A Review on Human Face Recognition Techniques

Shreyas Sridhar¹

¹Dept. of CSE, National Institute of Engineering, Mysuru.

Abstract - The human face is a unique feature of the human body. This uniqueness varies from person to person as each human face is different from the other in various aspects and hence is an unique biometric of an individual person. Face detection, a branch of artificial intelligence is a technique of detecting faces from pictures, video footages, etc. Detecting human face plays an important role in the variety of applications such as facial animation, face recognition, human image database management and human computer interface, face verification and validation and a pre-processing step for many high-end and complex computer vision tasks for which a number of state-of-the art detection methods have been formulated. But the complexity of the human face and the changes due to different effects such as illumination, pose, angle of the picture taken will also impact the recognition accuracy of the algorithms and that is where the advancement in deep learning came in rescue. Recently with the advancement is deep learning we are able to train the artificial neural network for face detection and recognition, which in certain case have a better accuracy and performance than the convensional algorithms. The main objective of this paper is to present a survey of face detection and face recognition methods and algorithms and the neural networks. In this paper, I have presented an overview of the various methods used for face recognition algorithms in our study such as Viola-Jones, Principle component analysis (PCA), Local Binary Pattern (LBP) and neural network like CNNs (Convolutional Neural Networks). The paper concludes with future research.

Key Words: face detection, feature extraction, image processing, neural networks, face detection, face recognition.

1. INTRODUCTION

The humans are able to establish their identity based on their face. Among various biometrics such as fingerprint, iris detection, etc used for a person’s recognition, the face is one of the most popular. Face not only tells the identity of the people also conveys the information about human behaviour such as emotions. Face detection of people is one of the most important topics in computer vision and pattern recognition and it plays an important role in the numerous fields such as to authenticate the users, security of homeland, smart home access security, identify the criminal, identifying the user in small scale applications. Face recognition includes four parts: face extraction, face image pre-processing, feature extraction, classification and recognition. Any face recognition method will generally be used for fundamental services like face recognition [2], face tracking [25], face verification [21], facial behavior analysis [23] which can be considered as a building block of the multistage computer vision system.

Fig. 1 Face recognition as a building block of a multistage computer vision system.

The human face can be considered as a dynamic object and has a high degree of variability in its appearance, which makes face detection a difficult problem in computer vision. Hence to overcome this difficulty a number of methods and algorithms for face detection are being designed and formulated. One such famous and widely algorithm would be the Viola-Jones algorithm [10]. The method is a learning based algorithm which uses Haar based features and a cascade of classifiers to identify the objects. Various algorithms which are available for face recognition such as Viola-Jones, Haar Feature Based Detection (HFFD) method [5], Principal Component Analysis (PCA) [12]. Ahmad Almadhor [3] has devised a face detection algorithm based on deep learning for mobile applications. M.Sushama [24] has brought about combing the SIFT and DRLBP feature extraction methods to a new algorithm which a better accuracy than the individual methods. But theses classical algorithms all have an accuracy which are similar to each other and some algorithms may have a high false positive rate.

To overcome this problems artificial neural networks are trained for facial recognition where certain networks [2] have a better efficiency than convensional algorithms mainly due to the enhancement of deep learning.

Shun Lei Myat Oo [20], has designed an neural network to detect children face which have a different facial feature than an adult human face. Rong Qi [7] built a cascaded convolutional network by combing the Mobilenet and Resnet
structure which has better accuracy while keeping the real
time performance.

The rest of the paper is divided into sections, section 2 gives
the background analysis of face recognition methods, section
3 talks about the methodology, section 4 tells about the how
was data collected for analysis, section 5 talks about the
neural network, section 6 gives details about the face
detection and recognition methods and it is concluded in
section 7.

2. Background Knowledge

The various available face detection methods can all be
generally be grouped into three classes, which are as follows,
geometric based, appearance based, template based, and
color based.

In case of the geometric based approach for face detection
the face parts, the detection algorithms starts of by
calculating the geometric relationship between the features
like the eyes, ears, nose etc. The most used geometric
relationship methods would be the Euclidian distance [28],
Manhattan distance [28]. The geometric based is highly
independent of the rotation and scaling. Viola-Jones
algorithm [10] works best with this type.

In case of the appearance based, statistical and machine
learning techniques are used by the system to learn the
template characteristics. Some of the most widely used
common techniques are neural network [2], principle
component analysis (PCA) [12] and support vector machine
(SVM) [14].

To have a much more simpler implementation we can go
with the template matching [6] where the correlation values
between the face features are calculated using standard face
patterns.

Colour based approach [27] uses colour models such as RGB
colour model, YCbCr colour model for face parts detection.
This approach uses skin color information for segmentation
of the face.

3. Methodology

Human face analysis includes extraction of human face
features such as eyes, nose, lips, eye color, beard, mustache,
hair, skin color etc. Extraction of human face features are
useful in many applications such as face recognition,
emotions or expression recognition, management of face
image database, human computer interaction, image
retrieval and many more. The general steps which will be
done during the face recognition process is briefly explained
below. The figure in 3 are the general steps to be carried out
by any face feature extraction.

Image acquisition : before the image can be given as a input
for either during the training for the neural network or for
the face recognition the image capture needs to pre-
processed into a standard format. In this stage a digital
image will be created.

Image pre-processing : This step is mainly carried out to
improve the quality of the image and includes steps like
color conversion [10], image resizing [20] where the models
accepts a image of size 224x224. This can be done by
removing the unwanted data called as the noise. Some noise
removal techniques such as Gaussian filter, median filter, low
pass filter, high pass filter are available to remove noise from
an image If in case the noise is not removed from the image,
the model during the training phase will even learn the noise
which may lead to over fitting of the model. Certain methods
require the input image to be converted into a grey scale
image [14] of a standard format before giving as input to
increase the image visualization. But certain other model can
handle the colored image itself, some colored image formats
include the RGB format [20], YCbCr colour format. In [9],
using a SVM classifier on the Lab Color space to train the
model to distinguish between skin and non-skin areas. In
figure 3 the general steps to be carried for face feature extraction is shown.

**Face Detection:** Before the face recognition, feature extraction can occur the face detection needs to be done. The system must detect the which part of the image has the face object. For this to occur the network will be trained how a human face would look like and what are the focal points (eyes, nose, mouth, ears). Some techniques such as Viola Jones, skin color detection, morphological operations are used for detecting face region.

**Image segmentation:** Image segmentation can play an important role in feature extraction of human face images. Image segmentation means to divide the image into particular sections or regions. In facial feature extraction we can apply image segmentation for getting different feature parts like eye segmentation, lip segmentation, eyebrows segmentation, nose segmentation, beard and mustache segmentation. Extraction of face features can be done by performing segmentation on image.

**Feature Extraction:** Once the face detection occurs the features of the face can extracted to bring about the process of face recognition. The feature extraction of an image focuses on identifying inherent characteristics or features of an object present in the image. These features can be used to describe the objects. In human face features, the most important one are the eyes, nose, ears, mouth and this can be carried out by a number of techniques like calculating the distance between two eyes, eyes and nose, etc in case of geometric based techniques, Scale Invariant Feature Transform (SIFT) and Dominant Rotated Local Binary Pattern (DRLBP) [24]. Color based techniques can be used for eye color detection, beard and mustache detection.

Once the face detection and the feature extraction steps are carried out the network and begin with the face recognition step where the face will be detected by using certain standard algorithms like Viola-Jones, PCA or an improved version of the PCA called the geometrical approximated PCA (gaPCA).

In OpenCV, the classifier class is used through the CV2 to create a face and eye classifier objects. The respective XML files and CascadeClassifier() are loaded. The cv2 is used to create camera image. And in order to capture the images the VideoCapture() commands are used. To match the various sizes and to find the location the CascadeClassifier.detectMultiScale() object is used.

**4. Data Acquisition**

Data plays a very important role in today’s world, it is very much an essential resource which is required for the better and accurate performance of any application in today digital world. Data collection occurs in many ways, primary device can collect their own data or a third party device can collect the data and send it to the primary device for the analysis part and good example is the cloud computing. If in case the data analysis all occurs at a central location then it poses a huge drawback on the network cause it will constrain the central located device and if the central located device fails then the entire system may fail. To avoid this in [8], the authors have devised a method to bring about edge computing. As said earlier a good data set is very important whether be it a supervised or unsupervised learning a good data set will led to a better performance of the system with good results. Different authors have used different dataset and the most common is the Labelled Faces in the Wild (LFW) [8] which contains more than 13,000 images which are collected from the web. FDD dataset [3] is one of the publicly available dataset for the task of face detection. This dataset consists of 5171 annotated faces in 2845 images taken from Faces in the Wild dataset. Certain authors used a combination of two or more datasets for a comparison validation of their technique [22] used the Yale Face Database which contains 165 greyscale images in GIF format of 15 individuals and the Cambridge Database of Faces which
contains ten different images of each of the 40 distinct subjects. Certain authors created their own dataset to fit to their methods to get a better performance [20] created their own Child face dataset and no publicly dataset was there. In [28], they have created a new generation dataset called the T-Dataset which is created by utilizing the correlation between the training images without using a conventional technique of image density.

5. Neural Network

Artificial Neural Network (ANN) is inspired by the human biological nervous system. ANN uses the processing of the brain as a basis to develop algorithms that can be used to model complex patterns and prediction problems. As they have the ability to learn and model non-linear and complex relationships, which is really important because in real-life, many of the relationships between inputs and outputs are nonlinear as well as complex. After learning from the initial inputs and relationship between the input, it can infer unseen relationships on unseen data as well, thus making the model generalize and predict on unseen data. The learning process of ANN is used with back propagation algorithm [28].

There are many types of ANN like the multi-layered feed forward neural networks consist of many layers as input layer, hidden layer and output layer. Nodes of the layers make independent computation of data and pass it to next layer up to the output layer. Computation of data on the layer nodes are the weighted sum of the input values and output of the nodes are pushed out by activation function such as threshold function, sigmoid function, hyperbolic function.

The above mentioned types of artificial neural network can belong to two classes of neural network supervised ANN and unsupervised ANN. In [17], supervised ANN is a machine learning task where the model will be initially trained with input variables and using an algorithm to learn the mapping function from the input to the output. In case of unsupervised learning process [11], we will have only the input data and no corresponding output data. The main goal is to design a model which will learn about the various pattern which will exist between the data and bring about the learning process.

6. Face Detection and Face Recognition

6.1 Face Detection and Face Recognition

Human face detection is a type technology that used to identify human face in any type of still or digital images and even in videos. To detect face first digital images may be converted into grey-scale images or a coloured image can also be used as input. Images may be any type like larger image, crowded image, individual style of pose, colour and background illumination, etc. Many algorithms were developed to detect human face in images and videos. Face detection algorithms used to detect the size and location of the face region. Depending upon the input images, poses, angles, features and appearance, face recognition different methods are formulated and here in the following sections I bring about analysis of some face recognition methods and a comparison of their accuracy and performance.

Mobile applications have a limited resources both in terms of hardware as well as the software. Hence keeping in mind these limitations in [2], a technique is developed based on the Mobilenet architecture, by Google, which is based on depth wise separable convolutions operations which are later followed by the batch normalization and Relu non-linearity generating the feature maps which are provided as input to the soft-max layer which will bring about labeling of the object as face or non-face. The proposed algorithms in [2], has a better performance than traditional boosting algorithms like Viola-Jones and ResNet (SSD) where additionally it was able to detect faces with different orientations of face and also detect multiple face in a single go.

Recently a new human feature has been selected to incorporate which can increase the detection accuracy, the human skin color. Many techniques have been developed like the Haar Feature Based Face Detection (HFFD) which also considers the skin colour feature. This method works well only with specific illumination intensity and works poorly when the images have a varying colour and illumination intensity. To overcome this problem in [3], the author devised a technique which uses a robust hybrid face recognition technique based on the HFFD. In this method the outcome of the HFFD will decide the skin colour filter to be applied to each 5-pixel interval rather than an individual pixel throughout the detected face area by Rectangular Haar-
like feature process. Then the colour matching is done for selected pixels. If the average light intensity is high the method “RGB with Specific Values” will be used as color filter. Otherwise the method “Simple RGB Ratio” with a minor modification will be used for filtering the outcome of standard HFFD. From the experimental results it is observed that proposed face detection technique is suitable for detecting face from almost all kind of skin color and illumination intensity. In [4], the authors devised a method which will bring about optimization by increasing the speed of computation of one of the most widely used classification algorithm the Sliding window method (SVM) by developing an Ant algorithm which brings about optimization of the heuristic information processing for the face detection phase. Dominant Rotate Local Binary Pattern (DRLBP) and Scale Invariant Feature extraction (SIFT) feature extraction procedure is an accurate and perfect solution for all identification problems hence it is widely used in service sectors like the banking services, ATM Services, Aadhaar linked programs, police investigations etc. In [24], a new method is formulated where a combination of both these methods is used for face detection by using the back propagation neural network which is a type of feed forward network. Here initially the face image will be pre-processed to remove noise and the facial features are extracted by the DRLBP by dividing the image into windows called as cell, each cell of size 16x16 pixel and then the LBP method is applied on each cell to get a binary pattern and the SIFT features extracted are combined and given as input along with the database of image to the back propagation neural network for the face recognition process and further authenticate the person. The accuracy for the proposed system increased to 75% when compared to 48% of the previous method which used DRLBP and SIFT methods as an individual methods rather than together. The principle component analysis (PCA) is a well-established technique which can be used for both feature extraction and face recognition, known as the eigenface approach. This technique, has a good recognition rate, has from the disadvantage of high computation cost due to the complexity and technical difficulties in algorithm parallelization. To overcome this problem in [22] a geometrical approximated PCA (gaPCA) algorithm for computing the eigenfaces is proposed and here three datasets like the Yale dataset, Cambridge and LFW dataset are used as dataset. For the Yale and Cambridge dataset the proposed algorithm showed an overall increase in accuracy by 4% and hence can be considered as a viable alternative to the classical statistical approach for computing the principal components.

Facial recognition algorithms like HOG, SIFT, PCA, Viola-Jones have a very good recognition rate. But these techniques have certain limitations when the input images have a different illumination condition, pose, angle of face rotation etc. Hence to overcome these limitations recently researchers started to experiment with artificial neural network where the network models would be initially trained with a training dataset and later used for face recognition. In [8], the authors have built a convolutional neural network which is a deep unified model for Face Recognition based on Faster Region Convolution Neural Network. For the purpose of face recognition the author has used a Region Propose Network (RPN) is a light weight network that scans the images with the help of the sliding windows over the anchors it returns the anchors (bounding box) that have the maximum probability of containing the face. Edge Computing have been utilized for processing the data at the edges of the nodes of the network to reduce the data latency and increase the real time response.

In [20], the accuracy and performance of three standard Convolutional Neural Networks (CNNs) such as VGG Face based on two architectures (VGG16 and ResNet50) and Mobilenet on child face dataset is tested. The CNN is built and trained by using the ChildDB. The classification is done by using the fully connected layers of the CNN or by K-Nearest Neighbour Algorithm (KNN). All the three techniques gave a good performance average of accuracy of 99%, out of which the Mobilenet had the highest accuracy of 99.75% and can process the images in a very short period of time.

In [29], a CNN is built which can bring face as well as gender recognition is the same model. Three datasets (LFW, YTF and VGGFace2) have been used for the training and recognition purpose. In the neural network the author used AdaBoost algorithm to detect the face and Global Average Pool (GAP) is used instead of a fully connected layer which reduces the size of the network for the classification. They achieve the best results that are ever achieved on the Adience benchmark, by increasing accuracy from 91.80% to 93.22%. Viola-Jones is one of the most widely used algorithm for face recognition with an accuracy rate of 78.4%. In [10], an improvement is done to increase the accuracy of the algorithm and also improvised the algorithm to detect the eyes even if people are wearing glasses by using an artificial neural network. Firstly the face recognition is done by using the Viola-Jones algorithm which uses Haar features and a cascade of classifiers to identify the objects and the best features are selected by using the AdaBoost algorithm. This process is carried out in every layer to remove incorrectly detected face. To improve accuracy every object detected will be checked for the presence of an eye, hence every detected image region is classified as a face if and only if eyes are present in it. For the detection of eyes even if the person is wearing sunglasses the neural network was trained Matlab by using the Cascade Object Detector tool. The proposed modified Viola-Jones algorithm has a better accuracy of 90% and even be able to detect eyes in the image even if people are wear glasses.
Table-1: Survey of face recognition methods

<table>
<thead>
<tr>
<th>Author Name</th>
<th>Feature Extraction</th>
<th>Dataset</th>
<th>Other feature</th>
<th>Results</th>
</tr>
</thead>
<tbody>
<tr>
<td>Monali et al.[10]</td>
<td>Haar feature and cascade of objects to classify objects.</td>
<td>LFW dataset</td>
<td>Viola-Jones for face recognition, Adaboost algorithm for feature classification</td>
<td>90.8%</td>
</tr>
<tr>
<td>Eyad et al.[12]</td>
<td>PCA algorithm</td>
<td>ORL dataset</td>
<td>Euclidian distance, Squared Euclidian Distance, and the Gty-Block Distance methods for feature extraction.</td>
<td>100% accuracy by using the cluster method with Squared Euclidian Distance method</td>
</tr>
<tr>
<td>Sujay et al.[14]</td>
<td>Extended LBP</td>
<td>FERET dataset Yale dataset</td>
<td>Viola-Jones for face recognition, SVM for classification.</td>
<td>Accuracy of FERET-86.33% Yale-93.33%</td>
</tr>
<tr>
<td>Jandaulyet et al.[15]</td>
<td>LBP with Neural network</td>
<td>Ondokuz Mayis University database.</td>
<td>Viola-Jones for face detection.</td>
<td>91.75 for 1300 epoch.</td>
</tr>
<tr>
<td>Juxiang et al.[17]</td>
<td>LBP, HOG</td>
<td>IIITD dataset</td>
<td>BPNN for face detection.</td>
<td>A simplified version of Alexnet with better accuracy.</td>
</tr>
<tr>
<td>Gaili et al.[19]</td>
<td>Histogram equalization</td>
<td>ORL database</td>
<td>Google deep learning framework, Tensorflow called LeNet-5 for face recognition.</td>
<td>100% recognition accuracy at 7 training sample number</td>
</tr>
<tr>
<td>Wenxian et al.[21]</td>
<td>GaPCA</td>
<td>Yale, LFW and Cambridge dataset</td>
<td>Euclidean distance used for face recognition</td>
<td>Yale-73.33%, Cambridge- 93.33%</td>
</tr>
<tr>
<td>An-Chao et al.[27]</td>
<td>CNN based on google inception V3</td>
<td>User Identity database</td>
<td>Google Inception V3 for face recognition</td>
<td>Recognition accuracy of 90.61%</td>
</tr>
</tbody>
</table>

6.2 Comparison of tools used

The table II bellows gives a survey of the commonly used software and tools like Matlab, OpenCV, OpenMAJ, and Microsoft Azure Cognitive Services, Tensorflow, Keras which are used for the face feature extraction and face recognition.

Table. 2 Comparison of tools

<table>
<thead>
<tr>
<th>Matlab</th>
<th>OpenCV</th>
<th>OpenMAJ</th>
</tr>
</thead>
<tbody>
<tr>
<td>Matlab it contains extensive libraries and functions for various steps of image processing such as image pre-processing, segmentation, feature extraction and classification or clustering or both.</td>
<td>OpenCV contains different functions for image processing steps such as noise removing(filters), feature extraction, classification and clustering.</td>
<td>OpenMAJ is a java library for image fundamentals, video processing, machine learning, deep learning techniques and facial feature extraction and analysis.</td>
</tr>
<tr>
<td>It also contain functions for face detection and face parts(eyes, nose, mouth) detection.</td>
<td>OpenCV 2.4 contains New FaceRecognize class for face recognition.</td>
<td>In facial analysis, OpenMAJ have functions for face detection, face reorganization and Eigen face method.</td>
</tr>
</tbody>
</table>

6.3 Performance Comparison

True positive (TP), False positive (FP) and False negative (FN) are some of the parameters which are used by many researchers to evaluate the performance and correctness of the proposed algorithms [3]. True positive (TP) is an outcome where the model correctly predicts the output, hence positive and is defined as follows.
R = \frac{TP}{TP+FN} \quad (1)

False positive is where the models correctly predicts the output but the output should have been negative and is defined as follows.

P = \frac{TP}{TP+FP} \quad (2)

False negative is where the model wrongly predicts the output and is defines as follows.

F = \frac{2 \times P \times R}{P+R} \quad (3)

Some of the other evaluation parameters used in [10], are precision (P) which is ratio of true positive to total number of predicted face, Recall (R) which is ratio of number of true positive to the sum of number of true positive and false negatives, simulation time (t) is the time taken by the algorithm for execution. In [15], the authors devised a new algorithm based on extended LBP and it accuracy increased and the performance was measure using parameters like False Acceptance Rate (FAR) which decreased, False Rejection Ratio (FER) which was increased when compared to the original algorithm. In [20], a confusion matrix is used to analyze the performance between the shallow network and deep network. A confusion matrix is a table which is used to describe the performance of classification models. Top 5-error rate is defined as the percentage of the time that the classifier did not include the correct class among its top 5 guesses. In [21], the algorithm devised had a top-5 error rate as low as 6.73%. The Receiver Operator Curve (ROC) is a graphical representation by plotting the true positive rate against the false positive rate and which would show the classification ability of the binary classifier.

7. Conclusion

In this paper, I have analyzed the various face recognition techniques currently available, the step by step procedures of the face detection. A brief description of methods is given in this paper in which the whole process of face recognition is divided into steps and then the review each step of the face recognition algorithm. It has also been analyzed and proved that artificial neural network the current hot topic in research, can lead to face recognition methods which have a faster recognition rate and better accuracy than convectional techniques.

References


[34] Ghaliya Alfarsi, Jasiya Jabbar, Ragad M Tawafak, Abir Alsidiri, Maryam Alsinali, ”Techniques For Face Verification: Literature Review “,International Arab
Conference on Information Technology (ACIT), International Arab Conference on Information Technology (ACIT), 2019.