Evaluation of Face Recognition Methods for Low Quality Images from Multiple Facial Images in a Group Photograph

1Nachamai M, 2Joy Paulose, 3Samuel Marandi

1Associate Professor, Department of Computer Science, Christ University, Hosur Road, Bengaluru-560029, Karnataka, India
2Professor, Department of Computer Science, Christ University, Hosur Road, Bengaluru-560029, Karnataka, India
3Research Associate, Department of Computer Science, Christ University, Hosur Road, Bengaluru-560029, Karnataka, India

ABSTRACT

Background: Implementing a face recognition model for the purpose of biometric identification and authentication is always a challenge. There are numerous approaches available which have allowed the efficiency of a facial recognition system (FRS) to gain an identification rate of 100%. In the recent years, the older techniques have been improved and newer approaches have been formed by combining two or more approaches. One of the most popular techniques in implementing an FRS is to extract features from the image, train a dataset and use the dataset to classify the images using a classifier. The technique to extract features from a face in an image may vary depending on the approach as well as the quality of the image available. However, the efficiency of an FRS depends on a lot of external factors ranging from the quality of the device with which the image or video is taken to the conditions in which they are acquired. As such it is often seen that the performance of an FRS model fails to provide identification to certain faces with extreme conditions which do not meet the prerequisite criteria of the FRS. Objective: In this paper we compare and study such four different approaches and quantify the performance and accuracy of the facial recognition models. These approaches are compared for low resolution images acquired from a group photograph.

INTRODUCTION

A facial recognition system is a computer application for automatically identifying or verifying a person from a digital image or a video frame from a video source. One of the ways to do this is by comparing selected facial features from the image and a facial database (Jafri and Arabnia, 2009).

Facial recognition algorithms identify facial features by extracting markers or features from a face in a photograph. For example, an algorithm may analyze the relative position, size, and/or shape of the eyes, nose, cheekbones and jaw. These features are then used to search for other images with matching features. Other algorithms normalize a gallery of face images and then compress the face data, only saving the data in the image that is useful for face recognition. The image is then compared with the face data. Recognition algorithms can be divided into two main approaches; geometric, which look at distinguishing features and photometric, which is a statistical approach that distills an image into values and compares the values with templates to eliminate variances. Popular recognition algorithms include Principal Component Analysis using eigenfaces, Linear Discriminant Analysis, Elastic Bunch Graph Matching using the Fisherface algorithm, the Hidden Markov Model and the neuronal motivated dynamic link matching.

Biometric identification and authentication systems employed for the purpose of tracking identified and unidentified individuals often require system models that make use of facial features for tracking an individual. The usage of biometric features to track an individual may be inspired from a requirement as simple as an attendance system to highly complex set of features being used, such as in a surveillance video. However, in any case the need for a good facial recognition technique cannot be undermined, since a good face recognition model can reduce the amount of work that may require to be processed manually. Rapid developments in the area of face detection and face recognition have led to numerous techniques, each better than the last and with highly promising features and results. In this paper we consider four techniques that are used for
face recognition, viz. eigenfaces with Karhunen-Loeve(KL) transform(Sandhu et al., 2009), aggregated features technique, wavelets and artificial neural networks(ANN) for face recognition(Agarwal et al.,2010) and 7 state Hidden Markov Model with singular value decomposition (SVD) coefficients for face recognition (Davari, 2008). The objective of this paper is to compare and analyze the efficiency of these four techniques in terms of their efficiency and accuracy when low quality images are given as input. Each of the models will have a training set and an input image. The goal is to quantify how accurately and efficiently an FRS model is able to identify and classify the input image based on the given training set. The dataset that is used for carrying out the analysis is a part of ongoing major research project titled “Facilitating Feedback Reckoning of Teaching Session using Artificial Intelligence” at Christ University, India. It contains multiple low resolution images of faces of multiple individuals for the training set grouped which have been grouped together in a single directory.

**Literature Review:**

Diverse systems require consistent recognition schemes to either confirm or determine the identity of an individual requesting their services. By using biometrics, it is possible to confirm or establish an individual’s identity based on "who she is", rather than by "what she possesses"(Jain et al., 2004). Facial recognition systems form a very integral part of any biometric system. Bevilacqua et al. (2009) in their work propose a face recognition model based on Pseudo 2-D Hidden Markov Model. Their work was based on multiple regions of interest and achieved success rate close to 100%. Phuong et al. (2012) in their work provide a method of locating facial features in face images for different applications. This work was based on Viola-Jones detector algorithm to generate regions of interest and extracting the Haar features based on Adaboost Algorithm. Cheung et al. (2004) in their work have proposed a low feature dimension scheme to solve the problem of face recognition. They have used 2-D Gabor filters to represent face images and claim the method to be more robust than PCA-based method under varying facial expressions. Cooray and O’Connor (2004) have tried to solve the problem by extracting facial features like eyes, nose and mouth apply eigenface theory to these features. Facial features were obtained using Recursive Shortest Spanning Tree (RSST) color segmentation.

**Face Recognition Models:**

In this section, the technical working of the four models is described. The attempt made is to demonstrate the differences in the execution of the four face recognition models. The first technique is based on eigenfaces and Karhunen-Loeve (KL) transform. The working of the system is based on projecting face images onto a feature space that extends significant dissimilarities among known face images (Boualleg et al., 2006).

The significant features are known as “eigenfaces” because they are the eigenvectors (principal components) of the set of faces. Eigenfaces are principal components of a distribution of faces. Eigenfaces for the base set of all images are used to create the covariance matrix. This results in dimension reduction by allowing the smaller set of base images to represent the original images (Zhang and Turk, 2008). The Karhunen–Loeve (KL) transform is a method which is used to approximate a set of vectors or images by a low dimensional subspace. The method provides the optimal subspace, spanned by the KL basis, which minimizes the Mean Squared Error (MSE) between the given set of vectors and their projections on the subspace. Face images are grouped into sets and every set is labeled a class.

It includes the different number of images for each individual and all the faces have different facial expressions in varied lighting conditions. Upon adding a new face image to the training database the class label is labeled, otherwise the new input image has to be processed and compared against every other image in the database. A fixed number of eigenvectors are chosen which is equal to the number of classes available in the training set. This image once classified can be added to the training dataset if not already present in the dataset.

The second technique, i.e. the aggregated features method presents an approach to facial recognition using Features from Accelerated Segment Test (FAST) and Speeded Up Robust Features (SURF). FAST is a corner detection method used to detect and extract feature points which can be used for tracking and mapping objects in a computer vision application at any later point of time. The computational efficiency of this corner detection is quite promising and it is faster than most of the popular feature extraction methods like Difference of Gaussians (DoG) used by scale invariant feature transform and Susan corner detector. SURF is a feature detector used in computer vision tasks like 3D reconstruction and object recognition.

It is mostly influenced by the scale invariant feature transform (SIFT) descriptor. SURF is based on sums of 2D Haar wavelet responses and makes an efficient use of integral images.

For feature detection, it uses the sum of the Haar wavelet response around the point of interest computed with the help of an integral image. By stacking images to "aggregate features" over the training sets a remarkable efficiency for facial recognition can be achieved.
After trying many combinations of detection and extraction methods, it is seen that FAST features and SURF extraction works exceptionally well. The FAST feature detection is tuned to be fairly "permissible"; that is, low values of MinQuality and MinContrast to return a lot of matches. The test image is then matched against each training set, and the match considered is recognized as the match minimizes the metric of Sum of absolute differences. Montages of each training set of the images are created and rather than matching to individual faces or person-averaged faces the features are extracted and are evaluated against the whole training set.

The third technique uses wavelets to extract the features from the input images and uses a neural network classifier for the identification of images. A wavelet is a mathematical function used to divide a given function or continuous-time signal into different scale components. Usually one can assign a frequency range to each scale component. Each scale component can then be studied with a resolution that matches its scale. A wavelet transform is the representation of a function by wavelets (Vijayalakshmi and Vidya, 2014). On the other hand neural networks are a family of statistical learning algorithms inspired by biological neural networks and are used to estimate or approximate functions that can depend on a large number of inputs and are generally unknown. In this method a wavelet transform is applied on the input image and a set of properties is extracted.

These properties are then fed into the Neural Network which has been trained against a database of images. Once the input is fed the network tries to identify the image based on its training and tries to classify the image consequently.

**Fig. 1**: Flow diagram of face recognition technique based on eigenfaces and KL transform
The fourth technique is based on 7-state HMM with SVD coefficients. In this method SVD coefficients are used to extract robust features from face images (Srinivasan, 2013). The first step in this method is to extract the features from the faces. An order statistic filter is applied to the face images to improve the efficiency and accuracy of the algorithm. The features are then extracted and selected from the images and the SVD coefficients are generated.

A quantization process is implemented on the SVD coefficient to reduce the number of observational vectors so that it can be modeled with a discrete HMM. After the representation of the images is complete a selected number of faces are used to train the related HMM. After the learning process is complete each face is associated to a HMM. Now each test image is represented by its own observation vector and the classification of a test image is determined by calculating the probability of an observational vector given each HMM face model. An illustrative diagram to show the working of HMM is given above.

RESULTS AND DISCUSSION

There were numerous challenges that were encountered during the execution of the analysis, chiefly because the above techniques are well suited and designed for high resolution images and the input as well as the training images are required to have consistent dimensions.

If the dimensions are not consistent the accuracy falls drastically. Therefore, to overcome these challenges a model is created wherein the images for both the training set and the input are normalized. An illustrative diagram at figure 5 for the same is provided for the reference.
Fig. 3: Flow diagram of face recognition technique based on wavelet and neural network

Fig. 4: Flow diagram of face recognition techniques based on HMM and SVD coefficients
The first step in this model is to pick an image from the dataset. Since every face recognition model considered for this paper has its own dimension specifications, the images are formatted based on those specifications. This image is then checked against the training set. This image is now given as input to the chosen facial recognition model and the image is classified based on the model. If the image is an existing image present in the training set, it is directly classified. If the image is a new image and not present in the training set but belongs to one of the categories present in the training set, the training set is updated. If the image does not belong to any of the categories of the training set, a new category is added to the training set and that category is updated with the new image. The images in the database are of full frontal view with lighting conditions at normal, low light and high contrast.

The analysis of the above elaborated techniques as mentioned earlier was done for a dataset with extremely low quality images taken under extreme conditions. The images that have been used for analysis are picked from a video file through a sophisticated frame extraction technique. The subject in the video file is a class of post graduate students of Christ University, attending a lecture session. The video file is approximately 11 minutes in length. It has a video frame rate of 25 frames per second and data bitrate of 7996 bits per second. The height and width specification of the video is 1920 and 1080 respectively. A total of 10 frames are picked from the video and a face detection operation is performed on the video. The detected faces are cropped and saved onto the memory in jpeg format to be used as the experimental dataset.

![Flow diagram of face recognition techniques analysis model](image-url)

**Fig. 5:** Flow diagram of face recognition techniques analysis model
Fig. 6: Samples of various images present in the database.

The final dataset obtained has images with the bit depth of 24 and vertical and horizontal resolution of 96 dots per inch (dpi). The width and height specification varies from image to image. The height specification of the images range from 101 to 133 and the width specification of the images range from 111 to 126. The dataset contains multiple images of 26 individuals which include full frontal as well as semi frontal images of the individuals. Since the images were collected in a very dynamic and extreme environment suited for a group images, most of the images also contain overlapping of multiple faces as well and partial faces. A screenshot of the sample images are provided in figure 6 for illustration.

Figures 6(a) and 6(b) are examples of full frontal images. Figures 6(c) and 6(d) are examples for overlapping images. Figures 6(e) and 6(f) illustrate semi frontal views and figures 6(g) and 6(h) are examples of partial faces in the dataset.

<table>
<thead>
<tr>
<th>Table 1: Comparison Face Recognition Techniques</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Approach Name</strong></td>
</tr>
<tr>
<td>Aggregated Features</td>
</tr>
<tr>
<td>Wavelets and Neural Networks</td>
</tr>
<tr>
<td>Eigen faces with KL</td>
</tr>
<tr>
<td>7 State HMM with SVD</td>
</tr>
</tbody>
</table>

Each of these techniques was tested for the respected success rate for trained images and untrained images and the feasibility of use of the techniques for tweaks to make it usable with the current database. The feasibility is the amount of freedom the approach allows in terms of changing the dimensions and classification criteria, to be tweaked in order to accustom it to the current database. The results in Table I are the consolidated figures after running the module multiple times with different inputs. The execution time recorded is the maximum time that was required by a facial recognition model to identify an untrained image. This was recorded using the tic/toc functions available in MATLAB since the implementation of these approaches were done in MATLAB.

**Conclusion:**
Among all the four techniques the 7 State HMM with SVD coefficient appears to have fared the best against untrained images. Even though the execution time of the aggregated features technique was the best it failed to classify untrained images and the success rate was quite low for this method. The model based on wavelets and neural network was very successful in identifying trained images but when it came to untrained images the network failed to classify the images correctly.
However, the reason for low rate of accuracy of the approaches may also be attributed to the low and degraded quality of the images. It may not always be possible for the images to be of ideal quality every time when face recognition model is implemented and hence there is a necessity to improve these current models for low quality images.

ACKNOWLEDGEMENT

Evaluation of Face Recognition Methods for Low Quality Images from Multiple Facial Images in a Group Photograph is part of major research project undertaken by Department of Computer Science, Christ University. We would like to thank the Center For Research, Christ University and the management for their incessant support and opportunity to work on the project. This work was carried out as part of the major research project titled ‘Facilitating Feedback Reckoning of Teaching Session using Artificial Intelligence’ fully supported and funded by Center For Research, Christ University under the Grant No: MRPDHS-1210.

REFERENCES


