easy to access and combine the scores presented by the different modalities [36]. The adopted matching score in this work is the same as the algorithm specified in [35] but, with a cosine similarity output of each facial component.

E. Score Normalization

Each component is treated independently, and its own score is computed. Then, each component's scores are normalized given a set of n raw component matching scores $\{S_k\}$, k = 1, 2... n. For the sake of this work, n should be 1 up to 6 the number of analysed facial components. The corresponding normalized scores S_k' is given by Min-max normalization as the following:

$$s'_k = \frac{s_k - \min(\{s_k\})}{\max(\{s_k\}) - \min(\{s_k\})} \quad (10)$$

Where min and max are the minimum and maximum seen at learning phase, respectively, of the given set $\{S_k\}$ of component matching scores.

F. Score Weighting

Each component score is given a weighing to achieve an overall fusion. Let s'_e s'_n , s'_m, s'_f, and s'_c be the normalized scores for a specific class - i and t_1, t_2, t_3, t_4, and t_5 are the thresholds of the eyes, nose, mouth, forehead, and cheeks respectively. Then the initial weights of components scores are computed as follows:

$$w'^i_e = \frac{s'^i_e}{t_1 + s'^i_e} \quad (11)$$

$$w'^i_n = \frac{s'^i_n}{t_2 + s'^i_n} \quad (12)$$

$$w'^i_m = \frac{s'^i_m}{t_3 + s'^i_m} \quad (13)$$

$$w'^i_f = \frac{s'^i_f}{t_4 + s'^i_f} \quad (14)$$

$$w'^i_c = \frac{s'^i_c}{t_5 + s'^i_c} \quad (15)$$

The preliminary weights w'^i_e, w'^i_n, w'^i_m, w'^i_f, and w'^i_c related to eyes, nose, mouth, forehead, and cheeks respectively. Next, the fusion weights for the i^{th} class are computed respectively, as follows:

$$w^i_e = \frac{w'^i_e}{w'^i_e + w'^i_n + w'^i_m + w'^i_f + w'^i_c} \quad (16)$$

$$w^i_n = \frac{w'^i_n}{w'^i_e + w'^i_n + w'^i_m + w'^i_f + w'^i_c} \quad (17)$$

$$w^i_m = \frac{w'^i_m}{w'^i_e + w'^i_n + w'^i_m + w'^i_f + w'^i_c} \quad (18)$$

$$w^i_f = \frac{w'^i_f}{w'^i_e + w'^i_n + w'^i_m + w'^i_f + w'^i_c} \quad (19)$$

$$w^i_c = \frac{w'^i_c}{w'^i_e + w'^i_n + w'^i_m + w'^i_f + w'^i_c} \quad (20)$$

Then the fusion score is computed as follows:

$$S_{fuse} = w_e^i s_e' + w_n^i s_n' + w_m^i s_m' + w_f^i s_f' + w_c^i s_c' \quad (21)$$

Cosine similarity between the two facial image pairs is the similarity score between the two faces. Lastly, this score is compared with a threshold to decide whether two faces belong to the same class or not. The threshold *is* selected from the training set so that False Acceptance Rate is equivalent to False Rejection Rate. The decision function defined in Equation (22) verifies the class.

$$Decision\left(S_{fuse}\right) = \begin{cases} Accept, if > threshold \\ Reiect. otherwise \end{cases} \quad (22)$$

The ROC curve of the experiments on MORPH dataset is depicted in Fig. 9. The curve shows that the cheek component has the lowest accuracy while the nose has the highest contribution that reflects its stability during aging.

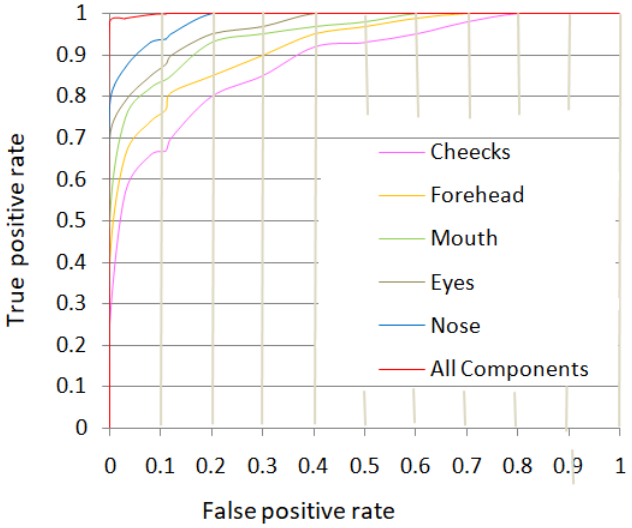

Fig. 9. MORPH Results on 1-5 Year Age Gap Data Set.

TABLE I. PER COMPONENT TAR (%) AT 1% FAR, MORPH 0-1 YEAR AGE GAP SET

Components	**Accuracy %**	
Eyes	90.31	
Nose	95.12	
Mouth	87.07	
Forehead	82.52	
Cheeks	77.45	
fusion	**Enhancement%**	**Accuracy%**
Forehead + Cheeks	0.58	83.1
Mouth + forehead	1.53	88.6
Eyes+ forehead	2.55	92.86
Eyes+ mouth	2.95	93.26
Nose + cheeks	1.46	96.58
Nose+ forehead	2.82	97.94
Nose + mouth	3.25	98.37
Nose +eyes	3.56	98.68
All components	4.83	99.95

The Tables I to IV display per component accuracy performance and detail how the proposed method improves the performance using two-facial component score fusion. Moreover, it is noticeable that the nose component has the highest accuracy, and the cheeks are the lowest. Performance is enhanced when scores of the two components are fused. Eyes and nose combination increases the accuracy for both data subsets. However, the accuracy decreased dramatically when forehead and cheek scores fused. But, when scores of three components or more are fused the accuracy is highly improved. We compared our work with the previous most related works. The comparison results are shown in Table V.

TABLE II. PER COMPONENT TAR (%) AT 1% FAR, MORPH 1-5 YEAR AGE GAP SET

Components	**Accuracy %**	
Eyes	86.92	
Nose	95.10	
Mouth	84.28	
Forehead	77.62	
Cheeks	70.33	
fusion	**Enhancement%**	**Accuracy%**
Forehead + Cheeks	0.10	77.72
Mouth + forehead	1.20	85.48
Eyes+ forehead	2.12	89.04
Eyes+ mouth	2.43	89.35
Nose + cheeks	1.05	96.15
Nose+ forehead	2.24	97.34
Nose + mouth	2.80	97.9
Nose +eyes	3.20	98.3
All components	4.80	99.9

TABLE III. PER COMPONENT TAR (%) AT 1% FAR FG-NET 0-1 YEAR AGE GAP SET

Components	**Accuracy %**	
Eyes	90.81	
Nose	95.8	
Mouth	88.12	
Forehead	86.72	
Cheeks	80.04	
fusion	**Enhancement%**	**Accuracy%**
Forehead + Cheeks	0.51	87.23
Mouth + forehead	1.5	89.62
Eyes+ forehead	2.87	93.68
Eyes+ mouth	3.23	94.04
Nose + cheeks	1.87	97.67
Nose+ forehead	2.91	98.71
Nose + mouth	3.25	99.05
Nose +eyes	3.67	99.47
All components	4.2	100

TABLE IV. PER COMPONENT TAR (%) AT 1% FAR, FG-NET 1-5 YEAR AGE GAP SET

Components	**Accuracy %**	
Eyes	89.24	
Nose	94.92	
Mouth	86.05	
Forehead	82.29	
Cheeks	73.68	
fusion	**Enhancement%**	**Accuracy%**
Forehead + Cheeks	0.44	82.73
Mouth + forehead	1.1	87.15
Eyes+ forehead	2.23	91.47
Eyes+ mouth	2.75	91.99
Nose + cheeks	1.73	96.65
Nose+ forehead	2.52	97.44
Nose + mouth	3.1	98.02
Nose +eyes	3.54	98.46
All components	5.08	100

TABLE V. COMPARISON OF OUR RESULT WITH COMPONENT BASED ALGORITHMS

Description	**Compone nt used**	**Component representation**	**Face database**	**Accuracy**	
Component-based LDA method with component Bunches [18]	L. eye, R. eye, nose, R. mouth, L. mouth	Pixel representation	FERET	93.57%	
Component-based with procruetes analysis [19]	Eyes, nose, and mouth	Random measurement matrix	Extended Yalu face database	96%	
Resized components to pre-trained model [20]	Periocular region, nose, mouth	Pre-trained AlexNet CNN	FG-NET database	98.31%	
Components with landmarks [2]	Eyes ,Nose, Mouth and Eyebrows	Multi-scale local binary pattern and scale-invariant feature transform	PCSO and MORPH database	97.60%	
The Proposed work	Eyes, nose, mouth, cheeks and forehead	Convolutional Neural Network	MORPH database	99.90 %-99.95 %	
			FG-NET data-base	100 %	

V. CONCLUSIONS AND FUTURE WORK

This paper addressed the challenge of facial recognition on aging subjects using Convolutional Neural Network. Facial components such as eyes, nose, mouse, forehead and cheeks are resized independently, each as ideal input image size. Each subset of the same type classified indecently using cosine similarity. Weighted fusion is utilized to sum up all facial components scores for a final decision. We tested our work using FG-NET and MORPH publicly available datasets. The proposed work achieved a state-of-the art accuracy of 100% on FG-NET dataset and the results obtained on MORPH dataset outperform accuracy results obtained in the literature. Our future work will focus on facial components that give high score contributions and updating the weights of the candidate components to improve overall performance.

REFERENCES

[1] Harakannanavar, Sunil S., C. R. Prashanth, Vidyashree Kanabur, Veena I. Puranikmath, and K. B. Raja, "Technical Challenges, Performance Metrics and Advancements in Face Recognition System," International Journal of Computer Sciences and Engineering 7, no., pp. 836-847, 3 2019.

[2] Otto, Charles, Hu Han, and Anil Jain, "How does aging affect facial components?.," in 7584, Berlin, Heidelberg, 2012.

[3] Liao, Shengcai, Anil K. Jain, and Stan Z. Li, "Partial face recognition: Alignment-free approach," IEEE Transactions on pattern analysis and machine intelligence, vol. 35, no. 5, pp. 1193-1205, 6 Sep 2012.

[4] Kortli, Yassin, Maher Jridi, Ayman Al Falou, and Mohamed Atri, "A novel face detection approach using local binary pattern histogram and support vector machine," in International Conference on Advanced Systems and Electric Technologies (IC_ASET), 2018.

[5] Ojala, Timo, Matti Pietikäinen, and David Harwood, "A comparative study of texture measures with classification based on featured distributions," Pattern recognition, vol. 29, no. 1, pp. 51-59., 1996.

[6] Napoléon, Thibault, and Ayman Alfalou, "Pose invariant face recognition: 3D model from single photo," Optics and Lasers in Engineering, vol. 89, no. 1, pp. 150-161, 1 Feb 2017.

[7] Khoi, Phan, Lam Huu Thien, and Hoai Viet Vo, "Face retrieval based on local binary pattern and its variants: a comprehensive study," International Journal of Advanced Computer Science and Applications, vol. 7, no. 6, 2016.

[8] Karaaba, Mahir, Olarik Surinta, Lambert Schomaker, and Marco A. Wiering, "Robust face recognition by computing distances from multiple histograms of oriented gradients," IEEE Symposium Series on Computational Intelligence, pp. 203-209), 7 Dec 2015.

[9] Prince, Simon, Peng Li, Yun Fu, Umar Mohammed, and James Elder., "Probabilistic models for inference about identity," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 34, no. 1, pp. 144-157., May 2011.

[10] "Face recognition using principal component analysis and log-gabor filters," arXiv preprint cs/0605025., 7 May 2006.

[11] Hassan, Adam, and Serestina Viriri, "Invariant feature extraction for component-based facial recognition," International Journal of Advanced Computer Science and Applications, vol. 11, no. 11, 2020.

[12] Zhai, Huanhuan, Chunping Liu, Husheng Dong, Yi Ji, Yun Guo, and Shengrong Gong., "Face verification across aging based on deep convolutional networks and local binary patterns," International conference on intelligent science and big data engineering, pp. 341-350, 14 Jun 2015.

[13] Yousaf, Adeel, Muhammad Junaid Khan, Muhammad Jaleed Khan, Adil M. Siddiqui, and Khurram Khurshid., "A robust and efficient convolutional deep learning framework for age-invariant face recognition," Expert Systems, vol. 37, no. 3, p. e12503, 2020.

[14] Ojala, Timo, Matti Pietikainen, and Topi Maenpaa, "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns," IEEE Transactions on pattern analysis and machine intelligence, vol. 24, no. 7, pp. 971-987., 7 Aug 2002.

[15] Klare, Brendan, Zhifeng Li, and Anil K. Jain, "Matching forensic sketches to mug shot photos," IEEE transactions on pattern analysis and machine intelligence, vol. 33, no. 3, pp. 639-646, 14 Oct 2010.

[16] Klare, Brendan, and Anil K. Jain, "On a taxonomy of facial features," in IEEE International Conference on Biometrics: Theory, Applications and Systems (BTAS), 2010.

[17] Heisele, Bernd, Purdy Ho, Jane Wu, and Tomaso Poggio., "Face recognition: component-based versus global approaches," Computer vision and image understanding,, vol. 91, no. 1-2, pp. 6-21, Jul 2003.

[18] Wang, Lijia, Hua Zhang, and Zhenjie Wang., "Component based representation for face recognition.," 2015.

[19] Sadr, Javid, Izzat Jarudi, and Pawan Sinha, "The role of eyebrows in face recognition," vol. 32, no. 3, pp. 285-293, 2003.

[20] Gumede, A., Serestina Viriri, and M. Gwetu., "[Hybrid Component-based Face Recognition," in Conference on Information Communication Technology and Society (ICTAS),, 2017.

[21] Seal, Ayan, Debotosh Bhattacharjee, Mita Nasipuri, Consuelo Gonzalo-Martin, and Ernestina Menasalvas, "À-trous wavelet transform-based hybrid image fusion for face recognition using region classifiers," Expert Systems, vol. 35, no. 6, p. e12307, 2018.

[22] Boussaad, Leila, and Aldjia Boucetta, "An effective component-based age-invariant face recognition using Discriminant Correlation Analysis," Journal of King Saud University-Computer and Information Sciences, 25 Aug 2020.

[23] Sahoo, Tapan Kumar, and Haider Banka, "Multi-feature-based facial age estimation using an incomplete facial aging database," Arabian Journal for Science and Engineering, vol. 43, no. 12, pp. 8057-8078, 2018.

[24] Panis, Gabriel, Andreas Lanitis, Nicholas Tsapatsoulis, and Timothy F. Cootes, ". Overview of research on facial ageing using the FG-NET ageing database," Iet Biometrics, vol. 5, no. 2, pp. 37-46, 2016.

[25] Viola, Paul, and Michael Jones, "Rapid object detection using a boosted cascade of simple features," in In Proceedings of the 2001 IEEE computer society conference on computer vision and pattern recognition. CVPR, 2001.

[26] Krizhevsky, Alex, Ilya Sutskever, and Geoffrey E. Hinton, "Imagenet classification with deep convolutional neural networks," Communications of the ACM, vol. 60, no. 6, pp. 84-90, 2017.

[27] Gu, J., Wang, Z., Kuen, J., Ma, L., Shahroudy, A., Shuai, B., Liu, T., Wang, X., Wang, G., Cai, J. and Chen, "Recent advances in convolutional neural networks," Pattern recognition, vol. 77, no. 1, pp. 354-377, 2018.

[28] Hu, Guosheng, Yongxin Yang, Dong Yi, Josef Kittler, William Christmas, Stan Z. Li, and Timothy Hospedales, "When face recognition meets with deep learning: an evaluation of convolutional neural networks for face recognition," in n Proceedings of the IEEE international conference on computer vision workshops, 2015.

[29] Nguyen, Hieu V., and Li Bai., "Cosine similarity metric learning for face verification," in In Asian conference on computer vision, Berlin, Heidelberg., 2010.

[30] El Khiyari, Hachim, and Harry Wechsler, "Age invariant face recognition using convolutional neural networks and set distances," Journal of Information Security, vol. 8, no. 3, p. 174, 2017.

[31] Van Dyk, David A., and Xiao-Li Meng, "The art of data augmentation," Journal of Computational and Graphical Statistics, vol. 10, no. 1, pp. 1-50, 2001.

[32] Hu, Junlin, Jiwen Lu, and Yap-Peng Tan., "Discriminative Deep Metric Learning for Face Verification in The Wild," in In Proceedings of the IEEE conference on computer vision and pattern recognition, 2014.

[33] Hu, Yiqun, Ajmal S. Mian, and Robyn Owens, "Sparse Approximated Nearest Points for Image Set Classification," In CVPR, pp. 121-128, 20 Jun 2011.

[34] Huang, Xixian, Xiongjun Zeng, Qingxiang Wu, Yu Lu, Xi Huang, and Hua Zheng., ".[Face Verification Based on Deep Learning for Person Trackingin Hazardous Goods Factories," Processes, vol. 10, no. 2, p. 380, 2022.

[35] Viriri, Serestina, and Jules R. Tapamo, "Integrating iris and signature traits for personal authentication using user-specific weighting. Sensors," Sensors, vol. 12, no. 4, pp. 4324-4338, 2012.

[36] Jain, Anil K., and Arun Ross, "Multi biometric systems," Communications of the ACM, vol. 47, no. 1, pp. 34-40., 1 Jan 2004.