

Age Invariant Face Recognition using Convolutional Neural Network

Mrudula Nimbarte, Kishor Bhoyar

IT Department, YCCE Nagpur, MH, India

Article Info

Article history:

Received Oct 29, 2017

Revised Dec 23, 2017

Accepted Apr 7, 2018

Keyword:

Aging model

AIFR

CNN

Deep learning

Face recognition

ABSTRACT

In the recent years, face recognition across aging has become very popular and challenging task in the area of face recognition. Many researchers have contributed in this area, but still there is a significant gap to fill in. Selection of feature extraction and classification algorithms plays an important role in this area. Deep Learning with Convolutional Neural Networks provides us a combination of feature extraction and classification in a single structure. In this paper, we have presented a novel idea of 7-Layer CNN architecture for solving the problem of aging for recognizing facial images across aging. We have done extensive experimentations to test the performance of the proposed system using two standard datasets FGNET and MORPH (Album II). Rank-1 recognition accuracy of our proposed system is 76.6% on FGNET and 92.5% on MORPH (Album II). Experimental results show the significant improvement over available state-of-the-arts with the proposed CNN architecture and the classifier.

Copyright © 2018 Institute of Advanced Engineering and Science.

All rights reserved.

Corresponding Author:

Mrudula Nimbarte,
Department of Information Technology,
Yeshwantrao Chavan College of Engineering,
Nagpur, MH, India.
Email: mrudula_nimbarte@rediffmail.com

1. INTRODUCTION

Recognizing an identity of a person from facial images is a challenging and interesting problem in many real-world applications. Variations in parameters like head pose, expressions, brightness and aging makes it difficult to recognize a person from the databases [1]-[3]. Many researchers are working on these parameters to improve the performance of the system for face recognition. Aging in itself, is very complicated process. It varies from person to person [4]. Again, it is highly dependent on factors like geographical location, living style, eating habits, use of cosmetics, physical and mental health etc. of a person [5]-[7]. So, considering all these factors, recognizing an identity of a person over a span of years is really difficult. It has many real-world applications like finding missing children, passport renewal system, driving license renewal system, finding criminals, providing securities to VIP's etc. [8]-[10].

One more important issue in this application is about the changes in the images of the same person over aging i.e. intra-subject variations and similarity in the images of other persons i.e. inter-subject similarities. This application basically consists of two categories: First is Age Invariant Face Recognition (AIFR) and second is Age Invariant Face Verification (AIFV). Face recognition is generally considered as multi-class problem and face verification is binary-class problem of classification. AIFR focuses on recognizing a single input image of a person with available images of the same person in gallery. Whereas, AIFV try to check the identity of two given images, whether they are same or different.

Though various methodologies have been proposed by researchers to solve this problem and improve recognition accuracy, there is still a scope to improve accuracy. In this paper, we have focused on Age Invariant Face Recognition problem by using Convolutional Neural Network (CNN), named as

AIFR-CNN. Deep learning using CNN has become very popular nowadays. It has the advantage that, it provides feature extraction and classification in a single structure. But, for deciding the details of the CNN architecture, there is no standard rule or any logic. Researchers proposed their methodologies using CNN also; it is focused only on their own architecture not on any specific reason behind using those details. These details include: (a) number of layers in the network, (b) sequence of these layers, (c) dimensions of the filters applied, and (d) number of neurons used etc. Hence, we also proposed our own methodology to design the CNN architecture for AIFR.

The main contributions of this paper are: (a) novel 7-layer CNN architecture for AIFR, and (b) the use of smaller image size of 32x32 pixels to reduce time and space complexity. The rest of the paper is organized as follows. Second section includes the related work done in this area. The next section i.e. third gives complete details of the proposed methodology for age invariant face recognition. It is followed by section four, for experimental details using both standard datasets FGNET and MORPH(Album II). Finally, fifth section presents conclusion.

2. RELATED WORK

This section presents the related work in this area. Some of the researchers focused their work on face identification or recognition problem and others on face verification problem. This problem is basically categorized in two types: Generative and Discriminative Methods. Generative methods need to develop synthetic images of the person at the required age and then perform matching of those images with given image. Discriminative methods need their own way for feature extraction and classification purpose so that two images of same person are matched.

2.1. Generative methods

Recently, the method in [11] presented hierarchical model based on two-level learning with new feature descriptor called as Local Pattern Selection (LPS) for solving the problem of aging face recognition. The method in [12], focused on the role of facial asymmetry in recognizing age-separated face images based on matching-score space (MSS). In [13], authors used minimal set of geometric features for age invariant face recognition. It was based on selected feature points and performance evaluated on FGNET dataset. Park *et al.* [14] proposed a generic method that consists of a 3-D aging model to improve the face recognition performance. They used pose correction step and separate modeling for shape and texture.

2.2. Discriminativem

Gong *et al.* [15] presented a novel feature descriptor named as maximum entropy feature descriptor (MEFD) to recognize age invariant face images. It is a discriminative feature descriptor. To improve recognition accuracy a new feature-matching framework is also presented as Identity Factor Analysis (IFA). Ali *et al.* [16] focused on a combination of shape and texture features for age-invariant face recognition. They adopted phase congruency feature for shape and LBP variance for texture feature. Bouchaffra [17] introduced a novel framework to reduce dimensionality and extracting topological features such as shape for age invariant face recognition. It is a combination of Kernelized Radial basis function (KRBF) for dimensionality reduction, construction of α -shape for feature extraction and mixture multinomial distributions for object classification.

Tandon *et al.* [18] attempted a novel approach using LBP of particular region as ROI for age invariant face recognition. Chi-square measure is used as a dissimilarity measure to calculate the distance between two histograms. Yadav *et al.* [19] presented a system to improve the results of face recognition across age progression by using bacteria foraging fusion algorithm. It reduces the aging effects by a combination of LBP features of global and local facial regions by using bacteria foraging fusion algorithm. Xiao *et al.* [20] presented a novel method for face recognition using a combination of texture and shape descriptors, called as Biview face recognition algorithm. For texture feature subspace learning methods are used and graph is constructed for shape topology for face images. Li *et al.* [21] proposed a discriminative approach for face recognition over aging. In this model, they used Scale-Invariant feature transform (SIFT) and Multi-scale Local Binary Patterns (MLBP) as feature descriptors and multiple LDA-based classifier to generate a decision via fusion rule. Ling *et al.* [22] proposed a discriminative method for face verification across age progression. In their study, they used Gradient Orientation (GO) and Gradient Orientation Pyramid (GOP) as feature descriptor and Support Vector Machine (SVM) as a classifier.

2.3. Using convolutional neural networks (CNN)

Recently CNN have become a very popular technique for Computer Vision applications. Many researchers used CNN for face recognition applications. In [23], a method is proposed using a fusion of 2-D

face images and motion history image (MHI) for face recognition based on 7-layer deep learning neural network. In [24], authors presented the novel use of deep learning using CNN for automatic feature extraction for robust face recognition across time lapse. They used VGG-very-deep 16 layer CNN architecture in their experiments. Li *et al.* [25], proposed a new deep CNN model for age-invariant face verification with 7-layer CNN architecture. Parkhi *et al.* [26] presented a model for face recognition either from a single image or from a series of faces traced from video. It was 11-layer architecture for face recognition. Sun *et al.* [27] proposed two very deep neural network architectures for face recognition named as DeepID3 net1 and DeepID3 net2. Half features from DeepID3 net1 and other half from net2 are concatenated into a long feature vector in this method. Hu *et al.* [28] proposed three CNN architectures and conducted extensive evaluation of CNN-based face recognition system.

These architectures are: small (CNN-S), medium (CNN-M) and large (CNN-L). They used LFW dataset for experimentation. Xinhua *et al.* [29] focused on face recognition problem using CNN on LFW dataset. They used Sobel operator to improve result accuracy. Taigman *et al.* [30] proposed a 9-layer deep neural network for face verification problem where they used alignment step and representation step to apply a piecewise affine transformation. Yi *et al.* [31] developed an effective representations for both face identification and verification with deep learning named as DeepID2. Many researchers presented their work on AIFR using various methods as discussed above, but only few studies reported for Age Invariant Face Recognition specially using Convolutional Neural Network.

3. PROPOSED METHODOLOGY FOR AGE INVARIANT FACE RECOGNITION

This section describes the proposed methodology for age invariant face recognition using Convolutional Neural Networks (AIFR-CNN). This network is designed for the recognition of the person having some aging variations. The overall process contains the same traditional steps: Image Preprocessing, Feature Extraction and Classification. Image preprocessing steps improve the performance of the system. We used three basic preprocessing steps. Feature extraction is the process of capturing the desired feature descriptors using CNN rather than extracting it manually. In this model, we used 7-layer CNN architecture. Classification is required to recognize the identity of the person. This work includes multi-class classification problem. The overall process for AIFR-CNN is shown in Figure 1.

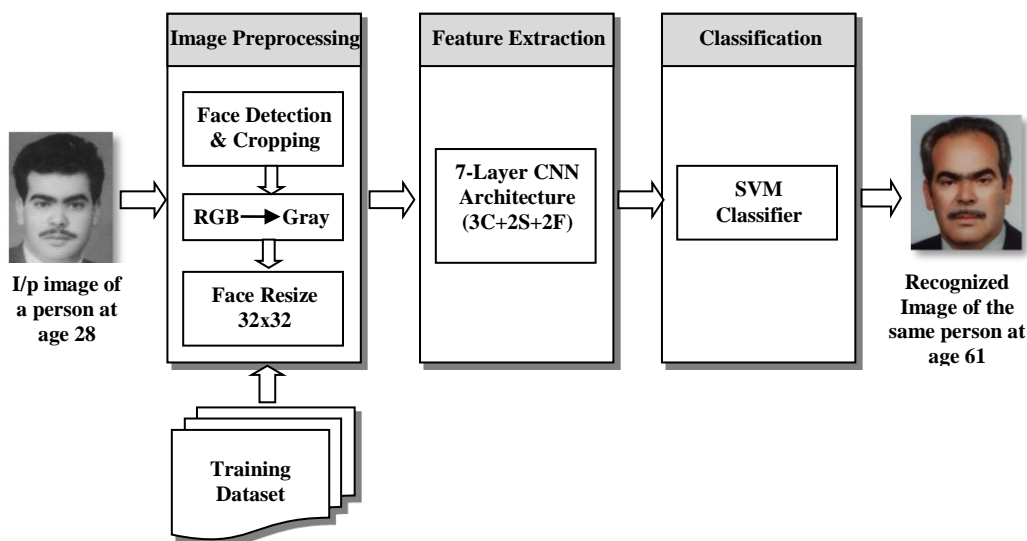


Figure 1. Basic Block Diagram for Proposed AIFR-CNN

3.1. Image preprocessing

Standard datasets may have images of different sizes and illumination. Hence, it may lead to some recognition problems. Image preprocessing helps to keep our dataset in normalized format. It includes detection and cropping of facial portion from the given image. For this purpose, we used popular Viola Jones algorithm for face detection. The next step is to convert the RGB image to gray scale image. Later, images are resized to 32x32 pixels and 64x64 pixels for some experiments. In this work, we have not performed any

complicated preprocessing steps as histogram normalization and head-pose correction. Figure 2 illustrates preprocessing steps on a sample image from FGNET dataset.



Figure 2. Stepwise Preprocessing for a Sample Image from FGNET Dataset

3.2. Network architecture for feature extraction

After basic preprocessing steps, the next step is to extract features as per our requirements. In our proposed work (AIFR-CNN), for this purpose we used deep learning approach using Convolutional Neural Network (CNN). It has many advantages. First, feature extraction and classification are concerns of CNN itself with single structure. Second, this network extracts deeper 2-D features. Third, it is fully adaptive and invariant to local and geometric changes in the image.

Three types of main layers are there in a CNN: (a) Convolution layer, (b) Pooling layer (Sub-sampling), and (c) Output layer. Feed-forward structure is used to arrange these layers in the network. Each convolution layer is followed by a pooling layer, whereas last convolution layer is followed by output layer. Convolution and pooling layers are 2-D layers whereas output layer is 1-D layer. Every 2-D layer of a CNN contains several planes. A plane of a 2-D layer consists of 2-D array of neurons. Feature map is the output of a plane. In AIFR-CNN, we proposed a 7-layer architecture that includes 3 convolution layers (C1,C3,C5), 2 pooling layers (P2,P4) and 2 fully connected output layers (F6,F7). The architecture for proposed methodology is shown in Figure 3.

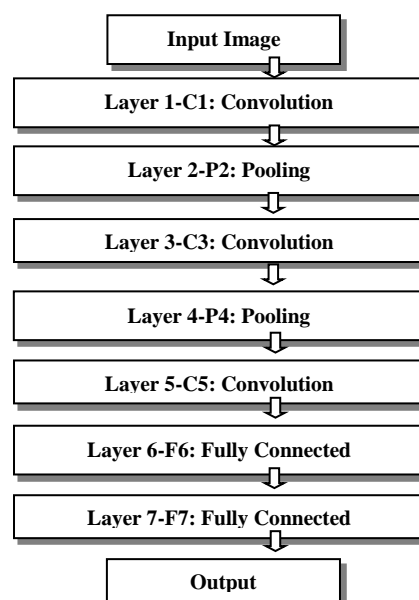


Figure 3. 7-Layer CNN Architecture for AIFR-CNN

3.2.1. Convolution layer

Each plane of convolution layer is associated with one or more feature maps of earlier layer. Convolution mask is used as an associated connection which is a 2-D weight matrix of adjustable entries. The convolution is computed in each plane between its 2-D inputs and its convolution mask. The outputs of convolution layers are added together with an adjustable scalar, called as bias. Lastly, an activation function is applied to obtain the planes output. This output of each plane is known as feature map. A convolution layer may have one or more maps. Each of these feature maps is connected to exactly one plane of next layer i.e. sub-sampling layer. Each plane in last convolution layer is associated with feature map of exactly one preceding layer. Each plane in the convolution layer produces one scalar output; these outputs from all planes are given to output layer. The purpose of this layer is to extract low-level features such as edges and texture. Feature map n of convolution layer l is calculated as:

$$y_n^l = f_l \left(\sum_{m \in V_n^l} y_m^{l-1} \otimes w_{m,n}^l + b_n^l \right) \quad (1)$$

Where $w_{m,n}^l$ is the convolution mask, b_n^l is the bias term, and V_n^l is the list of planes [32].

3.2.2. Pooling (sub-sampling) layer

The dimensionality of each feature map is reduced by spatial pooling by retaining the most valuable information. It can be of three different types: Max pooling - takes the largest element, Average pooling- takes the average of the elements, and Sum pooling - takes the sum of all the elements. The main function of pooling is to reduce the spatial size of the input representation progressively. It helps to make the input representations smaller and more convenient. A pooling and preceding convolution layers have the same number of planes. This result is then passed through the activation function to produce the outputs. This feature map is connected to one or more planes of the next convolution layer. It makes the output of convolution layer more robust to local distortions. Feature map of sub-sampling layer l is calculated as

$$y_n^l = f_l (Z_n^{l-1} \times w_n^l + b_n^l) \quad (2)$$

where Z_n^{l-1} is matrix obtained by summing all four pixels of a block, w_n^l is the weight and b_n^l is the bias term [32].

3.2.3. Output layer (fully connected layer)

In AIFR-CNN, the output layer is constructed from sigmoidal neuron. Generally, the outputs of this layer are the outputs of the network. In the output layer, softmax activation function is used by traditional multi layer perception. Other classifiers like SVM can also be used. These fully connected layers capture the correlations between features of various parts of the face like shape and location of eyes and mouth. The Convolution and Pooling layers in combination are used for feature extraction while fully connected layers are used for classification. The output of sigmoidal neuron n is calculated as

$$y_n^L = f^L \left(\sum_{m=1}^{N^{L-1}} y_m^{L-1} w_{m,n}^L + b_n^L \right) \quad (3)$$

where N^L is the number of output sigmoidal neurons, $w_{m,n}^L$ is weight from feature map m of the last convolution layer to neuron n of the output layer, and b_n^L is the bias of neuron n associated with layer L [32].

3.2.4. 7-Layer Architecture for AIFR-CNN

In our implementation, we used 7-layer CNN architecture for age invariant face recognition as shown in Figure 4. This network architecture consists of sequence of convolution, sub-sampling and fully connected output layers. The convolution layers use filters of 5x5 whereas sub-sampling with 2x2. The input to this network is 32x32 pixels grayscale image. This image used for performing convolution with a filter of 5x5 pixels in size. Convolution is a linear operation and performs element wise matrix multiplication and addition.

The filtered image of 28x28 pixels is obtained by convolving a 5x5 filter with 32x32 pixels image. It has 6 distinct planes. These 6 planes will generate 6 separate feature maps as output of first convolution layer C1 as 28x28x6 matrix. Layer 2 is sub-sampling layer S2. The pooling operation we used is summing and regions are 2x2 pixels. It results in reduced feature map by a factor of 2 in both dimensions and we obtain 14x14x6 matrix. Next layer is another convolution layer C3, and we applied a filter of same size of 5x5. In this layer, we have 16 distinct planes of 10x10. Layer 4 is again sub-sampling layer S4 with

2x2 block reduces feature map by factor of 2 and gives matrix of size 5x5x16. Last convolution layer C5 produces feature map of 120 scalar values. Finally, the next layers are two fully connected layers (F6 and F7) where each output unit is connected to all inputs. F6 contains 84 and F7 contains 10 neurons. The output of last fully connected layer F7 is provided to the classifier.

Table 1 shows the details of the CNN architecture used earlier for face recognition problem. Some of them are only for face recognition and verification while some of them are for age invariant face recognition. It shows all the details as size of input image taken, number of layers used in the architecture, datasets used and length of feature vector. From this table it is observed that some work has been done for face recognition, but for age invariant face recognition there is still a lot of scope to improve the performance. Again, variations are there in each architecture in size of image and number of layers in the architecture. So, the size of feature vector varies. And, there is no specific reason mentioned behind selecting these parameters.

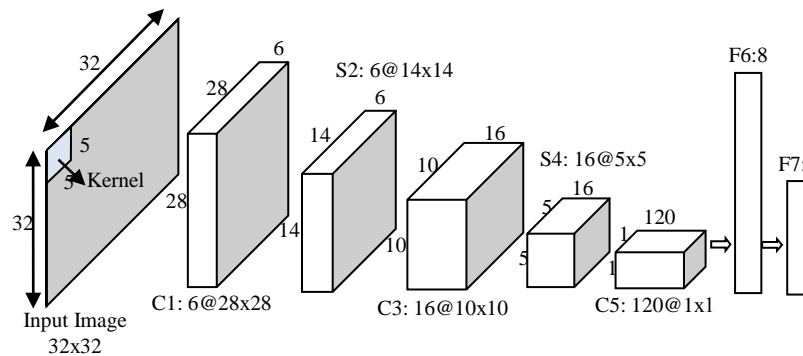


Figure 4. 7-Layer Architecture for AIFR using CNN

Table 1. Comparative Details of State-of-the-Arts for Face Recognition using only CNN

	Input Image ¹	Architecture ²	Dataset ³	Feature Vector
Face Recognition/Verification using CNN				
DeepFace [30]	152x152x3	7 Layer (2C+1M+2L+1F)	SFC, LFW, YTF	4096
Deep Face Recognition [26]	224x224x1	14 Layer (8C+3M+3F)	LFW, YTF	4096
DeepID	39x31x1	8 Layers (4C+3M+1F)	LFW	160
DeepID2[31]	55x47x3	8 Layers (4C+3M+1F)	LFW	160
DeepID3[27]	-	(10C+4P)	LFW	300
When FR meets with Deep Learning [28]	58x58x1	CNN-S(3C+1F) CNN-M(3C+1F) CNN-L(4C+1F)	LFW	160
FR based on Deep Neural Network [29]	60x48x1	6 Layer (2C+2P+2F)	LFW	-
Fusion FR based deep learning [23]	100x100x1	7 Layer	Private	20000
Age-Invariant Face Recognition/Verification using CNN				
FR across time lapse using CNN [24]	224x224x1	16 Layer (16C+ 5P+3F)	FGNET, MORPH	4096
Deep Joint Learning approach for AIFV [25]	180x130x3	7 Layer (2C+2M+3F)	MORPH	400
Proposed AIFR-CNN	32x32x1	7 Layer (3C+2M+2F)	FGNET, MORPH	961

¹Input image is represented as Width x Height x Channels, 1 and 3 mean Gray and RGB images respectively.

²Uppercase letters C, P, M, L and F represent Convolutional, Pooling, Max-Pooling, Locally connected and fully connected layers respectively.

³LFW- Labeled Faces in the Wild dataset, YTF-You Tube Faces, SFC-Social Face Classification, FGNET-Face and Gesture feature recognition NETWORK, MORPH-MORPH (album II).

3.3. Classification techniques

In this work, we use multi-class Support Vector Machine (SVM) as a classifier for the identification of a person over long period. It is a supervised learning algorithm as data labels are available. They are effective in high dimensional spaces. It is also memory efficient and versatile in nature. In another

experiment, we use Euclidian Distance with Nearest Neighbor (NN) as a classification rule for recognizing a person across aging.

4. EXPERIMENTAL DETAILS

In this section, we describe the implementation and experimental details for AIFR-CNN. In our experiments, we used LOPO(Leave One Person Out) scheme for testing, where one person from dataset is kept out for testing. In our earlier experiments, we used 3-fold cross validation that needs to keep 3 images of the same person at different ages in three separate folders. In this approach, instead of keeping one image of the person in testing folder, we kept two images of the same person having a gap of at least 10 years. The reason behind doing so is to avoid repetitive testing using different folders. As all images are in a single testing folder, all persons are considered as different individuals. Remaining images of the person are considered for training, in order to avoid the same person in both the folders. We use Rank-1 recognition rate as a performance evaluation parameter. The experiments are performed on MATLAB 2015a(64-bit) with 2.60 GHz Intel(R) CORE(TM) i-5 CPU and 8 GB of RAM.

4.1. Datasets

We use two publicly available datasets FGNET [33] and MORPH (Album II) [34] for AIFR-CNN. Both the datasets contain many images of the same person having variation in age, expressions, illumination and head position. FGNET consists of 1002 images of 82 subjects. It includes 12 images of a person in average. It has age range between 0 to 69 years. MORPH Album II contains more than 55000 images of 13000 subjects. It includes age range between 16 to 77 years. Figure 5 and Figure 6, show some sample images where we can observe the variations in illumination, expression, head position and age in both the datasets.



Figure 5. Sample images from FGNET dataset with age values [33]



Figure 6. Sample images from MORPH (Album II) dataset with age values [34]

4.2. Experiments on FGNET dataset

In our experiments, we used total 980 images of 82 subjects from FGNET dataset for AIFR-CNN among which 852 images are used for training and 128 for testing. We performed various experiments using this dataset. Firstly, we used images of size 32x32 after performing all preprocessing steps. These images include head pose variations also. We used Rank-1 recognition as a performance measuring parameter. From the results obtained, it is found that 98 images from 128 are truly recognized. It indicates 76.6% Rank-1 recognition. Secondly, we used the same procedure and same number of images but with different image size 64x64. In this experiment 87 images from 128 got correctly recognized. It shows 68.8% Rank-1 recognition accuracy. It may be because of the network architecture that is not enough capable for this image size. In the next experiment, we used only straight pose(frontal) images. We eliminated non-frontal images from our

dataset. Using FGNET dataset we extracted 654 frontal images for training and 96 frontal images for testing. In this experiment, we used resized images of 32x32 pixels as it has comparative good performance. In this case, 61.2% Rank-1 recognition is obtained for age invariant face recognition using proposed methodology. In the last experiment, we tested our proposed system without using SVM as a classifier. Here for classification, we used Euclidian Distance with Nearest Neighbor (NN) classification rule. As the image of 32x32 pixels gives better results, we used the same size in this experimentation. Here, we get Rank-1 recognition as 75%. Table 2, Table 3 and Table 4 show this comparative analysis for Rank-1 recognition using FGNET dataset.

Table 2. Comparative Rank-1 Recognition on different image sizes from FGNET Dataset
(All images with variation in head pose)

Image Size	Training Images	Testing Images	Rank-1 Recognition
32x32	852	128	76.6%
64x64	852	128	68.8%

Table 3. Rank-1 Recognition on Only Frontal Images from FGNET Dataset

Image Size	Training Images	Testing Images	Rank-1 Recognition
32x32	654	96	61.2%

Table 4. Comparative Rank-1 Recognition with SVM/NN from FGNET Dataset
(All images with variation in head pose)

Image	Training Images	Testing Images	Rank-1 Recognition
32x32 with SVM	852	128	76.6%
32x32 with NN	852	128	75%

Figure 7 shows correctly recognized results for some sample images using AIFR-CNN on FGNET dataset. First column shows the images used for testing and remaining columns show the images of the same person at different ages available in training folder. We used SVM, the supervised learning algorithm for classification as labels are available, it shows the class label to which test image belongs. It is observed from the results that AIFR-CNN is capable to recognize images of the same person at different ages. Hence, it is one of the good approaches. From Figure 7, we can see FGNET dataset contains many images of the same subject with large age gap and for more age gap more variations are there in the images. Figure 8 shows the Rank-1 results of some images other than the images from standard dataset also. For this, we added our sample images and tested on AIFR-CNN.

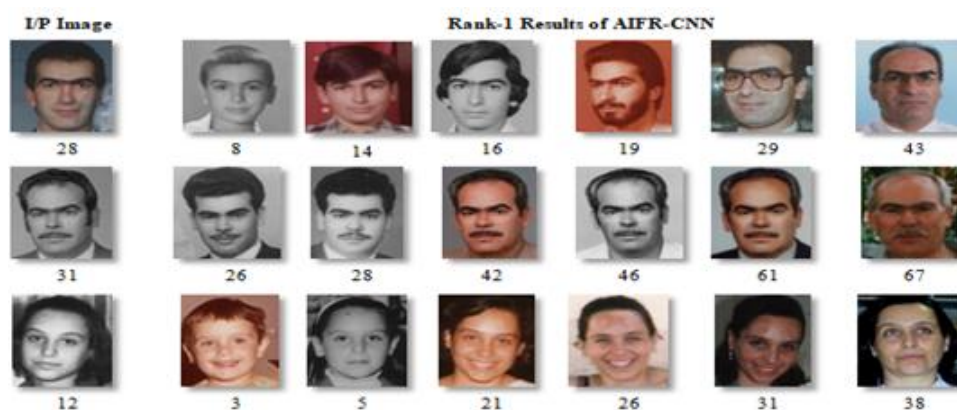


Figure 7. Some Correctly Recognized results using proposed method on FGNET Dataset. The numbers below these images show the age of the person



Figure 8. Some Correctly Recognized results using proposed method on our own images

4.3. Experiments on MORPH (Album II) dataset

For proposed age invariant face recognition using CNN, we used another publicly available dataset, MORPH(Album II). Using this dataset, we performed three experiments. Firstly, we have used 1005 images of 255 subjects with all head poses. Among these images 750 images are used for training and 255 images used for testing. Results show 92.5% Rank-1 recognition using CNN. Secondly, we used 2084 frontal images of 575 subjects. 1509 images for training and 575 images for testing. In this case, we obtain 92.8% Rank-1 recognition. In the last experiment, as in FGNET dataset, we tested performance using CNN with Euclidian Distance and Nearest Neighbor(NN) as classifiers. Here we obtain 91.3% Rank-1 recognition. Table 5, Table 6, Table 7 and Table 8 demonstrate this comparison.

Table 5. Rank-1 Recognition using MORPH (Album II) Dataset with SVM (All Images)

Images	Training Images	Testing Images	Rank-1 Recognition
All(Frontal+ Non-frontal)	750	255	92.5%

Table 6. Rank-1 Recognition using MORPH (Album II) Dataset with SVM (Only Frontal Images)

Images	Training Images	Testing Images	Rank-1 Recognition
Only Frontal	1509	575	92.8%

Table 7. Comparative Rank-1 Recognition using MORPH (Album II) Dataset with NN

Image Size	Training Images	Testing Images	Rank-1 Recognition
32x32	750	255	91.3%
64x64	750	255	90.2%

Table 8. Comparative Rank-1 Recognition using MORPH (Album II) Dataset with SVM and NN

Image Size	Training Images	Testing Images	Rank-1 Recognition
32x32 with SVM	750	255	92.5%
32x32 with NN	750	255	91.3%

Figure 9 shows some correctly recognized results for AIFR-CNN on MORPH(Album II) dataset. From this figure, we observe that there is no more variation in the age of the person as compared to FGNET dataset. Secondly, MORPH dataset contains less number of images per person.



Figure 9. Sample Correctly Recognized results using proposed method on MORPH Dataset

Figure 10 shows the Cumulative Match Characteristic(CMC) curve of the proposed AIFR-CNN with different methods as mentioned using FGNET and MORPH (Album II) dataset. Figure 11 shows the comparative performance analysis of proposed AIFR-CNN over FGNET and MORPH dataset.

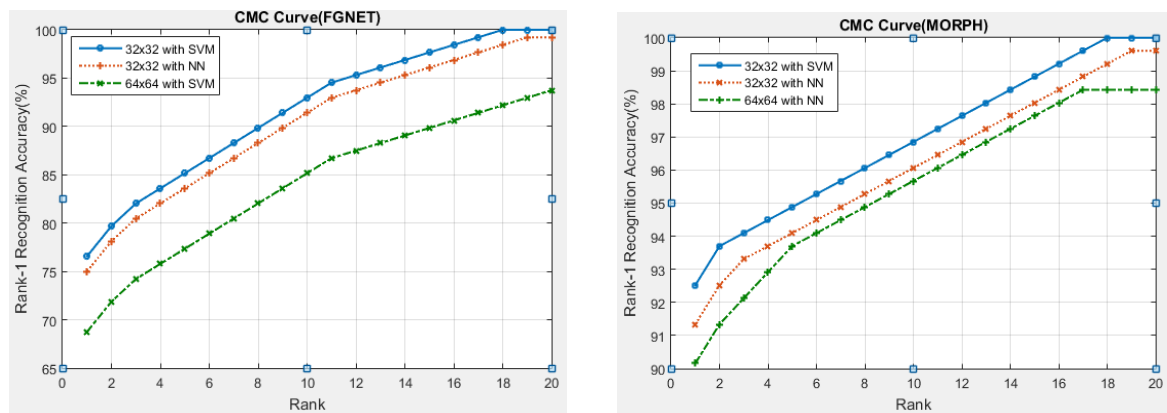


Figure 10. Cumulative Match Characteristic (CMC) Curves with different methods on Both Datasets

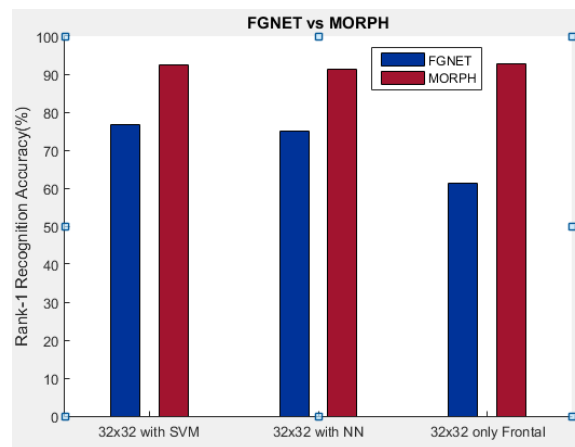


Figure 11. Comparative Performance Analysis (Rank-1 Recognition) of FGNET Vs MORPH Dataset

4.4. Overall comparative discussions

Here in this section, we compare our proposed methodology with existing state-of-the-arts. Face recognition is very vast area in the field of image processing and pattern recognition. There are various parameters that make it really difficult like variations in head position, facial expressions and aging effects. Many researchers proposed their methodologies for solving the problem of recognizing facial images across aging. It generally includes the steps: face detection, preprocessing, feature extraction and classification. Performance of the system is directly dependent on algorithms used for feature extraction and classification. But, the beauty of Convolution Neural Networks is that, it provides feature extraction and classification in a single structure. Although CNN is a very powerful tool, it makes difficult to decide number of layers, number of neurons, and the size of input image provided to CNN architecture. Unfortunately, there is no way or formula available. Nobody focused on these issues rather they proposed architecture by their own way. We also follow the same process to decide number of layers, their dimensions, and size of the image provided as input to CNN.

Figure 12 and Figure 13 show some of the failed Rank-1 retrieval results from FGNET and MORPH (Album II) respectively. First row shows the input images used for testing, second row shows the output of our method i.e. failed to recognize correctly and third row shows the ground truth images available in the gallery. It is seen from the results that, there are more intra-class differences and inter-class similarities in both the datasets. Manually also, it is difficult to identify the persons, as some of them look similar to others.



Figure 12. Some examples of Rank-1 failed retrievals from FGNET dataset. First row shows input images, the second row is the rank-1 results of our method, and the third row is the ground-truth, i.e. correct matched images available



Figure 13. Some examples of Rank-1 failed retrievals from MORPH dataset. First row shows input images, the second row is the rank-1 results of our method, and the third row is the ground-truth, i.e. correct matched images available

In this study, we used images after preprocessing, of size 32x32. To the best of our knowledge, this image size is used for the first time to recognize facial images across aging. It needs less time for computation and space for a large database. So, compared to other studies, it gives better performance on both the datasets as shown in Table 9. We also performed other experimentations on: (a) all images (Frontal and Non-frontal), (b) Only Frontal images on both datasets, and (c) same architecture with SVM and NN as a classifier. The results are tested on FGNET with 980 images of all 82 subjects, on MORPH (album II) with 1005 (Frontal and Non-frontal) and 2084 (Only Frontal) images.

Table 9. Comparative Analysis of AIFR-CNN with State-of-the-Arts

FGNET Dataset		MORPH(Album II) Dataset	
Methods	Rank-1 Recognition (%)	Methods	Rank-1 Recognition (%)
NTCA [17]	48.96	Facial Asymmetry [12]	69.40
Graph based view [35]	64.47	NTCA [17]	83.80
MDL [36]	65.2	HFA [38]	91.14
PCA & WLBP [37]	67.30	MDL [36]	91.8
HFA [38]	69.0	CNN [24]	92.2
Facial Asymmetry [12]	69.51	MEFD [39]	92.26
MEFD [37]	76.2	Proposed using CNN	92.5
Proposed using CNN	76.5		

From the experimentations, it is observed that,

- a. CNN is better in Rank-1 recognition than available methods with no complicated preprocessing steps like histogram normalization and head pose correction.
- b. It gives better performance on 32x32 image size as compared to 64x64 image size. As we increase the image size, it needs more time for execution.
- c. It gives better result on MORPH dataset over FGNET dataset as it contains less age variant images. Moreover, FGNET contains large intra-personal differences and MORPH contains less inter-personal similarities.
- d. It gives better results with SVM, as it is supervised learning algorithm over NN as a classifier.
- e. There is not much difference in the performance of CNN, if we consider only frontal images and exclude non-frontal images.

5. CONCLUSION

In this paper, we proposed a novel methodology for age invariant face recognition using Convolutional Neural Network named AIFR-CNN. Experimentation has been performed on two image datasets FGNET and MORPH-II. In this approach, our goal is to provide a simple network by using less number of layers, small image size(32x32) for processing. This system preserved simplicity as no separate algorithm is required for feature extraction. The results have demonstrated that it is better than current state-of-the-arts in Rank-1 recognition on both the datasets. Moreover, no complicated preprocessing steps are used for head pose correction. Resized images of size 32x32 pixels show better results as compared to images of size 64x64 pixels on both datasets. AIFR-CNN with SVM as a final classification stage, shows significant improvement in the performance over AIFR-CNN with NN as a final classification stage.

REFERENCES

- [1] Petra G. R. D., "Introduction to Human Age Estimation using Face Images," Research Papers, Faculty of Materials Science And Technology in Trnava Slovak University of Technology in Bratislava, Slovak University of Technology, 2013.
- [2] G. Mahalingam and K. Ricanek, "LBP-based Periocular Recognition on Challenging Face datasets," *EURASIP Journal on Image and Video Processing*, 2013.
- [3] P. Thukral, *et al.*, "A Hierarchical Approach For Human Age Estimation," *IEEE International Conference on Acoustics, Speech and Signal Processing*, pp. 1529-1532, 2012.
- [4] M. Bereta, *et al.*, "Local descriptors in application to the aging problem in face recognition," *IEEE Transactions on Pattern Recognition*, vol. 46, no. 10, pp. 2634-2646, 2013.
- [5] T. H. Le, "Applying Artificial Neural Networks for Face Recognition," *Hindawi Publishing Corporation Advances in Artificial Neural Systems*, vol. 2011, pp. 1-16, 2011.
- [6] J. M. Guo, *et al.*, "Human Face Age Estimation with Adaptive Hybrid Features," *International Conference on System Science and Engineering*, 2011.
- [7] Y. Fu, *et al.*, "Age Synthesis and Estimation via Faces: A Survey," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 32, no. 11, pp. 1955-1976, 2010.
- [8] D. Hunter and B. Tiddeman, "Facial Ageing," Cambridge University Press, 2012.
- [9] J. Suo, *et al.*, "A Concatenational Graph Evolution Aging Model," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 34, no. 11, pp. 2083-2096, 2012.
- [10] A. Lanitis, *et al.*, "Toward Automatic Simulation of Aging Effects on Face Images," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 4, pp. 442-455, 2002.
- [11] Zhifeng L., *et al.*, "Aging Face Recognition: a Hierarchical Learning Model Based on Local Patterns Selection," *IEEE Transactions on Image Processing*, vol. 25, no. 5, pp. 2146-2154, 2016.
- [12] M. Sajid, *et al.*, "The Role of Facial Asymmetry in Recognizing Age-Separated Face Images," *Journal of Computers & Electrical Engineering*, pp. 1-12, 2016.
- [13] S. Bijarnia and P. Singh, "Age Invariant Face Recognition using Minimal Geometrical Facial Features," *Chapter in Advanced Computing and Communication Technologies*, vol. 452, pp. 71-77, 2016.
- [14] U. Park, *et al.*, "Age-Invariant Face Recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 32, no. 5, pp. 947-954, 2010.
- [15] D. Gong, *et al.*, "A Maximum Entropy Feature Descriptor For Age Invariant Face Recognition," *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2015.
- [16] A. S. O. Ali, *et al.*, "Age-Invariant Face Recognition System Using Combined Shape And Texture Features," *IET Biometrics*, vol. 4, no. 2, pp. 98-115, 2015.
- [17] Bouchaffra D., "Nonlinear Topological Component Analysis: Application to Age-Invariant Face Recognition," *IEEE Transaction on Neural Networks and Learning Systems*, 2014.
- [18] Tandon A., *et al.*, "An Efficient Age-Invariant Face Recognition," *International Conference on Software Intelligence Technologies and Applications & International Conference on Frontiers of Internet of Things*, 2014.

- [19] Yadav D., *et al.*, "Bacteria Foraging Fusion for Face Recognition across Age Progression," *IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pp. 173-179, 2013.
- [20] B. Xiao, *et al.*, "Biview Face Recognition in the Shape-Texture Domain," *Elsevier Journal on Pattern Recognition*, vol. 46, no. 7, pp. 1906-1919, 2013.
- [21] Zhifeng L., *et al.*, "A Discriminative Model for Age Invariant Face Recognition," *IEEE Transactions on Information Forensics And Security*, vol. 6, no. 3, pp. 1028-1037, 2011.
- [22] H. Ling, *et al.*, "Face Verification Across Age Progression Using Discriminative Methods," *IEEE Transactions on Information Forensics and Security*, vol. 5, no. 1, pp. 82-91, 2010.
- [23] J. Liu, *et al.*, "A Fusion Face Recognition Approach Based on 7-Layer Deep Learning Neural Network," *Journal of Electrical and Computer Engineering*, pp. 1-7, 2016.
- [24] H. E. Khiyari and H. Wechsler, "Face Recognition across Time Lapse using Convolutional Neural Networks," *Journal of Information Security*, vol. 7, no. 3, pp. 141-151, 2016.
- [25] Y. Li, *et al.*, "A Deep Joint Learning Approach for Age Invariant Face Verification," *Chapter in Computer Vision*, vol. 546 of the series Communications in Computer and Information Science, pp. 296-305, 2015.
- [26] O. M. Parkhi, *et al.*, "Deep Face Recognition," *British Machine Vision Conference*, pp. 1-12, 2015.
- [27] Sun Y., *et al.*, "Deepid3: Face Recognition with Very Deep Neural Networks," *arXiv preprint arXiv: 1502.00873*, 2015.
- [28] Guosheng H., *et al.*, "When Face Recognition Meets with Deep Learning: an Evaluation of Convolutional Neural Networks for Face Recognition," *arXiv preprint arXiv:1504.02351*, pp. 4321-4329, 2015.
- [29] L. Xinhua and Y. Qian, "Face Recognition based on Deep Neural Network," *International Journal of Signal Processing, Image Processing and Pattern Recognition*, vol. 8, no. 10, pp. 29-38, 2015.
- [30] Y. Taigman, *et al.*, "DeepFace: Closing the Gap to Human-Level Performance in Face Verification," *International Conference on Computer Vision and Pattern Recognition (CVPR)*, 2014.
- [31] S. Yi, *et al.*, "Deep Learning Face Representation by Joint Identification-Verification," *arXiv:1406.4773v1*, 2014.
- [32] S. L. Phung and A. Bouzerdoun, "MATLAB Library for Convolutional Neural Networks," *Technical Report, Visual and Audio Signal Processing Lab University of Wollongong*, 2009.
- [33] The FG-NET Aging Database, <http://www.fgnet.rsunit.com>.
- [34] MORPH Non-Commercial Release Whitepaper, <http://www.faceaginggroup.com>
- [35] Yang H., *et al.*, "Age Invariant Face Recognition Based on Texture Embedded Discriminative Graph Model," *IEEE International Joint Conference on Biometrics*, 2014.
- [36] D. Sungatullina, *et al.*, "Multiview discriminative learning for age-invariant face recognition," *10th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition (FG)*, 2013.
- [37] Patel P. and Ganatra A., "Investigate Age Invariant Face Recognition Using PCA, LBP, Walsh Hadamard Transform with Neural Network," *Elsevier International Conference on Signal and Speech Processing (ICSSP-14)*, pp. 266-274, 2014.
- [38] D. Gong, *et al.*, "Hidden Factor Analysis for Age Invariant Face Recognition," *IEEE International Conference on Computer Vision*, pp. 2872-2879, 2013.
- [39] Yang H., *et al.*, "Age Invariant Face Recognition Based on Texture Embedded Discriminative Graph Model," *IEEE International Joint Conference on Biometrics*, 2014.

BIOGRAPHIES OF AUTHORS



Mrudula Nimbarte received B.E. degree in Computer Science and Engineering from Govt. College of Engg. Amravati in 2001 and M.E. degree from PRMIT&R Amravati (Badnera) in 2009. She is pursuing Ph.D. from YCCE, Nagpur. She is currently working as Assistant Professor in the dept. of Computer Engineering at BDCE, Sevagram (Wardha), India. She has more than 15 years of teaching experience. She has published many research papers in different National and International Conferences and Journals. Her areas of interests include Image Processing and Pattern Recognition.



Dr. Kishor K. Bhoyar completed B.E. in Computer Science and Engineering in 1990 and M.Tech. in Computer Technology in 2001, both from SGGSIET Nanded, an autonomous Institution funded by Govt. of India. He has been awarded Ph.D. degree in Computer Science & Engg. from VNIT, Nagpur in July 2010, for his research work titled 'Performance Enhancement of Color Based Classification and Segmentation for Image Retrieval using JND Approach'. He is presently working as Professor, in the dept. of Information Technology at Yeshwantrao Chavan College of Engineering, Nagpur (India). His areas of interests include Image Processing and Machine Learning.