A Comprehensive Survey of Feature Extraction and Feature Selection Techniques of Face Recognition System

Swati Saini¹, Dr. Preeti Malhotra²

(mail to: sainiswati39@gmail.com¹, preetianeja2005@gmail.com²)

Student, Computer Science and Engineering Department, Hindu College of Engineering, Sonepat, Haryana, India¹,
Assistant Professor, Computer Science and Engineering Department, Hindu College of Engineering, Sonepat, Haryana, India²

Abstract. Among various types of biometric techniques it is the face recognition that gains huge popularity in the last few decades due to its vast applications and more user friendly nature. Face Recognition is the procedure of identification and verification of a person's identity using his face. Main steps involved in a FR System are Face Detection, Feature Extraction, Feature Selection and Recognition. Feature Extraction and Feature Selection are the two main phases to be focused on in order to get a good FR System. In this paper, we have carried out an extensive literature review on face recognition focusing on feature extraction and feature selection phases a bit more.

Keyword: Face recognition (FR), feature extraction, principal component analysis (PCA), feature selection.

1. Introduction

Biometrics is the metrics related to human characteristics. Biometrics identify humans using their two types of characteristic namely- physiological and behavioral. Physiological traits involves face, finger, retina, DNA, etc. whereas behavioral traits involves gait, signature, keystroke, etc.[1]. Among all the above mentioned biometrics, it is the face recognition which has been the most fascinating topic for researchers since the past few decades being it convenient, more users friendly and an inexpensive technique of identification. The main phases of an FR System are: Face Detection, Feature Extraction, Feature Selection and Face Recognition. In the Face Detection phase, a human face is detected from the captured image/video. After detection, features are extracted from the captured face in feature extraction phase and only relevant features are selected from all the extracted features in feature selection phase in order to get much accurate results. Then comes the Face Recognition phase in which the captured face is compared with the models already present in database and recognition is done. The goodness of an FR System depends majorly on Feature Extraction and Feature Selection phase. To combine the process of Face Recognition altogether we can say, in an FR System, a new face is given to the system and is compared with the face
templates/models already stored in the database and then is classified as a known individual if a correspondence is found [2].

Following this introduction is Section 2 presenting basics of anFR System including the brief introduction to various phases of a FRS. Then, Section 3 and Section 4 respectively discuss the basics and techniques of Feature Extraction and Feature Selection phase along with their literature review. Section 5 represents results inferred from literature survey and finally paper concludes in Section 6.

2. **Face Recognition System**

AnFR System involves various steps shown in Figure 1. In this, an image or a video is fed as an input which is then pre-processed in order to improve quality of the input data. After that face is cropped from the whole image and background is removed in order to detect the face which is then followed by the Feature Extraction and Feature Selection phase in which the facial features are first extracted and then the most relevant ones are chosen. In the last phase, i.e., Face Recognition phase, matching is done with a help of a templates generated for enrollment in the FRS for verification/identification of a particular person [3]. Verification answers "Is the person who they say they are?" whereas Identification give answer to "Who is this person?".

![Figure 1: Key steps involved in Face Recognition Process](image)

The primary purpose of Face Detection phase is to check whether human faces are present in the input image/video or not; if yes, then what is the location of those faces. The survey written by Wei-Lun Chao [4] classified face detection algorithms in four categories, namely: Knowledge-based methods, Feature Invariant Approach, Template Matching and Appearance-based Methods. *Knowledge-based methods*[5] are the rule based methods and deduce the relationship between facial features using some pre-defined rules regarding facial features and their relationship with each other. They use the human knowledge to determine how human face actually looks like. Algorithms following *Feature Invariant approach*[6] focuses on to look for structural features even in bad illumination conditions and pose variations and then locate faces in the input image using those features. In the *Template Matching*[7] methods category, various standard patterns of human faces (either whole or partial) are already stored in the system and then the input image is matched with those patterns to detect whether a human face is present in the input image or not. The training images which capture the representative variability of facial appearance are used to determine models which are then used to detect faces in *Appearance-based methods* [8]. Face Recognition is extensively getting used in various sectors be it security-law enforcement, health or marketing & retail. Face Recognition is cheaper, restricts people involvement, convenient and less time consuming. 'Illumination' and the 'pose variation' are two most common issues present in face recognition. Same pictures appear to be a different picture under several
illumination effects. The functioning of an FR System degrades when large pose variations are present in the input images [9].

3. Feature Extraction

After detecting face in an image, the next step is to extract useful features only from the input image. Feature extraction is done to get rid of superfluous and beside the point features from the input image. The methods used to extract useful feature are divided into two types namely, Holistic Matching methods and Feature-based matching methods. The HolisticMatching Methods [10] or Appearance-based methods use entire face as an input to a recognition system whereas in the Geometric Feature[11] based Matching Methods the location of different features such as eyes, mouth etc. are first extracted and then fed as an input to an FR System [9]. Some very popular feature extraction techniques are:

3.1 Principal Component Analysis (PCA)

Karhunen-Loeve method [2] is an another name for PCA and is one of the most popular methods for solving the problem of overfitting. PCA is a dimension reduction technique with the goal to extract the most relevant information contained in the image. This reduced data is then used for recognition purpose[12]. A face contains certain set of features which are known as either the principal components or the Eigenfaces[13]. In PCA based face recognition, increasingeigen value is directly proportional to the recognition rate but the recognition rate become stagnant after a certain increase in the eigen value[12]. It is an unsupervised learning algorithm.

Various methods for face recognition are present in the literature such as Gabor Features[14], Diagonal Principal Component Analysis (DiaPCA)[15], PCA, ICA, LDA[16], a combination of PCA and LDA (PCA-LDA)[17][18]. Peng et al.[14] in 2004 carried out face recognition experiment on FERET database using Ada-Boosted Gabor Features and results showed that the rank 1 recognition rate improved to 95.2% from 37.5% when 700 features were selected against 6 features; also when re-sampling strategy was applied rank 1 recognition rate shoots to 92.8% (1740 features) from 90.6% (700 features). The Gabor Feature method was then combined with PCA and KPCA (Kernel PCA) by Vinay et al.[19] in 2015 and the results proved that on Euclidean distance measure, the linear Gabor-PCA outperformed the non-linear Gabor-KPCA by 6.67%, 4.17% using MAHCOS, 0.83% on the cosine measure and 12.00% using the City-Block distance measure.

PCA being the simplest of all techniques is widely used for the dimension reduction process. However various new methods were proposed that were not just the extension of widely used PCA but also proved their superiority over PCA. Daoqiang et al. [15]in 2006 presented diagonal principal component analysis (DiaPCA) which as compare to standard PCA looks for best projective vectors from diagonal face images without image-to-vector transformation. DiaPCA(90.50% recognition accuracy) performed better than both PCA (85.50% accuracy) and 2DPCA (85.50% accuracy) whereas the performance of DiaPCA further improved when it was combined with 2DPCA(DiaPCA+2DPCA=91.50% accuracy).

To solve the memory requirement burden and computational cost issues of existing PCA, Haitao et al. [20]in 2006 proposed a method named Incremental PCA (IPCA) whose average recognition rate difference with PCA was less than 1%. 
3.2 Linear Discriminant Analysis (LDA)

LDA is a supervised learning algorithm for dimension reduction which tries to preserve class discrimination information as much as possible. In case where number of samples per class are less, PCA performs better, whereas with large datasets having multiple classes, LDA works better[21]. LDA is also called the Fisherface method (13). The LDA is considered as a better option than PCA and the main aim of LDA is to find a base of vectors which provides best discrimination among classes[2].

Hidayat et al. [22] in 2011 carried out a comparative study between PCA and LDA for feature extraction in face recognition on various datasets such as FACES94, FACES95, FACES96, JAFFE, etc. Results show that PCA outsmarts LDA where time is the constraint whereas LDA(FACES94- 99.90%, FACES95- 90.80%, FACES96-97.20%, JAFFE-100.00%) performed better than PCA(FACES94- 99.90%, FACES95- 87.00%, FACES96-94.00%, JAFFE-92.60%) when recognition rate is to be considered. Then Tomesh et al. [17] in 2013 combined PCA with LDA where former was used for dimension reduction whereas latter extracted the features. This combination on ORL face database gave 96.35% recognition rate beating various previously adopted methods. Bala et al. [23] in 2016 combined LDA(feature extraction) and Euclidean Distance classifier(for classification). Experiments carried over ORL database consisting 400 images of 40 people having 10 different poses each showed a recognition rate of 93.70%.

3.3 Discrete Cosine Transform (DCT)

DCT is one of the most popular techniques of signal as well as image processing. It was initially introduced in early 70s. Once the DCT is enforced on input image, the DCT coefficients are followed in zig-zag manner in order to convert image matrix into feature vector. Only a few DCT coefficients are enough for reducing redundancy and recovering original image from the selected coefficients[24].

Azam et al. [25] in 2010 presented a method of face recognition in which DCT was used for feature extraction and Artificial Neural Network(ANN) was used for recognition purpose. ORL, YALE and FERET databases were used for carrying out experiments where the result of the proposed method gave 98.01% recognition rate on ORL database and 92.77% on YALE database whereas in 2012 a combination of DCT for feature extraction and Nearest Neighbor Discriminant Analysis for face recognition by Tyagi et al.[26]. This method outperformed the one presented in [25] and gives a relatively high recognition rate of 99% (on ORL database) and 98.50% (over Yale database). In 2017 Preeti et al. [18] applied DCT-PCA+ CS and LDA+CS on FR System over ORL and Yale databases. The results proved the efficiency of proposed method over existing PCA and LDA.

4. Feature Selection

After feature extraction, the most important phase in face recognition process is the 'feature selection' phase. In Feature Selection phase, the most appropriate features that are needed by an FR system to deduce the desired results
are selected. In other words, we can say that an FR System works on the "Garbage in Garbage Out" strategy. The key steps of the feature selection process are shown in Figure 2.

Subset Genesis is a heuristic lookup in which each state defines a test subset for evaluation. The nature of subset genesis procedure is determined by: successor generation and search organization. The new generated subset is then evaluated by some evaluation criteria to determine the goodness of generated candidate subset. Hence, a stopping criteria has to be determined to stop the selection process. At last, the generated subset is validated in the validation phase by carrying out various comparisons and tests with already established results. The three general approaches to feature selection are: Filter Approach, Wrapper Approach and Embedded Approach. Filter Approach selects the feature subset on the basis of scores of features in different statistical test instead of any learning algorithm. These methods are generally used as a pre-processing step. Wrapper Approach uses a learning algorithm for evaluating generated subset and depending upon the results of previous model, features are added or removed from the feature subset. Embedded Approach is a combination of both the Filter Approach and the Wrapper Approach. Using embedded approach methods, the feature subset are generated using the algorithms having their own built-in feature selection methods.

Figure 2: Key steps involved in Feature Selection Process

Various optimization techniques present in literature such as Bat Algorithm, Particle Swarm Optimization (PSO), Genetic Algorithm (GA), Cuckoo Search Optimization Algorithm (CS), Ant Colony Optimization (ACO), Salp Swarm Algorithm (SSA) etc. helps in selecting features in a way that FRSystem gives as much accurate results as possible. Feature selection finds its application in fields like Text Categorization, Intrusion Detection, Image Retrieval, Remote Sensing etc.

As we know PCA is used for extracting features and Memetic Algorithm is used for selecting features, Dinesh et al. in 2009 combined both the strategies and applied PCA-MA approach on feature selection. The proposed approach clearly and completely outperformed the eigenface method as well as PCA led Genetic Algorithm (PCA-GA). The eigenface method with 30 features produced 97.14% recognition rate whereas PCA-MA produced 100% recognition rate with just 17 features. Despite PCA being the widely used and easiest algorithm it has some limitations as well such as large computational load and poor discriminatory power. To overcome these limitations Manisha et al. in 2014 proposed a fresh face recognition method based on the combination of PCA, Wavelet and SVM where low frequency component of wavelet were used for PCA representation, features were selected...
using Genetic Algorithm (GA) and Support Vector Machine along with nearest neighbor classifier (ND) was used for classification purpose. Experiments conducted on various databases such as ORL, FERET, Yale and YaleB proved the superiority of proposed method (ORL- 98.00%, FERET- 97.30%, Yale- 100.00%, YaleB- 100.00%) over other leading methods. Preeti et al. [30] in 2017 showed an FR System using combination of DCT and PCA for dimension reduction and feature extraction succeeded by Bat Algorithm to select the best feature set. Experimental results proved the superiority of proposed method (97.5% recognition rate with just 24 features) over other leading optimization techniques such as CS (97% with 48 features), GA(96.5% with 28 features) and PSO (96.5% with 28 features). They further carried forward their research in 2017 in whichPreeti et al. [33] suggested an FR system using the Cuckoo Search (CS) optimization technique where features were extracted using combination of PCA and DCT in order to achieve high accuracy rates. Results show that CS gave 88% recognition rate with 10 features only which are equivalent to just 10% of the features used in PCA and gave 96.50% recognition rate with just 34 features which is approximately 35% of the features used in DCT-PCA approach. Moving further Preeti et al. [39] in 2018 studied effects of various distance classifiers on performance of an FR System over ORL database. Features were extracted using DCT-PCA and LDA and further selection was done using Cuckoo Search Algorithm. Among various distance classifiers Euclidean distance classifier produced the utmost-93% recognition rate followed by Manhattan Distance-92%, Chebyshev Distance-89% and Canberra Distance-83% in absence of motion blur whereas in presence of motion blur results were: Euclidean Distance Classifier-91%, Manhattan Distance-86%, Chebyshev Distance-83% and Canberra Distance generated 74% recognition rate. In 2019 Preeti et al. [40] further studied the effects of various distance classifiers on the performance of FR system based on Bat and CS algorithm over Yale_Original and Yale_Extended databases and results proved the supremacy of DCTPCA + Bat over PCA and DCTPCA + CS algorithms[41][42].

5. Observations

The following observations have been summarized from the literature review and computed in a tabular form.

Table 1: Results summarized from literature survey.

<table>
<thead>
<tr>
<th>Reference Paper</th>
<th>Technique Used</th>
<th>Database</th>
<th>Recognition Rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>(14)</td>
<td>Ada-Boosted Gabor Features Resampling Technique</td>
<td>FERET</td>
<td>37.50% (6 features) to 95.20% (700 features) 90.60% (700 features) to 92.80% (1740 features)</td>
</tr>
<tr>
<td>(15)</td>
<td>PCA 2DPCA DiaPCA</td>
<td>FERET</td>
<td>85.50% 85.50% 90.50%</td>
</tr>
<tr>
<td>Method</td>
<td>Description</td>
<td>Database</td>
<td>Accuracy</td>
</tr>
<tr>
<td>--------</td>
<td>-------------</td>
<td>----------</td>
<td>-----------</td>
</tr>
<tr>
<td>DiaPCA + 2DPCA</td>
<td></td>
<td>ORL</td>
<td>91.50%</td>
</tr>
<tr>
<td>PCA + LDA</td>
<td></td>
<td>ORL</td>
<td>96.35%</td>
</tr>
<tr>
<td>PCA</td>
<td>FACES94, FACES95, FACES96, JAFFE</td>
<td></td>
<td>99.90%, 87.00%, 94.00%, 92.60%</td>
</tr>
<tr>
<td>LDA</td>
<td>FACES94, FACES95, FACES96, JAFFE</td>
<td></td>
<td>99.90%, 90.80%, 97.20%, 92.60%</td>
</tr>
<tr>
<td>LDA + Euclidean Distance Classifier</td>
<td>ORL</td>
<td></td>
<td>93.70%</td>
</tr>
<tr>
<td>DCT + ANN</td>
<td>ORL, Yale</td>
<td></td>
<td>98.01%, 92.77%</td>
</tr>
<tr>
<td>DCT + Nearest Neighbor Distance Analysis</td>
<td>ORL, Yale</td>
<td></td>
<td>99.00%, 98.50%</td>
</tr>
<tr>
<td>DCT + PCA + BAT</td>
<td>ORL</td>
<td></td>
<td>97.50% (24 features)</td>
</tr>
<tr>
<td>PCA + Wavelet + SVM + GA + ND</td>
<td>ORL, FERET, Yale, YaleB</td>
<td></td>
<td>98.00%, 97.30%, 100.00%, 100.00%</td>
</tr>
<tr>
<td>PCA + DCT + CS</td>
<td>ORL</td>
<td></td>
<td>88.00% (10 features), 96.50% (34 features)</td>
</tr>
<tr>
<td>PCA + GA, PCA + MA</td>
<td>ORL</td>
<td></td>
<td>97.14% (30 features), 100.00% (17 features)</td>
</tr>
</tbody>
</table>

### 6. Conclusion

This paper seeks to go through considerable number of research and review papers in order to cover developments made till date in the field of Face Recognition. Face Recognition has been an attractive field for researchers from past few decades and is still the most favored topic for research owing to its vast practical applications. In this paper, we presented an extensive review of significant work done in the field of feature extraction and feature selection as well. The main challenges for face recognition mainly include pose variation, illumination, expressions and occlusion. Some of the most important techniques such as PCA, LDA, DCT, Bat Algorithm, Cuckoo Search Algorithm, etc. are explained well in this paper. The results obtained by various researchers using numerous different techniques are also compiled together in a tabular form in this paper. The objective of this literature survey is to help the potential researchers in the field of face recognition to understand the various methods and techniques present and to support their further researches.
References


