International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 08 Issue: 02 | Feb 2021 www.irjet.net p-ISSN: 2395-0072

A Survey on Different Approaches to extract Facial Attributes

Raj Baldania[1], Prof. Barkha Bhavsar[2]

[1]*M.Tech Student, Dept. of Computer Engineering, LDRP Institute of Technology and Research, Gandhinagar, Gujarat, India 382015*
[2]*Professor, Dept. of Computer Engineering, LDRP Institute of Technology and Research, Gandhinagar, Gujarat, India 382015*

Abstract - Face detection and analysis systems has been growing in last few years for various applications. Since the hardware performance increase in last few years, use of Deep Learning, Convolution Neural Network, Face detection, Face analysis techniques is increasing and day by day developed models are breaking accuracies of previous models and research in various tasks. Facial analysis system with age, gender and emotion recognition have been proposed with good accuracies for real-time and non-real time both. The present dissertation focuses to provide a robust system architecture for age, gender and emotion recognition in real time which can be use in commercial, healthcare, and many more industries. To achieve this a literature survey is done on the same topic with previous researches to compare their results. The final model architecture proposed in this dissertation is efficient and fast and provides accurate results as compare to previous researches

KeyWords: Facial Expression Recognition, Feature Extraction, Support Vector Machine, Deep Learning, Convolution Neural Network, Age Recognition, Gender Recognition, Emotions Recognition.

1. INTRODUCTION

In previous researches, studies have used various techniques to know about human feeling by measuring stress, voice and retinas. Same way for more accurate data of what a human being is feeling can be measure by recognizing the emotions by detecting their face with the help of computer vision and machine learning. It is difficult to detect voice of particular human in a crowded environment and also to scan their body for stress which will take more time than face detection. For this only solution to measure the feelings and thinking of human is to recognize face and extract features.

Face detection and recognition both have only one difference that is in face detection the computer detects the faces but in recognition the computer detects the face and tries to recognize that person by scanning the face in existing database. In real world, there are so many number of applications and scope regarding this research. One can use this to measure that how many people likes their advertisement in a street screen having a camera above it. We can measure then number of happy and sad audience in a live show in real time and many more applications.
Analysis of facial attributes in by an artificial intelligence system can change the way of living for people. It can easily predict the mood of a person based facial attributes like emotions, age, gender and other factors. Same as a person can easily say the mood of another person by just watching the face. Using these attributes, we can decide how to react with that person, like a restaurant waiter can recognize that customers liked the meal or not, which music to play in café to make comfortable to the customer. Same way in a clothing store by predicting customer ‘s age, gender and emotion the system can suggest the perfect outfit. In short machines will be able of taking care of things what the customer will like or not without asking customers. Not only in commercial industry, it can also judge what a patient feels in hospital and clinics because sometimes the patient afraid to tell the doctor how they feel Machines will be able of taking care of the uninteresting part of life, end-to-end. It’s possible to make this happen by combining computer vision and machine learning for facial analysis. Deep learning and Convolution Neural Networks are widely used to extract features from images and identify as an output. Being a Computer Engineer and capability of programming it motivates me to create a Multi Model approach for human face feature extraction.

This paper will give a literature survey of different methods and approaches to extract facial features like age, gender and emotions. Different types of Convolution Neural Networks (CNN), Deep Learning approaches and other algorithms are used.

Datasets such as IMDB-WIKI - 500k+ [1] face images with age and gender labels, FER 2013 [2] , MORPH [3], FG-NET [4], CVPR2016 LAP [5] challenge dataset and AffectNet [6] emotion dataset are used by the methods which are introduced in this paper.

The remainder of this paper is organized as follows: Section 2 present a brief explanation of facial analysis, age estimation, gender recognition and emotion recognition. The different methods or techniques of facial expression recognition are given in Section 3. Result as comparison between methods introduced in this paper is presented in Section 4, and the conclusions are given in Section 5.

2. FACIAL ANALYSIS

The term facial analysis itself gives definition of analyzing person's attribute using its facial features. In recent years facial analysis is growing faster in research and commercial sector both, as increasing of computing power the evolution of Deep Learning system increases and neural networks are outperforming humans in estimation and recognition tasks.

In the next three sub sections, an overview of tasks for age, gender and emotion recognition through computer vision and machine learning is introduced

2.1 Age Estimation

In recent years age estimation using computer vision and machine learning becomes attractive and challenging topic by developing an automatic system to identify age. There are many approaches to detect age of a person, but using machine learning by feeding age labelled images to it gives more accurate output. Predicting the age from a face image has been a complex challenge in computer vision. There are many applications to this task, such as precision advertising, intelligent surveillance, face retrieval and recognition, etc. Tradition age estimation method usually consists of two parts, first one includes an input of face image which starts by feature extraction to represent aging information, the second part includes performing the age estimation task. The feature extraction task has been performed with several methods, such as using based on LBP and GLCM Features Using SVM [7].

2.2 Gender Recognition

Analyzing face using computer vision and machine learning has increased the attention to identify attributes like age and gender. Many researchers also provided algorithms and CNN architectures for gender prediction by different approaches. This gender attribute has so many applications such as human-computer interaction, surveillance, intelligent marketing and for commercial use also. Previous methods for gender recognition used several different approaches, such as AdaBoost [8], Neural Networks, SVM [9] and many more.

2.3 Emotion Recognition

Automatic emotion recognition system is growing day by day as there are so many researches done and still going for this. In Human-computer interaction area, recognition of emotions is having more interest which focused on the interface between computers and humans. There is wide application of automatic facial emotion recognition such as audience analytics, marketing, humanoid robots, entertainments, etc. The way researches are done on this system, complexity and variability also increases which makes it difficult task to make it more accurate than previous researches. There are two main steps for the task of automatic facial emotion analysis: feature extraction and classification of facial expression. There are many models proposed such as Automatic Facial Expression Recognition Using Facial Animation Parameters and Multistream HMMs [10], Real Time Facial Expression Recognition in Video using Support Vector Machines [11], Neural-AdaBoost based facial expression recognition system [12], etc. Methods including this one are compared in A Survey on Human Facial Expression Recognition Techniques [13] briefly. Many models with different classification methods have been proposed for the task of facial emotion recognition. Most of the models and databases for emotion recognition use seven main human emotions: anger, disgust, fear, happiness, sadness, surprise, neutral.

3. DIFFERENT TECHNIQUES OR METHODS TO EXTRACT FACIAL ATTRIBUTES

3.1 DAGER: Deep Age, Gender and Emotion Recognition Using Convolutional Neural Networks. [14]

Authors: Afshin Dehghan, Enrique G. Ortiz, Guang Shu, Syed Zain Masood

Technique/Method: Deep CNN, Detailed proposed method not mentioned

Description: - In this paper they proposed an end to end pipeline with novel deep networks to detect human face and extract features like emotions, age and gender in real-time.

In Figure 2.1 it shows the proposed method pipeline of the system. The first deep model is trained by feeding a large dataset of 4 million images for the task of face detection as it serves the backbone of the system. After training the images by their own Sighthound's Face Recognition model. After this step, all images are fine-tuned by Age, Gender and Emotion recognition model.

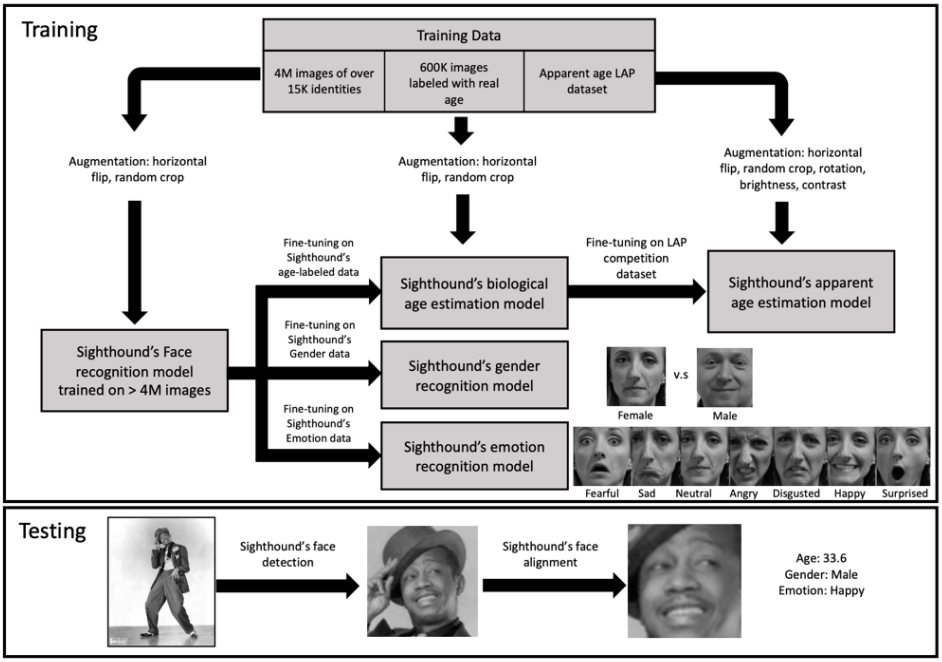

Figure 1 : DAGER proposed system architecture of method 1

The above system consists of 4 models. The first Sighthound's Face Recognition model is trained on 4 million images and this face recognition model will serve images as input to other 3 Sighthound's models for age, gender and emotion recognition. That' why Sighthound's Face Recognition model is the backbone of the system.

In Training phase, the data containing images with labels is pre-processed which includes face detection, facial landmark detection and alignment before feeding them to their DNN (Deep Neural Network). The face recognition model achieves outstanding results on the LFW dataset.

Results: -

(i) Real Age Estimation

Table 1: Real Age Estimation Mean Absolute Error results of method 1

Methods	MAE
Sighthound	5.76
Rothe et al. [4]	7.34
Microsoft. [13]	7.62
Kairos [14]	10.57
Face++ [15]	11.04

The table shows Mean Absolute Error (MAE) comparing with other methods also.

(ii) Gender Recognition

Table 2: Gender Recognition Accuracy Rate results of method 1

Methods	Accuracy
Sighthound	91.00%
Microsoft. [13]	90.86%
Rothe et al. [4]	88.75%
Levi and Hassner [18]	86.80%
Kairos [14]	84.66%
Face++ [15]	83.04%

The above table shows the accuracy of model on Adience benchmark by comparing to other stare-of-the-art researches

3.2 Audience Analysis System on the Basis of Face Detection, Tracking and Classification Techniques [15]

Authors: Vladimir Khryashchev, Member, IAENG, Alexander Ganin, Maxim Golubev, and Lev Shmaglit

Technique/Method: P. Viola and M. Jones for face detection, for face tracking an algorithm proposed by B. Lucas and T. Kanade was chosen as the basic approach to solve the problem of optical flow calculation, AF-SVM algorithm for Gender Recognition, Hierarchical approach using Binary Classifier for Age Estimation

Description: - In this to solve the task of face detection AdaBoost classifier, described in paper is utilized. This procedure consists of three parts: 1) Integral image representation 2) Learning classification functions using AdaBoost. 3) Combining classifiers in a cascade structure.

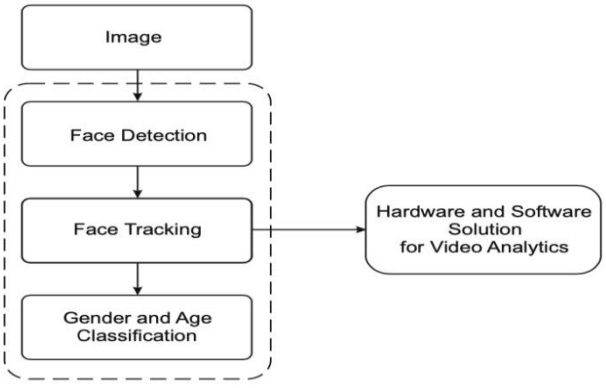

Figure 2: A block diagram of the proposed audience gender and age classification system of method 2

After face detection, the next step is for face tracking using an algorithm, proposed by B. Lucas and T. Kanade in paper, was chosen as the basic approach to solve the problem of optical flow calculation. With the help of this algorithm the coordinates of feature pixels on the current image frame are calculated out of their coordinates on the previous frame.

Gender recognition algorithm, proposed in this paper, is based on non-linear SVM classifier with RBF kernel. To extract information from image fragment and to move to a lower dimension feature space we propose an adaptive feature generation algorithm which is trained by means of optimization procedure according to LDA principle. Thus, the proposed classifier is based on Adaptive Features and SVM (AF-SVM).

AF-SVM algorithm consists of the following steps: color space transform, image scaling, adaptive feature set calculation and SVM classification with preliminary kernel transformation. Data, required for the calculation of adaptive feature set, is generated during training. The training procedure of the proposed AF-SVM classifier can be split into two independent parts:

1. Feature generation
2. SVM construction and optimization.

The proposed age estimation algorithm realizes hierarchical approach (fig. 4). First of all input fragments are divided into three age groups: less than 18 years old, from 18 to 45 years old and more than 45 years old. After that the results of classification on the first stage are further divided into seven new groups each of which is limited to one decade. Thus, the problem of multiclass classification is reduced to a set of binary "one-against-all" classifiers (BC). These classifiers calculate ranks for each of the analyzed classes. The total decision is then obtained by the analysis of the previously received histogram of ranks.

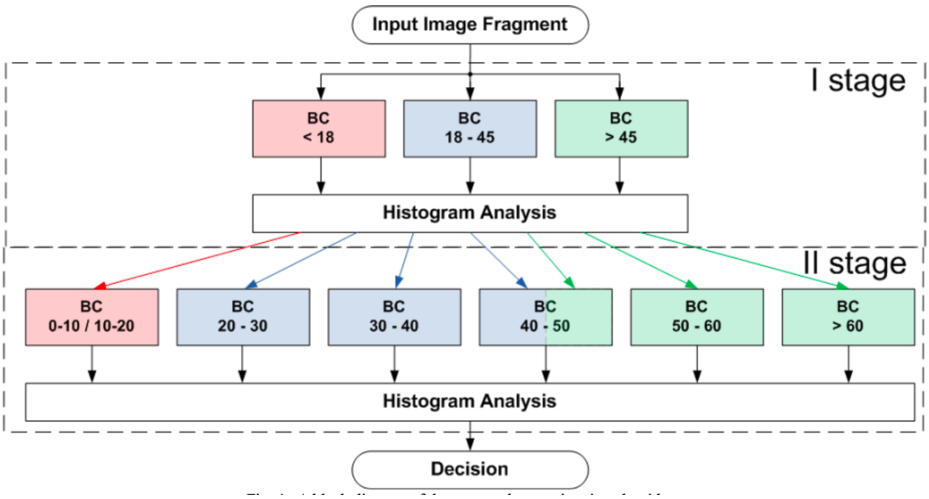

Figure 3: A block diagram for the proposed age estimation algorithm of method 2

Results: -In conclusion of this research the authors didn't mentioned accuracy or any other format of result of this proposed system.

3.3 Age, gender and emotion detection using CNN [16]

Authors: Manasa SB, Jeffy.S. Abraham, Anjali Sharma, Himapoornashree KS

Technique/Method: Haar Cascade classifier for face detection, CNN for emotion detection, ResNet for Age and Gender Classification

Description: -

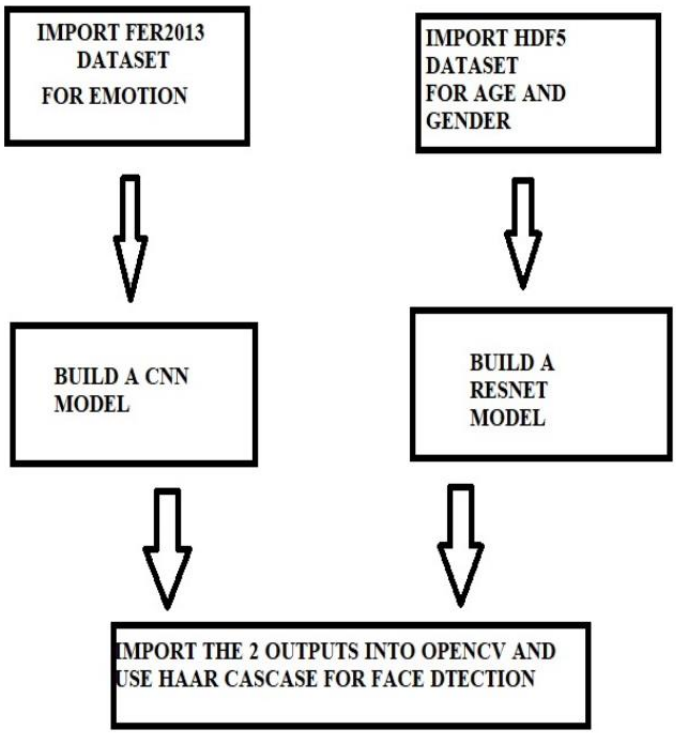

Figure 4: Blueprint of proposed system of method 3

A. Face Detection:

Haar cascade XML file which is a classifier used to identify a specific object from the webcam. The haarcascasde_frontalface_default.xml provided by OpenCV used to recognize frontal face.

B. Emotion Detection:

FER2013 is a Kaggle dataset that contains labelled 3589 test images, 28709 train images.

They used 3 convolutional layers. Input [48x48x1] carries the pixel values of given image. Hence images have width equal to 48, height equal to 48, and with one color channel.

Step 1: Normalizing the data between 0 and 1.

Step 2: They used 3 convolutional layers. In each layer, we do Batch Normalization, MaxPooling. In fully connected layer they used RELU activation function and SOFTMAX function.

Step 3: Calculate the loss function using Adam optimizer

Step 4: RELU activation function and use to use the trained model later, save the weights in fer.h5.

C. **Age and Gender Classification**:

Here they used an hdf5 file as dataset which has 10,000 images. The images have been labelled and classified into male, female and various age groups. Here they used ResNet neural network architecture to train this model. ResNet is used to add a

large number of layers with strong performance.

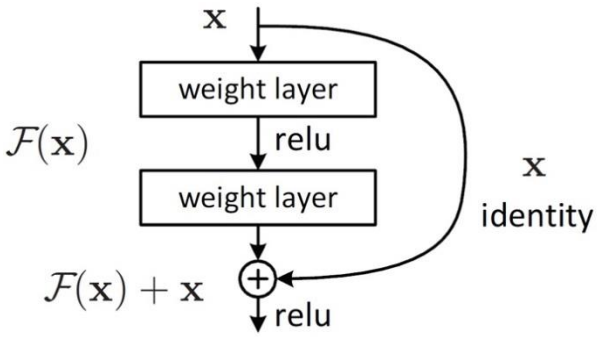

Figure 5: Single residual block

A network containing residual blocks, each layer is an input to the next layer and also inputs into the layers about 2-3 hops away. If the number of layers keeps increasing, we will notice that the accuracy will tend to saturate at certain point and after a point it will gradually decrease. This is not just because of over fitting. It indicates that shallow networks are better at learning than deeper networks. So, the skip- connections skips the training of few layers and are also called residual connections as shown in above image. By including skip connections in our network architecture, they are making the network to skip training for the layers which are not useful and don't add much value in overall accuracy.

Results: -

They have used resnet architecture instead of VGG16 for age or inception v3 for gender classification. Resnet helps in handling the training error generated as the networks get deeper. We were able to achieve 95% accuracy rate. Replacing VGG-16 layers in Faster R- CNN with ResNet, we can observe a relative improvement of 28%.

3.4 A Convolutional Neural Network for Real-time Face Detection and Emotion & Gender Classification [17]

Authors: Md. Jashim Uddin, Dr. Paresh Chandra Barman, Khandaker Takdir Ahmed, S.M. Abdur Rahim, Abu Rumman Refat, Md Abdullah-Al-Imran

Technique/Method: Sequentially fully CNN, Back propagation

Description: -

At first, they take a real-time video frame then convert it as an image and then extract face from image to detect a human face. After extracting face, we consider each face part of the image as a full image for further process. Each extracting face image is then providing as input to pre-process step of classification model and each pre-processing step takes some operation on its input to resize as model input and data augmentation as input to their proposed convolutional neural network (CNN) model for classification of the emotion and gender. The resulting label that is the output of the CNN is then used for making a description of gender {"man" or "women"} and facial emotion classification {"angry", "disgust", "fear", "happy", "sad", "neutral"}.

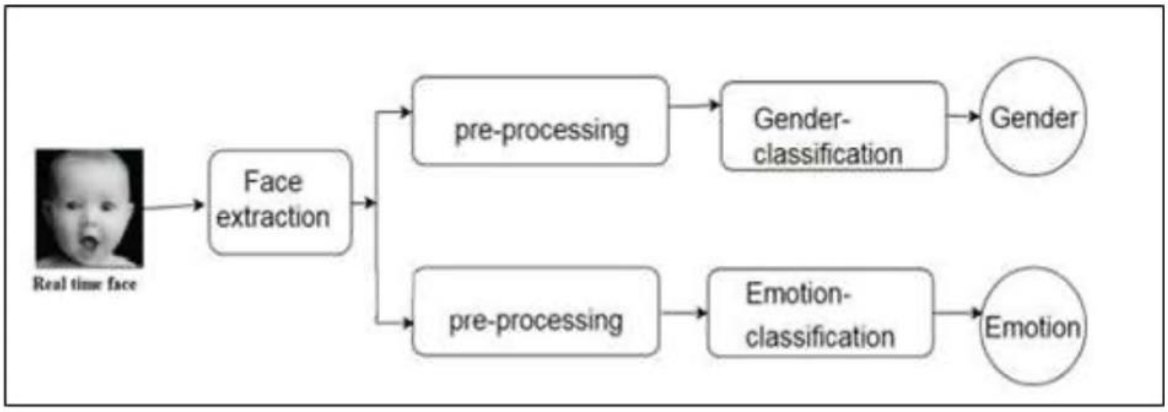

Figure 6: Blueprint of proposed system of method 4

Their method is a totally convolutional neural network that contains 4 residual depth-wise separable convolutions where each convolution is followed by a batch normalization operation and a ReLU activation function. The last layer applies a global average pooling and a soft-max activation function to create a prediction. This architecture has approximately 60; 000 parameters; which corresponds to a reduction of 10× when compared to our initial naïve implementation, and 80× when compared to the original CNN.

Results: -

The proposed architecture obtains an accuracy of 95% in the gender classification task.

They tested this method in the FER-2013 dataset and we gained the same accuracy of 66% for the emotion classification task.

The complete pipeline including the OpenCV face detection module, the gender classification, and the emotion classification takes 0:22 ± 0:0003 ms on an i55200U CPU.

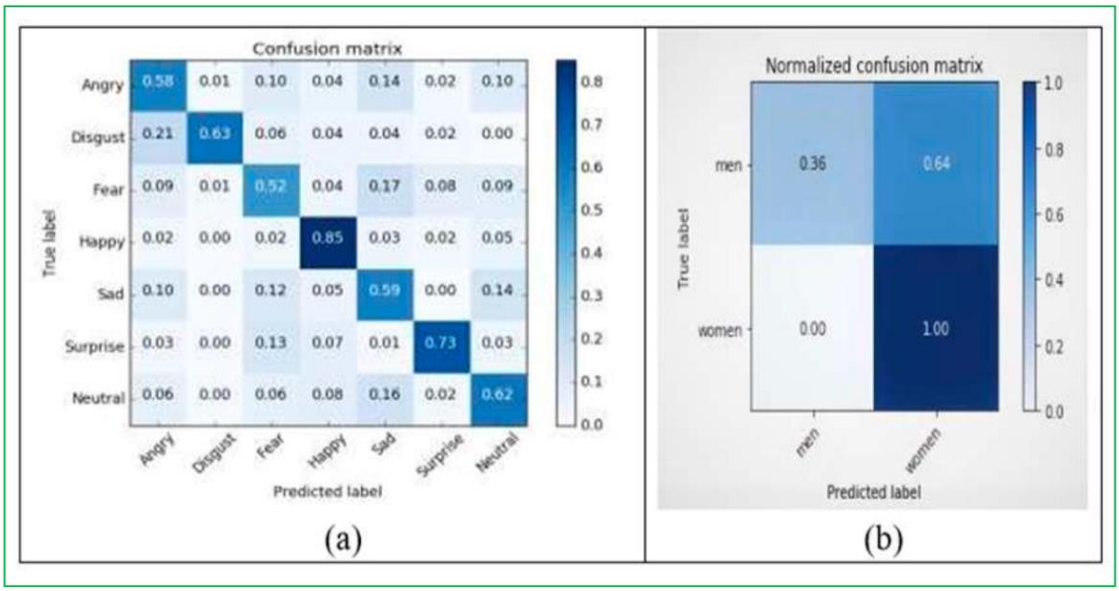

Figure 7: Normalized confusion matrix of our mini-Xception network (a) confusion matrix for facial emotion recognition (b) confusion matrix for gender recognition.

3.5 Real Time System for Facial Analysis. [18]

Authors: Janne Tommola, Pedram Ghazi, Bishwo Adhikari, Heikki Huttunen

Technique/Method: SSD for face detection and MobileNet for age, gender and emotion recognition

Description: -

A. Face detection

The face detection uses the SSD detector with MobileNet backbone, depth parameter α = 0.75 and input size 240x180. The depth parameter was chosen for fast performance and the input size matches the camera's aspect ratio 4:3. The network is initialized using COCO-pre-trained weights and trained with FDDB face database.

B. Age recognition

The age estimation uses MobileNet with depth parameter α = 1.0. The network is initialized using Imagenet-pretrained weights and fine-tuned in two stages: first with the large but noisy 500K IMDB-WIKI dataset and then using the small but accurate CVPR2016 LAP challenge dataset.

C. Gender recognition

MobileNet with α = 1.0 is used again. The last three layers were removed and replaced by a layer containing a single neuron and the network is trained from scratch in two stages: first with the 500K IMDB-WIKI dataset and then fine-tuned with the CVPR2016 LAP challenge dataset.

D. Facial expression recognition

MobileNet with α = 1.0 is used here, as well. The network is initialized with ImageNet pre-trained weights and fine-tuned with AffectNet database containing 7 different expressions.

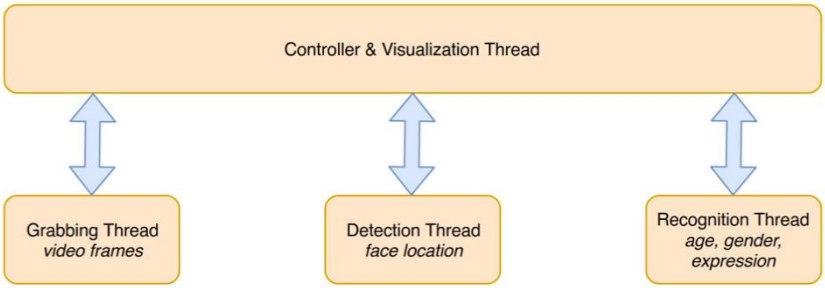

Figure 8: System architecture of method 5

Results: -

The accuracies are comparable with the state-of-the-art on most cases. However, our computational budget is limited, and we cannot use a combination of many networks, as *e.g.,* the best methods in age estimation do.

Table 3: Accuracies of the different stages of method 5.

Stage	*Network*	*Accuracy*
Detection	SSD-MobileNet, α=.75	67.2% (AP @0.5IoU)
Age	MobileNet, α=1.0	4.9 years (MAE)
Gender	MobileNet, α=1.0	88.3% (accuracy)
Expression	MobileNet, α=1.0	55.9% (accuracy)

4. RESULTS

Human Facial Expression Recognition by using different methods is studied in this paper. The following Table I gives a short idea of comparison between these methods. The table also gives a brief idea of the databases used by each method. The conclusion of each technique is also provided in the table. Also, future scope if there are included in Table I.

Table 4: Comparison table for all proposed methods given by above mentioned methods.

Sr.	Methods	Accuracy Rate	Advantage	Disadvantage
1.	DAGER: Deep Age, Gender and Emotion Recognition Using Convolutional Neural Networks	Top -1 accuracy for age prediction is 61.3 %, accuracy for emotion recognition is 76.1 % and for gender it is 91%	End-to-end complete pipeline for real-time	Accuracy for age prediction is low as compare to Paper 3
2.	Audience Analysis System on the Basis of Face Detection, Tracking and Classification Techniques	Not mentioned by author	Age classification is done by hierarchical approach which speeds up the classification process	Not combined pipeline. All models are presented in paper individually
3.	Age, gender and emotion detection using CNN	95 % accuracy for age and gender prediction. Accuracy rate for emotion recognition is not mentioned	Accuracy rate is high for age and gender	Not a real-time system
4.	A Convolutional Neural Network for Real-time Face Detection and Emotion & Gender Classification	Accuracy rate for gender prediction is 95% and for emotion recognition it Is 66%	Gender prediction accuracy is very good as compare to others	There is no age prediction model mentioned in the paper
5.	Real Time System for Facial Analysis	Accuracy rate for gender is 88.3 % and for emotions it is 55.9%	The whole system is real time	Accuracy of emotion recognition is very slow as compare to other methods

5. CONCLUSION

The objective of this paper is to give a survey and compare different methods for human facial feature extraction. In order to do this, the paper has looked into the details of every method introduced in section 3 and had given a short description with results. After a complete literature survey and comparative analysis of different approaches, it is concluded that there are some disadvantages such as accuracies of some models are low compare to others, face detection is done by Haar cascade or with basic CNN which takes more time than modern CNN models and libraries and also, we can use a single CNN for different tasks.

REFERENCES

[1] R. T. a. L. V. G. Rasmus Rothe, "DEX: Deep EXpectation of apparent age from a single image," in IEEE International Conference on Computer Vision Workshops (ICCVW), 2015.

[2] "FER 2013 Dataset," [Online]. Available: https://www.kaggle.com/c/challenges-in-representation-learning-facial-expression-recognition-challenge/data.

[3] K. Ricanek and T. Tesafaye, "MORPH: a longitudinal image database of normal adult age-progression," in 7th International Conference on Automatic Face and Gesture Recognition (FGR06), Southampton, 2006.

[4] Y. F. a. T. M. H. a. T. X. a. Y. Y. a. S. Gong, "Interestingness Prediction by Robust Learning to Rank," 2014.

[5] X. B. H. J. E. a. I. G. S. Escalera, "ChaLearn looking at people: A review of events and resources," in 2017 International Joint Conference on Neural Networks (IJCNN, Anchorage, 2017.

[6] A. a. H. B. a. M. M. H. Mollahosseini, "AffectNet: A Database for Facial Expression, Valence, and Arousal Computing in the Wild," IEEE Transactions on Affective Computing, vol. 10, no. 2371-9850, p. 18–31, January 2019.

[7] C. S. V. Rampay, "Automatic Age Estimation Based on LBP and GLCM Features Using SVM," International Journal of Advanced Research in Computer Science and Software Engineering, vol. 7, no. 11, pp. 186-190, 2017.

[8] H. L. H. Lu, "Gender Recognition using Adaboosted Feature," Haikou, 2007.

[9] L. Fei, "Gender Identification Using SVM Based on Human Face Images," in International Conference on Virtual Reality and Visualization 2014, Shenyang, 2015.

[10] P. Aleksic, "Automatic facial expression recognition using facial animation parameters and multistream HMMs," IEEE Transactions on Information Forensics and Security, vol. 1, no. 1, pp. 3-11, 2006.

[11] P. Michel, "Real time facial expression recognition in video using support vector machines," 5th international conference on Multimodal interfaces, 2003.

[12] E. Owusu, "A neural-AdaBoost based facial expression recognition system," Expert Systems with Applications, vol. 41, no. 7, pp. 3383-3390, June 2014.

[13] K. P. a. B. B. Raj Baldania, "A Survey on Human Facial Expression Recognition Techniques," International Research Journal of Engineering and Technology, vol. 7, no. 2, pp. 2644-2654, February 2020.

[14] A. Dehghan, "DAGER: Deep Age, Gender and Emotion Recognition Using Convolutional Neural Networks," 2017.

[15] V. Khryashchev, "Audience Analysis System on the Basis of Face Detection, Tracking and Classification Techniques," in roceedings of the International MultiConference of Engineers and Computer Scientists, Hong Kong, 2013.

[16] M. SB, "Age, gender and emotion detection using CNN," International Journal of Advanced Research in Computer Science, vol. 11, no. 1, pp. 68-70, May 2020.

[17] D. P. C. B. K. T. A. S. A. R. A. R. R. M. A.-A.-I. Jashim Uddin, "A Convolutional Neural Network for Real-time Face Detection and Emotion & Gender Classification," SR Journal of Electronics and Communication Engineering, vol. 15, no. 3, pp. 37-46, May 2020.

[18] P. G. B. A. H. H. Janne Tommola, "Real Time System for Facial Analysis,," Computer Vision and Pattern Recognition.