JOURNAL OF MECHANICS OF CONTINUA AND MATHEMATICAL SCIENCES
www.journalimcms.org

ISSN (Online) : 2454 -7190 Vol.-15, No.-2, February (2020) pp 118-128 ISSN (Print) 0973-8975

PREDICTION OF GENDER FROM FACIAL IMAGE USING DEEP LEARNING TECHNIQUES

Ramalakshmi K[1], T. Jemima Jebaseeli[2], Venkatesan R[3]

[1,2,3]Assistant Professor, Department of Computer Science and Engineering, Karunya Institute of Technology and Sciences, Coimbatore, Tamil Nadu, India.

[1]jemima_jeba@karunya.edu, [2]ramalakshmi@karunya.edu, [3]rlvenkei_2000@karunya.edu

Corresponding Author: T. Jemima Jebaseeli

https://doi.org/10.26782/jmcms.2020.02.00010

Abstract

Gender recognition is a process of recognizing a person's gender from their facial image using deep learning. The posed variation, illumination, and occlusion are some of the factors that affect in recognizing faces. These are reduced by increasing the accuracy of prediction. The network used for training the system is Convolutional Neural Network (CNN). For improving accuracy, the faces are detected and cropped from the image. Face detection is done using Open CV which detects the face by the frontal features of the face. This is done during training the network. The dataset used for training has cropped images. The proposed system predicts the person's gender without compromising accuracy.

Keywords: Gender recognition, convolutional neural network, VGGNet.

I. Introduction

Gender recognition is an active area of research in machine learning. It has a wide range of applications such as man-machine communication, security, surveillance, law enforcement, demographic studies, education, and telecommunication [XI], [XIX], [XVI], [XX].Gender classification can be done by reading facial features, body language or even from the voice of a person [IV], [VII], [XIII], [XIV], [XVII], [II]. Classification requires training the network with a dataset. Training is done using an artificial neural network which is similar to the neural system in the human brain.

Machine learning is the process of training software to predict more accurately by learning from the data given without being explicitly programmed. It is categorized into supervised and unsupervised algorithms. The supervised learning uses known set of input and output data and train the model to generate the expected prediction. The unsupervised learning finds hidden pattern or it doesn't need to be trained with the

Copyright reserved © J. Mech. Cont.& Math. Sci.
Ramalakshmi K et al

expected output data. Clustering is the technique used for data analysis to find the pattern and grouping it in unsupervised learning. Gathering data, preparing the data, choose the model, training the model, evaluating the model, parameter tuning, and prediction are the seven steps in machine learning. The data is converted into a table during gathering and divided into two groups: one for training and another for testing/evaluating. The best model for performing the task has to be identified. In this stage, the model will be initialized with some values; it is then compared with the prediction to adjust the values accordingly. This step is repeated and updated; each cycle updated is called one training cycle. This is then evaluated by testing the model. The evaluated parameters are tuned for improving training. This model is used for predicting the outcome.

Deep learning is a subset of machine learning which is inspired on function of a brain with artificial neural networks [I]. This technology is used in driverless car, phones, tablets, and hands-free speakers. It will learn to classify the image, text or sound which are stored in a dataset used for training. It requires a large amount of labelled data and computational power. All deep learning technique uses neural networks and hence it is called deep neural networks [XII]. The main advantage of this is to improve the training accuracy as the data size increases.

There are mainly three steps in gender recognition: pre-processing, feature extraction, and classification. Pre-processing includes resizing, cropping, and normalization etc. The feature extraction is done using the network built and classification model is done using the classification algorithm such as SVM [VI], CNN [V], and K-Nearest Neighbour etc.

There are several challenges in gender recognition which makes face image analysis difficult such as pose, scaling, image capturing factors like blurring, noise, and resolution [VIII], [XX]. Even, age can make a difference in the classification accuracy. The adult face will give more efficient results than kids.

The proposed network is created using tensor flow and image processing is done using opencv. Tensor flow is used to setup, train and deploy neural network with large dataset. Opencv will detect the face and process the image by converting the image to its required input model to predict the output.

II. Literature Review

The method used for gender recognition or face recognition is by extracting the feature using an algorithm and classifying using a classifier. The extracted features will be given to the classifier. As shown in fig.1, the feature extraction is classified into two, geometry based and appearance based.

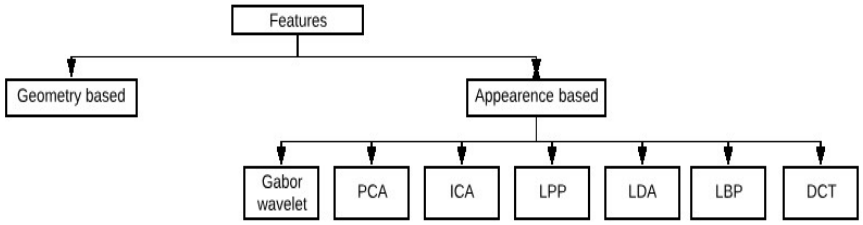


Fig.1: Classification of feature extraction.

Copyright reserved © J. Mech. Cont.& Math. Sci.
Ramalakshmi K et al

The various methods used for extracting the facial features are explained as follows based on the exiting literature.

Geometrical Based Extraction

In this method features are extracted based on various features of face such as eyes, nose, chin, and eye brow. These features are also called as local features. From the position of these features the vectors are created. The feature points are selected depending on the image to make extraction reliable and more significant. This method can be used even if there is any position variation in the image, since it identifies the geometrical relationship by locating the feature points [XX].

Appearance Based Extraction

In this method, the global feature of face is used. The features are extracted based on the pixels in the image. There are many algorithms for appearance based extraction such as Principle Component Analysis (PCA), Linear Discriminant Analysis (LDA), and Independent Component Analysis (ICA), etc.

Principle Component Analysis (PCA)

This method uses algebraic methods for extracting feature. It recognizes the pattern and uses the data to predict the similarities and dissimilarities. PCA is used in face recognition after converting a 2-Dimensional image into a 1-Dimensional vector for moving it into a feature space. This vector defines a face space which is a part of the image. The advantages of PCA areto compresses the data along with reducingdimension. It also has less data loss [XV].

Linear Discriminant Analysis (LDA)

This is used mainly in image processing and machine learning techniques to find the features to separate it to its specific class it belong to. This deals with only linear data. It converts the original data space into low dimensional feature space where data is separated. For good extraction, the dataset given for training should be very large. The advantages of LDA are that it works better with different illuminations than ICA and it is sensitive to occlusion [II], [X], [IX].

Independent Component Analysis (ICA)

This method converts signals into additive subcomponents. These signals are non-Gaussian signals which are statically independent from each other [XX].There are different types of classifiers used for extracting feature. The extracted feature will be given to the classifier where the features will be further processed for classification. Each technique will have different accuracy rate and processing method. There are different hand crafted classifiers used in gender recognition. The min-max modular support vector machine[VI], KNN classifier, and similarity measurements are some of the existing classifier methods.

Min-Max Modular Support Vector Machine

In this method, first the facial features from the image are extracted and then it divides the training data into smaller subsets. The facial features are extracted using facial point detection and Gabor wavelet transform method. The training data is divided into smaller sets using ‘part-versus-part’ task decomposition method [VI]. The advantage of this task decomposition is that it will have a prior knowledge about the data, and depending on that the data will be decomposed. Similar to the traditional SVM[V], M^3-SVM it will take the data as a one class during training and after that it is divided into small subsets for each class. It implements parallel training easily and

Copyright reserved © J. Mech. Cont.& Math. Sci.
Ramalakshmi K et al

can solve pattern classification efficiently. But one of the main problems is identifying the optimal method for decomposition. The imbalanced dataset due to decomposition leads to performance degradation. The lack of prior knowledge, affect the accuracy since it will have issues regarding creating subsets.

K-Nearest Neighbour Classifier

K-Nearest Neighbour (KNN) classifier is used to identify and classify thedata in the database. SVM is more accurate than KNN model. In this process, the test set is identified from the label assigned to it. The label is given in such a way that it is closer to the learning set and distance is measured in the image space. This distance is measured by the Euclidean metrics and it is called the Euclidean distance between two pixels [III].KNN classifies the pixels using the training data instead of learning from the data. Another problem is finding K value for closest neighbouring points and it does not work for a large set of data.

Similarity Measurement Classifier

In similarity measurement classifier, the similarity measure or the distance is used to compare with the samples. This method is used to improve KNN classification. Descriptors like Scale-Invariant Feature Transform (SIFT) and Local Binary Pattern are used to improve the performance of similarity measurement. A wide variety of similarity measures can cause confusion and difficulties in choosing suitable measures. This can be used only in a high dimensional datasets with hierarchical approach. The similarity measurement does not require direct access to the features of the sample. The similarity functions are asymmetric and fail to satisfy the mathematical properties which are required for metrics [XIX].

III. The Proposed System

The proposed system is implemented using Tensorflow and OpenCV. Tensorflow is used to build the neural network and opencv is used for image processing. The neural network is trained for extracting the features from the images. During the feature extraction, the filters in the convolutional layer will extract the local features at first and then the high level features are extracted using the neural network and this makes the comparison easier as well as improves efficiency. Activation function used in the network is ReLU (Rectified Linear Unit) which activates only if the input data is zero or a positive value. As more convolutional layer is added, more features are extracted and accuracy will increase which reduces the loss in data while training and testing.

As shown in fig. 2, the raw input is pre-processed by reducing the dimensions and resizing it. After that the image will be trained in the neural network. The Convolutional Neural Network used here is called VGGNet. It consists of multiple layers of convolutional layer, pooling layer and an activation functions. As the image goes through each layer, the features are extracted from the image.

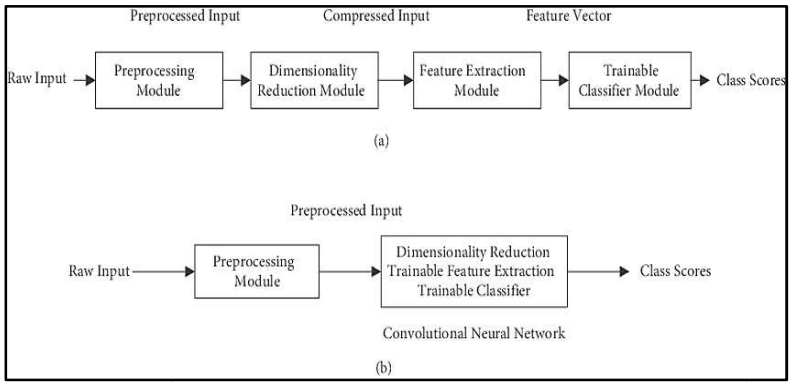
Small VGG Network

The VGGNet is one of the top five models in image processing. It works well for dataset with more than 14 million images and 10000 classes. Here we have used only a small part of VGG network. It have a small 3x3 sized convolutional layers, one after the other are stacked prior to perform pooling operation. This layer increases the depth of the network and that helps in understanding complex features. Pooling ReLU

Copyright reserved © J. Mech. Cont.& Math. Sci.
Ramalakshmi K et al

(Rectified Linear Unit) is the activation function used in this network. In the first layer, the input shape of height, width and depth have to be given. The number of convolution layers in this network starts from 32 and increases to 64 and 128 after every sub-sampling layer. The final layer is fully connected with the output classes. The gender recognition is done using only a small part of VGGNet and hence it is called SmallVGG Network. There are four steps process in this network.

Input Image



The input image is resized to 64x64 pixel images. All the images used are RGB images which are focused on face.

Pre-Processing

The pre-processing is the part where the input images are resized and reshaped. This process translates the image into an array matrix of pixel values.

Fig. 2:(a) Conventional approach (b) CNN based approach in gender recognition.

Feature Extraction

The feature is extracted using convolutional layers in the network. There are two parts in this network, a convolutional layer anda pooling layeror subsampling layer. The input is given to the convolutional layer. As the filter slides around the image, it is compared with the original values of the image. This is repeated for all the locations. From that an activation map is created as output. The output of the first convolutional layer is the input for the next convolutional layer.

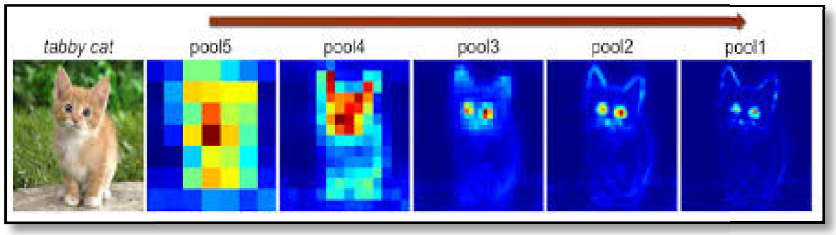


Fig.1:CNN feature extraction activation map.

After passing the image through convolutional layer, it is passed through sub sampling. There will be many subsamples created during this and only the best one will be taken. The features extracted in each convolutional layer are given in Fig 3.

Copyright reserved © J. Mech. Cont.& Math. Sci.
Ramalakshmi K et al

Prediction

The input image will go through the network then the image will be classified into certain class. The detected feature will be attached to the end of the network using fully connected layer. While comparing the pixel values of the input image with the original image; if the difference is high, then the chances of the image to correlate with a particular class will be high.

Convolutional Neural Network

Convolutional neural network uses the VGG net architecture for recognition and detection of faces. The first convolutional layers have 32 filters, the next have 64 filters, and third convolutional layer have 128 filters. All the filters are of size 3x3. In the first convolutional layer, the edges and curves are identified. In the next convolutional layer it will read high level features from the activation map of previous convolution layer. The output from the first convolution layer is the input for the next convolutional layer. The filter will be sliding from the top left corner of the image till the last location. The filter will move only one unit towards right at a time. Every unique location of input has a numerical value [XX]. After going through all the locations the activation map is created by finding the free spaces. Each filters acts like a feature identifier. There are two types of features, high level features and low level features. High level feature affects the classification, detection and recognition of an object: face or pose variation. Whereas the low level features detect the edges, curves, and lines in the image. The pixel value will be given in the location along with the area that has any edges or cures. The filter values and pixel values are multiplied and summation is done. If the obtained value is a large number and then the shape will resemble to the one in the filter. If the value is low, then chances for resembling will be low. More filters can be added to find more features, since as the filter increases, depth of the activation map will increase and that's leads to read more information about the input image. The output of the second convolutional layer is the activation map that represents the high level features.

After detecting the high level features, the fully connected layer will be attached to the end of the network. It takes input volume and gives an N-dimensional vector output, where 'N' is the number of classes. The full layer will look at the previous layer output and determines which feature correlates the most to the given classes. Training is done by adjusting the weights randomly. This process is called back propagation. At the beginning of training, the weights are initialized with a random value. Without weights, the filter won't be able to know what process have to be done. While training, the image has to be given as well as labels to process. There are four steps in back propagation forward pass, loss function, backward pass, and weight update.

Forward Pass

During this stage, the training images which are stored in the form of array of numbers are passed through the network. Since the weights are initialized randomly, the output can't be classified in this stage.

Loss Function

Output of the forward pass is given to the loss function. The loss function is defined through Mean Squared Error (MSE) value.

Copyright reserved © J. Mech. Cont.& Math. Sci.
Ramalakshmi K et al

MSE, E_{total} =Σ½ (target – output)^2 (1)

Loss will be high for the first few training images. If the predicted label and the training label are same may reduce the loss. For that the weights which contributes directly to loss have to be identified which can be expressed as dL/dW. Then the backward pass is done.

Backward Pass

Through this backward pass, the network will determine the weights that cause more loss. This also finds way to adjust or to reduce the loss.

Weight Updation

Here while changing the filter values; it will change in the opposite direction of the gradient. This can be represented as,

$$W = W_i - \eta\frac{dL}{dW} \quad (2)$$

Where W is the weight, W_i is the initial weight and Ƞ is the learning rate. The learning rate is selected by the programmer. Higher the learning rate, bigger the steps taken during weight updation. Thus it may take less time for the model to identify an optimal set of weights. When the learning rate is high, it might cause a jump in the training which results in less precision to reach the expected output. Each training process consists of these four iterative steps. This is then iterated multiple times for the set of training images. As it finishes its training, the last training sampleis expected to be trained enough for identify correctly. Then the output is compared with the ground truth to test the performance of the algorithm. More training image, more iteration, more weights updates, may fine tune the CNN.

CNN in Gender Recognition

To create a gender recognition algorithm, it needs a model and thendesign the network using, inputs, process and the expected outputs etc. After choosing the model, the network has to be trained with the training dataset. During the training process like pre-processing, feature extraction, classification models are done. The trained data is further tested using test dataset. As the number of images in the training set increases the accuracy will also increase. There are three modules in the proposed system: building network, training network and prediction.

Building Network

VGGNet is the network used in this gender recognition system. To build a VGGNet the model need to know the height, depth and width of the input image, and the number of classes. The depth of the image is the number of channels in the input image. Since RGB images are used in this research, the colour space of the image will be three. The sequential model is used to stack the data into layers. There are four layers in the VGG network created in this process.

The first layer convolutional network is passed to the activation function ofReLu (Rectified Linear Unit). The convolution layer has 32 filters of size 3x3in. In the first layer, the input shape is given to process the dimensions.The input image is normalized before passing it into the next layer of network; this is done using batch normalization function. It stabilizes the training by reducing the number of epochs. One epoch is when the entire dataset is passed forward and backward through the neural network only once. Since the dataset is too big, it is divided into smaller

Copyright reserved © J. Mech. Cont.& Math. Sci.
Ramalakshmi K et al

batches. The total number of training sample in a single batch is the batch size. One epoch is not enough for getting optimised learning, as the epoch increases the result will go from under-fitting to optimal fitting. Pool layer will reduce the spatial size of the input. To make the network more robust dropout function is used, since it removes the random neurons between layers. It also removesover fitting and increase the accuracy. During each iteration, the dropout is done. Iteration is the number of batches needed to complete one epoch. The same process is repeated with more number of filters. In the second convolutional layer, there were 64 filters and in third layer there were 128 filters. As the network layer increases, the volume of the input is reduced. The final layer will be then fully connected to the outputs using dense function in keras. To get the class probability for each label, softmax is used.

Training Network

The network training process requires arguments regarding the dataset, name of model, name of graph for plotting accuracy, and loss etc. These are read using argument parser and processed using ImageDataGenerator. ImageDataGenerator. It does data augmentation by adding rotations, shifts, and scaling. This reduces duplication of files since it is done during execution. The dataset, model, label binarizer, and plot should be mentioned with their path to be stored. The label binarizer and plot is required to generate the report regarding accuracy and loss in training and testing. The data along with their labels are sorted and shuffled. Through various iterationprocess each images are loaded, resized, and stored in the dataset.

During the looping process, the class labelsare extracted and add in the label list. The input pixels are converted into arrays using Numpy. This data is then divided into training and testing. In this proposed system, tereare 90% of the inputsused for training and remaining 10% is used for testing. The stored labels are then converted from integer to vectors. This is for encoding and storing the label in a pickle file so that it can be used later. Then the data is augmented to avoid over fitting and to generalize the model. The network is built by calling smallVGGNet with required parameters like width, height, depth, and number of classes.

To compile and train the model, the network needs to initialize the learning rate, number of epochs, and batch size. To initialize the model and SGD optimizer, the categorical_crossentropy is used. The network is trained using fit_generator with training data as their first parameter. The generator will then produce the batches of augmented training data according to the parameter settings made. Finally the proposed model is evaluated plotting loss and accuracy curve using matplotlib functions and numpy. For plotting the statistical charts, theclassification_report is used with the labels and range. The serialized network and the pickled file are stored into the system using file open and write functions. The accuracy of testing is displayed after all the epochs. As the number of epoch increases, the accuracy of the network also increases. This can sometimes cause over fitting when the number of epoch is very high. Hence an optimum value should be identified for epoch. Here the epoch value used is 50, the batch size is 15, and the learning rate is 0.01.

Prediction

To give the input to the system, cascade for detecting face have to be initialized. It is done using haarcascades classifiers in opencv. Haar cascade have the features to detect face depending on the type of classifier used. There are different

Copyright reserved © J. Mech. Cont.& Math. Sci.
Ramalakshmi K et al

types of classifiers available in haarcascade such as,haarcascade_frontalface, haarcascade_eye, haarcascade_lefteye_2splits, etc. Here to detect only face haarcascade_frontal face_default is used. After initializing the cascade, the image is loaded using argument parser. Using detect MultiScale () method in the classifier, the face is detected from the input image. It has the input image, scale factor and minNeighbors as parameters. The scale factor reduces the image size using min Neighbors. It determines the number of neighbours that each rectangle should have to retain it. Then the number of detected faces in the images is identified. If the face count is one, then it will crop the detected face and resize it into a 64x64 image. The model files and pickle file created while training is loaded. The face prediction is done using model.predict(). This is then displayed as the input image with the prediction probability. If the face count is more than one, then it will crop all the detected faces and it go through the same process for prediction and it show the output for every single face detected in the image

IV. Results and Discussions

A graph is plotted according to the accuracy and loss during each epoch. From that graph we can understand that how well the network is trained during each step. This also shows the person in the input image is a male, female or neither.

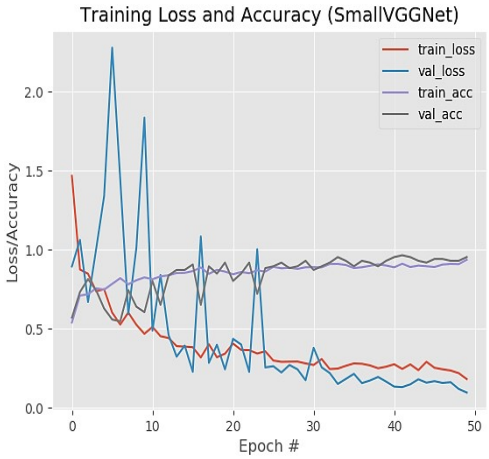


Fig. 2: The graph indicating the training accuracy and loss of the network after training the network.

The database used for training is called gender color FERET, which consist of 418 images for both male and female. The images in the database are of faces and not the full image of a person. The training images are cropped after detecting their face. When it is given for training, it will take 42 (10%) images for testing and remaining 376 (90%) images are used for training. As shown in fig. 4, the number of training images increases, the accuracy of the algorithm will increase while testing.

V. Conclusion

The gender recognition using Tensorflow and deep learning is one of the efficient methods. The proposed system learns the complex features and processes it.

Copyright reserved © J. Mech. Cont.& Math. Sci.
Ramalakshmi K et al

It produces an output of accuracy of 97.41% during training and 95.40% accuracy while testing. The training is done with images focused on face, when it comes to other pictures face have to be detected, cropped and then process it further. AlsoVGGNet is considered to be one of the best networks for gender classification. But it requires good memory and time since it have huge computations. Due to the huge computational requirements, it is difficult to deploy VGG in GPU. It becomes inefficient as the width of convolutional network increases. This can be improved in the future to make it more accurate and efficient.

VI. Acknowledgements

The authors would like to thank the management of Karunya Institute of Technology and Sciences for all the supports to complete the work.

Conflict of Interest

The authors have no conflicts of interest.

References

I. Amit Dhomne, Ranjit Kumar and Vijay Bhan. Gender Recognition through Face using Deep Learning. Procedia Computer Science. 2018; 132: 2-10.

II. Biao Shi, HuaijuanZang, RongshengZheng and ShuZhan.An efficient 3D face recognition approach using Frenet feature of iso-geodesic curves. Journal of Visual Communication and Image Representation.2019; 59: 455-460.

III. Dhriti. K-Nearest Neighbor Classification Approach for Face and Fingerprint at Feature Level Fusion.International Journal of Computer Applications. 2012; 60(14): 0975 – 8887.

IV. DongshunCui, GuanghaoZhang, KaiHu, WeiHan and Guang-BinHuan, Face recognition using total loss function on face database with ID photos. Optics & Laser Technology. 2019; 110: 227-233.

V. Gil Levi and Tal Hassner. Age and Gender Classification Using Convolutional Neural Networks. IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW). 2015.

VI. Hui-Cheng Lian, Bao-Liang Lu, ErinaTakikawa, and Satoshi Hosoi. Gender Recognition Using a Min-Max Modular Support Vector Machine. Lecture Notes in Computer Science. 2006; 210-215.

VII. HuiZhi and Sanyang Liu.Face recognition based on genetic algorithm. Journal of Visual Communication and Image Representation.2019; 58: 495-502.

VIII. Kevin Santoso and Gede Putra Kusuma.Face Recognition using Modified OpenFace. Procedia Computer Science.2018; 135: 510-517.

Copyright reserved © J. Mech. Cont.& Math. Sci.
Ramalakshmi K et al

IX. MaafiriAyyad andChougdali Khalid.New fusion of SVD and Relevance Weighted LDA for face recognition, Procedia Computer Science.2019; 148: 380-388.

X. MaafiriAyyad andChougdali Khalid.New fusion of SVD and Relevance Weighted LDA for face recognition, Procedia Computer Science.2019; 148: 380-388.

XI. Rai P and Khanna P. Gender Classification Techniques: A Review. Advances in Computer Science, Engineering & Applications. Advances in Computer Science, Engineering & Applications. 2012; 51-59.

XII. Ranjeet Singh and Mohit Kumar Goel. Gender Classification Techniques-From Machine Learning to Deep Learning. International Journal of Computer Technology and Applications. 2016; 9(41): 77-88.

XIII. RupaliSandipKute,VibhaVyas and AlwinAnuse. Component-based face recognition under transfer learning for forensic applications.Information Sciences.2019; 476: 176-191.

XIV. Samik Banerjee and Sukhendu Das.LR-GAN for degraded Face Recognition.Pattern Recognition Letters.2018; 116: 246-253.

XV. SwaroopGuntupalliJandM. Ida Gobbini. Reading Faces: From Features to Recognition.Spotlight. 2017; 21(12): 915-916.

XVI. Vito Santarcangelo, Giovanni Maria Farinella and SebastianoBattiat. Gender Recognition: Methods, Datasets and Results. Conference: International Workshop on Video Analytics for Audience Measurement. 2015.

XVII. YangLi, WenmingZheng and ZhenCuiTong Zhang.Face recognition based on recurrent regression neural network. Neurocomputing. 2018; 297: 50-58.

XVIII. Yan Liang, Yun Zhang andXian-Xian Zeng.Pose-invariant 3D face recognition using half face. Signal Processing: Image Communication. 2017; 57: 84-90.

XIX. Yihua Chen, Ali Rahimi and Luca Cazzanti. Similarity-based Classification: Concepts and Algorithms, Journal of Machine Learning Research. 2009; 747-776.

XX. ZahraddeenSufyanu, FatmaSusilawatiMohamad, Abdulganiyu Abdu Yusuf, AbdulbasitNuhu Musa and RabiuAbdulkadir. Feature Extraction Methods for Face Recognition.International Review of Applied Engineering Research. 2016; 5(3): 5658-5668.

Copyright reserved © J. Mech. Cont.& Math. Sci.
Ramalakshmi K et al