**CHAPTER 2**

**REVIEW OF RELATED LITERATURES AND SYSTEMS**

*This chapter presents the review of related literature and studies from both local and foreign sources, articles, forums and other knowledge bases about the study and the overall synthesis.*

**I. Foreign Literatures**

According to **Jayanta Kumar Basu, Debnath Bhattacharyya, and Tai-hoon Kim in their “Use of Artificial Neural Network in Pattern Recognition” of 2011**, the main characteristics of neural networks are that they have the ability to learn complex nonlinear input-output relationships, use sequential training procedures, and adapt themselves to the data. The most commonly used family of neural networks for pattern classification tasks is the feed-forward network, which includes multilayer perceptron and Radial-Basis Function (RBF) networks. Another popular network is the Self-Organizing Map (SOM), or Kohonen-Network, which is mainly used for data clustering and feature mapping. The learning process involves updating network architecture and connection weights so that a network can efficiently perform a specific classification/clustering task. The increasing popularity of neural network models to solve pattern recognition problems has been primarily due to their seemingly low dependence on domain-specific knowledge and due to the availability of efficient learning algorithms for practitioners to use. Artificial neural networks (ANNs) provide a new suite of nonlinear algorithms for feature extraction (using hidden layers) and classification (e.g., multilayer perceptrons).

In addition, existing feature extraction and classification algorithms can also be mapped on neural network architectures for efficient (hardware) implementation. An ANN is an information processing paradigm that is inspired by the way biological nervous systems, such as the brain, process information. The key element of this paradigm is the novel structure of the information processing system. It is composed of a large number of highly interconnected processing elements (neurons) working in unison to solve specific problems. An ANN is configured for a specific application, such as pattern recognition or data classification, through a learning process. Learning in biological systems involves adjustments to the synaptic connections that exist between the neurons.

According to **Wei-Lun Chao, “Face Recognition” of 2010**, Feature Extraction: After the face detection step, human-face patches are extracted from images. Directly using these patches for face recognition have some disadvantages, first, each patch usually contains over 1000 pixels, which are too large to build a robust recognition system1. Second, face patches may be taken from different camera alignments, with different face expressions, illuminations, and may suffer from occlusion and clutter. To overcome these drawbacks, feature extractions are performed to do information packing, dimension reduction, salience extraction, and noise cleaning. After this step, a face patch is usually transformed into a vector with fixed dimension or a set of fiducial points and their corresponding locations. We will talk more detailed about this step in Section 2. In some literatures, feature extraction is either included in face detection or face recognition.

Face Recognition: After formulizing the representation of each face, the last step is to recognize the identities of these faces. In order to achieve automatic recognition, a face database is required to build. For each person, several images are taken and their features are extracted and stored in the database. Then when an input face image comes in, we perform face detection and feature extraction, and compare its feature to each face class stored in the database. There have been many researches and algorithms proposed to deal with this classification problem, and we’ll discuss them in later sections. There are two general applications of face recognition, one is called identification and another one is called verification. Face identification means given a face image, we want the system to tell who he / she is or the most probable identification; while in face verification, given a face image and a guess of the identification, we want the system to tell true or false about the guess.

Also, In order to achieve automatic recognition, a face database is required to build. For each person, several images are taken and their features are extracted and stored in the database. Then when an input face image comes in, we perform face detection and feature extraction, and compare its feature to each face class stored in the database. There have been many researches and algorithms proposed to deal with this classification problem. There are two general applications of face recognition, one is called identification and another one is called verification. Face identification means given a face image, we want the system to tell who he / she is or the most probable identification; while in face verification, given a face image and a guess of the identification, we want the system to tell true or false about the guess.

According to **Omaima N. A. AL-Allaf’s, “Review of Face Detection Systems Based Artificial Neural Networks Algorithms” of 2014**, Zoran and Samcovic (2006) used ANN for face detection for video surveillance. The ANN is trained with multilayer back propagation neural networks (BPNN). Three face representations were taken (pixel, partial profile and eigenfaces) representation. Three independent sub-detectors are generated based on these three face representation. Aamer Mohamed, et al (2008) proposed face detection system based on BPNN via Gaussian mixture model to segment image based on skin color. In this approach start from skin and non-skin face candidates' selection. After that the features are extracted from discrete cosine transform (DCT) coefficients. Based on DCT feature coefficients in Cb and Cr color spaces, BPNN was used to train and classify faces. The BPNN used to check if the image include face or not. DCT feature values of faces that represent the data set of skin/non-skin face candidates obtained from Gaussian mixture model are fed into BPNN to classify whether original image includes a face or not.

Face detection is first step in face recognition systems to localize and extract the face region from the image background. The literature studies related to face detection systems which were based on ANN were described earlier in this research can be summarized as follows:

• The face detection techniques presented above were based on 2D image detection.

• Many of these literature researches did not give overview about the exactly used database for system training and testing.

• Many of these researches did not give sufficient information about performance measures used for face detection. There is lack of equations related to these performance measures.

• Most of these studies of face detection systems were adopted ANN in combination with other approaches and algorithms to obtain better results for detection and improve the performance of face detection system. But this may increase the system complexity, required memory and time for face detection.

• Lack of using other significant ANN architectures such as self-organizing map, PatternNet ANN, Fast BPNN and so on.

• Lack of literature related to face detection based on combination of ANN and genetic algorithm (GA). According to above points, many recommendations must be taken in our consideration when we suggest to build a strong face detection system such as:

• Try to design a real time face detection system that is based on video taken in real time camera. • Give sufficient details about the exactly used database for system training and testing.

• Give sufficient details about performance measures and equations used for face detection.

• ANN can be adopted in combination with other algorithms to obtain better results for face detection. At the same time, we must focus on how to simplify the combined algorithms steps to reduce the memory required and processing time.

• Try to use other ANN architectures: self-organizing map, PatternNet, FitNet and Fast BPNN.

• Try to use different optimization ANN training algorithms such as: Levenberg-Marquardt (TRAINLM); TRAINBFG; Bayesian regularization (TRAINBR); TRAINCGF algorithm; Gradient descent (TRAINGD); and Gradient descent with momentum (TRAINGDM) to obtain best results for face detection system.

• Try to use genetic algorithm (GA) as an optimization algorithm to obtain the best values of ANN algorithm parameters that result to optimal results.

According to **Shamla Mantri and KalpanaBapat’s, “Neural Network Based Face Recognition Using Matlab” of 2011**, Face recognition scenarios can be classified into two types, Face verification (or authentication) and Face identification (or recognition). 1) Face verification: It is a one-to-one match that compares a query face image against a template face image whose identity is being claimed. To evaluate the verification performance, the verification rate (the rate, at which legitimate users are granted access) vs. false accepts rate (the rate at which imposters are granted access) is plotted, called ROC curve. A good verification system should balance these two rates based on operational needs. 2) Face identification: It is a one-to-many matching process that compares a query face image against all the template images in a face database to determine the identity of the query face.

The identification of the test image is done by locating the image in the database that has the highest similarity with the test image. The identification process is a “closed” test, which means the sensor takes an observation of an individual that is known to be in the database. The test subject’s (normalized) features are compared to the other features in the system’s database and a similarity score is found for each comparison. These similarity scores are then numerically ranked in a descending order. The percentage of times that the highest similarity score is the correct match for all individuals is referred to as the “top match score”.

According to **M.Nandini, P.Bhargavi and G.RajaSekhar’s, “Face Recognition Using Neural Networks” of 2013,** The Back propagation training takes place in three stages: 1. Feed forward of input training pattern 2. Back propagation of the associated error and 3. Weight adjustment. During feed forward, each input neuron (p1) receives an input value and broadcasts it to each hidden neuron, which in turn computes the activation and passes it on to each output unit, which again computes the activation to obtain the net output.

Input image is acquired by taking photographs using the digital camera. These images are taken in color mode and saved in JPG format.

During training, the net output is compared with the target value and the appropriate error is calculated. From this, the error factor is obtained which is used to distribute the error back to the hidden layer. The weights are updated accordingly. In a similar manner, the error factor is calculated for units. After the error factors are obtained, the weights are updated simultaneously. The output layer contains one neuron.

According to **Athira Prem and Bineesh** **V.’s, “Anti-Spoofing of a Face Image Using Diffusion Speed Model” of 2016,** the imposter can rotate, shift, and bend the valid user’s photo before the camera like a live person to fool the authentication system. It is a challenging task to detect whether the input face image is from a live person or from a photograph.

According to **Sushma Jaiswal, Dr. (Smt.) Sarita Singh Bhadauria and Dr. Rakesh Singh Jadon’s, “Comparison between Face Recognition Algorithm-faces, Fisherfaces and Elastic Bunch Graph Matching” of 2011,** eigenface is a practical approach for face recognition. Due to the simplicity of its algorithm, we could implement an Eigenface recognition system easily. Besides, it is efficient in processing time and storage. PCA reduces the dimension size of an image greatly in a short period of time. The accuracy of Eigenface is also satisfactory (over 90 %) with frontal faces. However, as there has a high correlation between the training data and the recognition data. The accuracy of Eigenface depends on many things. As it takes the pixel value as comparison for the projection, the accuracy would decrease with varying light intensity. Besides, scale and orientation of an image will affect the accuracy greatly.

Pre-processing of image is required in order to achieve satisfactory result Advantages of this algorithm are that the eigentfaces were invented exactly for that purpose what makes the system very efficient. A drawback is that it is very sensitive for lightening conditions and the position of the head, it Fast on Recognition, and Easy to implement Disadvantages-Finding the eigenvectors and eigenvalues are time consuming on PPC The size and location of each face image must remain similar PCA (Eigenface) approach maps features to principle subspaces that contains most energy.

According to **Thai Hoang Le’s, “Applying Artificial Neural Networks for Face Recognition” of 2011,** One of the popular methods having the same achievement as well is artificial neural networks (ANNs). ANN is the term on the method to solve problems by simulating neuron’s activities. In detail, ANNs can be most adequately characterized as “computational models” with particular properties such as the ability to adapt or learn, to generalize, or to cluster or organize data, and which operation is based on parallel processing. However, many of the previously mentioned properties can be attributed to non-neural models.

One of the most important steps in the face recognition problem is the facial feature extraction. A good feature extraction will increase the performance of face recognition system. Various techniques have been proposed in the literature for this purpose and are mainly classified in four groups. (1) *Geometric feature-based method group*: the features are extracted by using relative positions and sizes of the important components face such as eyes, nose, mouth and other important component of face. The advantage of these methods is the concentration on important components of face such as eyes, nose, and mouth but the disadvantage is not to remain face global structure. (2) *Template-based method Group*: based on a template function and appropriate energy function, this method group will extract the feature of important components of face such as eyes and mouth, or face shape. An image region is the best appropriateness with template (eyes, mouth, etc.) which will minimize the energy. Advantages of this group method are using template and determining parameter for important components of face, but disadvantage is not to reflect face global structure. (3) *Color segmentation-based method group*: this group method is based on skin’s color to isolate the face. (4)*Appearance-based method group*: The goal of this method group is using linear transformation and statistical methods to find the basic vectors to represent the face.

Methods have been proposed in the literature for this aim such as PCA and ICA. In detail, goal of PCA method is to reduce the number of dimensions of feature space, but still to keep principle features to minimize loss of information. PCA method uses second-order statistic (covariance matrix) in the data. However, PCA method has still disadvantages. High-order dependencies still exist in PCA analysis, for example, in tasks as face recognition, much of the important information may be contained in the high-order relationships among the image pixels, not only second order. Therefore, we need to find a method more general than PCA; ICA is a satisfying method. Instead of principle component analysis, ICA uses technique-independent component analysis, an analysis technique that not only uses second-order statistic but also uses high-order statistic (kurtosis).

PCA can be derived as a special case of ICA which uses Gaussian source models. In this case, the mixing matrix cannot determine. PCA is not the good method in cases of non-Gaussian source models. In particular, it has been empirically observed that many natural signals, including speech and natural images, are better described as linear combinations of sources with “super-Gaussian” distributions (kurtosis positive). In this case, ICA method is better than PCA method because (1) ICA provides a better probabilistic model of the data. (2) It uniquely identifies the mixing matrix. (3) It finds an unnecessary orthogonal basic which may reconstruct the data better than PCA in the presence of noise such as variations lighting and expressions of face. (4) It is sensitive to high-order statistics in the data, not just to the covariance matrix.The appearance-based method group has been found the best performer in facial feature extraction problem because it keeps the important information of face image, rejects redundant information, and reflects face global structure.

In many applications of face recognition such as identity authentication for credit card and video surveillance, accuracy of face recognition problem is the best important factor. Therefore, besides principle components, independent components of data and face global structure are kept by PCA and ICA method. Combination of these features with geometric features such as nose, eyes, and mouth in recognition will increase accuracy, confident of face recognition system.In this section, we present architectures of ICA for face recognition and combining ICA method with geometric feature-based method (GICA), then comparison of GICA method and GPCA method on CalTech database. Since then, practicability of GICA method for face recognition problem is demonstrated.

According to **Saptarshi Chakrabortym and Dhrubajyoti Das’, “An Overview of Face Liveness Detection” of 2014**, Imposters will try to introduce a large number of spoofed biometrics into system. With the help of liveness detection, the performance of a biometric system will improve. It is an important and challenging issue which determines the trustworthiness of biometric system security against spoofing. In face recognition, the usual attack methods may be classified into several categories. The classification is based on what verification proof is provided to face verification system, such as a stolen photo, stolen face photos, recorded video, 3D face models with the abilities of blinking and lip moving, 3D face models with various expressions and so on.

Anti-spoof problem should be well solved before face recognition systems could be widely applied in our daily life.

The basic purpose is to differentiate between live face and fake face (2-D paper masks) in terms of shape and detailedness. The authors have proposed a single image-based fake face detection method based on frequency and texture analyses for differentiating live faces from 2-D paper masks. The authors have carried out power spectrum based method for the frequency analysis, which exploits both the low frequency information and the information residing in the high frequency regions.

Moreover, description method based on Local Binary Pattern (LBP) has been implemented for analyzing the textures on the given facial images. They tried to exploit frequency and texture information in differentiating the live face image from 2-D paper masks. The authors suggested that the frequency information is used because of two reasons. First one is that the difference in the existence of 3-D shapes, which leads to the difference in the low frequency regions which is related to the illumination component generated by overall shape of a face. Secondly, the difference in the detail information between the live faces and the masks triggers the discrepancy in the high frequency information. The texture information is taken as the images taken from the 2-D objects (especially, the illumination components) tend to suffer from the loss of texture information compared to the images taken from the 3-D objects. For feature extraction, frequency-based feature extraction, Texture-based feature extraction and Fusion-based feature extraction are being implemented.

For extracting the frequency information, at first, the authors have transformed the facial image into the frequency domain with help of 2-D discrete Fourier transform. Then the transformed result is divided into several groups of concentric rings such that each ring represents a corresponding region in the frequency band. Finally, 1-D feature vector is acquired by combining the average energy values of all the concentric rings. For texture-based feature extraction, they used Local Binary Pattern (LBP) which is one of the most popular techniques for describing the texture information of the images. For the final one i.e. fusion-based feature extraction, the authors utilizes Support Vector Machine (SVM) classifier for learning liveness detectors with the feature vectors generated by power spectrum-based and LBP-based methods. The fusion-based method extracts a feature vector by the combination of the decision value of SVM classifier which are trained by power spectrum-based feature vectors and SVM classifier which are trained by LBP-based feature vectors.

The authors have used two types of databases for their experiments: BERC Webcam Database and BERC ATM Database. All the images in webcam database were captured under three different illumination conditions and the fake faces (non-live) were captured from printed paper, magazine and caricature images. Experimental results of the proposed approach showed that LBP based method shows more promising result than frequencybased method when images are captured from prints and caricature.

Overall, the fusion-based method showed best result with error rate of 4.42% compared to frequency based with 5.43% and LBP-based method with 12.46% error rate. Similar technique of face spoofing detection from single images using micro-texture analysis was implemented by Jukka et al. The key idea is to emphasize the differences of micro texture in the feature space. The authors adopt the local binary patterns (LBP) which is a powerful texture operator, for describing the micro-textures and their spatial information. The vectors in the feature space are then given as an input to an SVM classifier which determines whether the micro-texture patterns characterize a fake image or a live person image. The first step is to detect the face, which is then cropped and normalization is done and converted into a 64 × 64 pixel image. Then, they applied LBP operator on the normalized face image and the resulting LBP face image is then divided into 3×3 overlapping regions. The local 59-bin histograms obtained from each region are then computed and collected into a single 531-bin histogram. Then, two other histograms obtained from the whole face image are computed using LBP operators. Finally, a nonlinear SVM classifier with radial basis function kernel is used for determining whether the input image is a fake face or live person image. The experimental results showed that LBP has the best performance with equal error rate (EER) of 2.9% in comparison with other texture operators like Local Phase Quantization and Gabor Wavelets with EER of 4.6% and 9.5% respectively.

Another method for texture based liveness detection based on the analysis of Fourier Spectra of a single face image or face sequence image was introduced by Li et al. Their method is based on structure and movement information of live face. Their algorithm is based on two principles: first, as the size of the photo is smaller than that of live face and the photo is flat, high frequency components of photo images is less than those of real face images and secondly, even if a photo is held before a camera and is in motion, as the expressions and poses of the face contained in the photo does not vary, the standard deviation of frequency components in a sequence is small. The authors have suggested that an effective way to live face detection is to analyze 2D Fourier spectra of the input image.

They calculated the ratio of the energy of high frequency components to that of all frequency components as the corresponding high frequency descriptor (HFD). According to the authors, high frequency descriptor of the live face should be more than a reasonable threshold Tfd. The high frequency components of an image are those whose frequencies are greater than two third of the highest radius frequency of the image and whose magnitudes are also greater than a threshold Tf (generally, the magnitude of high frequency components caused by the forgery process is smaller than that of original image.). The authors have found out that the above the above method will be defeated if a very clear and big size photo is used to fool the system. To solve this problem, motion images were exploited for the live face detection. So, via monitoring temporal changes of facial appearance over time, where facial appearance is represented by an energy value defined in frequency domain, is an effective approach to live face detection.

The authors have proposed an algorithm which is of three steps to solve this problem. In the first step, a subset is constructed by extracting image from an input image sequence every four images. In the second step, for each image in such subset, an energy value t is computed. The frequency dynamics descriptor (FDD) that is the standard deviation of the resulting flag value, is calculated for the representation of temporal changes of the face. Compared to the other works, which look for 3-D depth information of the head, the proposed algorithm has many advantages such as it is easy to compute.

**II. Local Literatures**

According to **Prospero C. Naval, Jr, “Recognizing Faces using Kernel Eigenfaces and Support Vector Machines” of 2003,** Principal Component Analysis (PCA) is used for extracting relevant features from high-dimensional data sets. It performs an orthogonal transformation of the coordinate system in which the data is originally described. After coordinate transformation, it is often the case that only a subset of the new coordinate values is necessary to describe most of the data. This subset is called the principal components of the data. The principal components possess large variance.

According to **Jerome Paul N. Cruz, et al., “Object recognition and detection by shape and color pattern recognition utilizing Artificial Neural Networks” of 2013,** the value of the weight and bias varies in every neuron. The process of determining the value of weight and bias is called learning or training. The algorithm used for learning is called back propagation algorithm. In this learning method, a desired output or target is given with a corresponding set of inputs. In the architecture of the artificial neural networks, back propagation algorithm requires 52 input elements and five output or target elements per set.

According to **Ma. Christina D. Fernandez, et al., “Simultaneous Face Detection and Recognition using Viola-Jones Algorithm and Artificial Neural Networks for Identity Verification” of 2014,** there are 7 facial features to be extracted and these are the skin color, color of the eye, the distance between the two eyes, the width of the nose, the height and width of the lips, and the distance between the nose and the lips. These are then detected, extracted, and measured from the person’s processed face image. These measurements are then passed through processes which will produce a representation of these characteristics in numerical vector form.  According to **Young Kyung Lee, Eun Ryung Lee. and Byeong U. Park's, “Principal Component Analysis in Very High-Dimensional Spaces” in 2011,** A particular disadvantage of PCA is that the principalcomponents are typically linear combinations of all variables Xj , which makes the results difficult to interpret, especially when d is very large. Recent years have seen several proposals that give ‘sparse’ solutions, that is, solutions that 934 YOUNG KYUNG LEE, EUN RYUNG LEE AND BYEONG U. PARK involve only a few nonzero loadings; see Jolliffe, Trendafilov, and Uddin (2003),Zou, Hastie, and Tibshirani (2006), d’Aspremont et al. (2007), d’Aspremont,Bach, and Ghaoui (2008), Shen and Huang (2008), Leng and Wang (2009), andWitten, Tibshirani, and Hastie (2009).  
 We are concerned with the case where d, the dimension of X, is comparable to, or even larger than, the sample size n. The standard PCA is known to yield inconsistent results in such a high-dimensional case, seeJohnstone and Lu(2009). We propose a method that gives consistent estimators of the principal component loading vectors.