2023

Improved Human Face Recognition by Introducing a New Cnn Arrangement and Hierarchical Method

Soroosh Parsai
windsor

Improved Human Face Recognition by Introducing a New CNN Arrangement and Hierarchical Method

By

Soroosh Parsai

A Dissertation
Submitted to the Faculty of Graduate Studies
through the Department of Electrical and Computer Engineering
in Partial Fulfillment of the Requirements for
the Degree of Doctor of Philosophy
at the University of Windsor

Windsor, Ontario, Canada

Improved Human Face Recognition by Introducing a New CNN Arrangement and Hierarchical Method

By

Soroosh Parsai

APPROVED BY:

____

M.M. Jamali, External Examiner

The University of Texas Permian Basin

____

R. Riahi

Department of Mechanical, Automotive and Materials Engineering

____

E. Abdel-Raheem

Department of Electrical and Computer Engineering

____

S. Alirezaee

Department of Electrical and Computer Engineering

____

M. Ahmadi, Advisor

Department of Electrical and Computer Engineering

September 12, 2022

Declaration of Co-Authorship and Previous Publications

I. Co-Authorship

I hereby declare that this dissertation incorporates material that is result of joint research with Dr. Majid Ahmadi and under his supervision.

- Chapter 2 to 5 of the thesis include the outcome of publications which have the following other co-author: Dr. Majid Ahmadi. In all cases only my primary contributions towards these publications are included in this thesis, and the contribution of Dr Majid Ahmadi was primarily through review, editing, and supervision.

I am aware of the University of Windsor Senate Policy on Authorship, and I certify that I have properly acknowledged the contribution of other researchers to my thesis and have obtained written permission from each of the co-authors to include the above materials in my research.

I certify that, with the above qualification, this dissertation, and the research to which it refers, is the product of my own work.

II. Previous Publication

This dissertation includes four original papers and one modification on a technique from a paper that have been previously submitted/ accepted for publication in peer reviewed conferences, and two journals as follows:

Publication title and full citation	Publication Status
Employing New Automatic Contrast-Limited Adaptive Histogram Equalization with Adaptive Average Dual Gamma Correction in A Face Recognition System	Submitted to Pattern Recognition Letters
Flow Pattern Recognition using Spectrogram of Flow Generated Sound with New Adaptive LBP Features	Presented at ICICT and Published in Springer
New Local Binary Pattern Feature Extractor with Adaptive Threshold for Face Recognition Applications	Published in International Journal of Artificial Intelligence & Applications
A Low Error Face Recognition System Based on A New Arrangement of Convolutional Neural Network and Data Augmentation	Accepted for presentation, TENCON 2022
A Novel Hierarchical Face Recognition Method Based on the Geometrical Face Features and Convolutional Neural Network with a New Layer Arrangement	Presented at ISBM 2022, Springer

I certify that I have obtained a written permission from the copyright owner(s) to include the above published material(s) in my dissertation. I certify that the above material describes work completed during my registration as graduate student at the University of Windsor.

III. General

I declare that, to the best of my knowledge, my thesis does not infringe upon anyone's copyright nor violate any proprietary rights and that any ideas, techniques, quotations, or any other material from the work of other people included in my thesis, published or otherwise, are fully acknowledged in accordance with the standard referencing practices. Furthermore, to the extent that I have included copyrighted material that surpasses the bounds of fair dealing within the meaning of the Canada Copyright Act, I certify that I have obtained a written permission from the copyright owner(s) to include such material(s) in my thesis.

I declare that this is a true copy of my thesis, including any final revisions, as approved by my thesis committee and the Graduate Studies office, and that this thesis has not been submitted for a higher degree to any other University or Institution.

Abstract

Human face recognition has become one of the most attractive topics in the fields of biometrics due to its wide applications. The face is a part of the body that carries the most information regarding identification in human interactions. Features such as the composition of facial components, skin tone, face's central axis, distances between eyes, and many more, alongside the other biometrics, are used unconsciously by the brain to distinguish a person. Indeed, analyzing the facial features could be the first method humans use to identify a person in their lives.

As one of the main biometric measures, human face recognition has been utilized in various commercial applications over the past two decades. From banking to smart advertisement and from border security to mobile applications. These are a few examples that show us how far these methods have come. We can confidently say that the techniques for face recognition have reached an acceptable level of accuracy to be implemented in some real-life applications. However, there are other applications that could benefit from improvement. Given the increasing demand for the topic and the fact that nowadays, we

have almost all the infrastructure that we might need for our application, make face recognition an appealing topic.

When we are evaluating the quality of a face recognition method, there are some benchmarks that we should consider: accuracy, speed, and complexity are the main parameters. Of course, we can measure other aspects of the algorithm, such as size, precision, cost, etc. But eventually, every one of those parameters will contribute to improving one or some of these three concepts of the method. Then again, although we can see a significant level of accuracy in existing algorithms, there is still much room for improvement in speed and complexity. In addition, the accuracy of the mentioned methods highly depends on the properties of the face images. In other words, uncontrolled situations and variables like head pose, occlusion, lighting, image noise, etc., can affect the results dramatically.

Human face recognition systems are used in either identification or verification. In verification, the system's main goal is to check if an input belongs to a pre-determined tag or a person's ID.

Almost every face recognition system consists of four major steps. These steps are pre-processing, face detection, feature extraction, and classification. Improvement in each of these steps will lead to the overall enhancement of the system. In this work, the main objective is to propose new, improved and enhanced methods in each of those mentioned steps, evaluate the results by comparing them with other existing techniques and investigate the outcome of the proposed system.

To my wife for her sacrifices and to daughter for bringing a light to my life, which made this possible.

To my parents for their endless supports.

Acknowledgement

First, and foremost, I would like to give my sincere thanks to my advisor Dr. Majid Ahmadi for his remarkable assist, wonderful guidance, and great support throughout this research and for believing in me. I will always be thankful to him for that his energy and enthusiasm have been my motivation through this path. As well, I would like to appreciate my committee, Dr. Esam Abdel-Raheem, Dr. Shapour Alirezaee, and Dr. Reza Riahi for their constructive and helpful comments and feedback that have had a determining role in the quality of this work.

Table of Contents

List of Tables

List of Figures

List of Abbreviation

PCA	Principal Component Analysis
SQI	Self-Quotient image
SVM	Support Vector Machine
SVD	Singular Value Decomposition
RF	Random Forest
LDA	Linear Discriminative Analysis
HOG	Histogram of Gradients
LBP	Local Binary Pattern
HE	Histogram Equalization
DoG	Difference of Gaussian
GIC	Gamma Intensity Correction

ANN	Artificial Neural Networks
CNN	Convolutional Neural Network
AAM	Active Appearance graph Models
CLAHE	Contrast-Limited Adaptive Histogram Equalization

1 *Introduction to Face Recognition*

1.1. Introduction

Due to its nature, human face recognition is the most familiar biometric method among humans. We can confidently say that it is the only method we use daily from the early ages of our lives in our communications with others. When encountering another person, our mind searches through various parameters, such as age, gender, skin color, height, hair type, etc., to recognize them. In the meantime, that person's facial features play the most important role in this task. There are many cases in which we confuse a person, despite the difference in other characteristics, only because of the similarity in the face with another person.

Many commercial and security applications can be imagined for face recognition in both verification and identification. In fact, these days, we consider face recognition systems practical and use them in many applications. One of the contributing facts is the latest improvement in the required infrastructure for those systems, such as video surveillance, computer networks, and databases. The other factor is the increasing demand for security in the digital world around us. All this together makes human face recognition an interesting subject in the field of image processing and motivates researchers to further improve the system. One of the highlights of face recognition is the diversity of the field of researchers attracted to the topic. Examples include computer vision [1], image processing, neural networks, and pattern recognition to neurology [2], humanology, and network security [3].

1.2. Face Recognition

These days, it is hard to find a person that has not been photographed for biometrics in visa application or for issuing a passport. Almost every cell phone is equipped with one of the biometric facial methods. These are a few examples of the practical application of face recognition. These methods are used to either verify a person or identify a target from a photo or a few frame of a video. One distinguishing feature of face recognition is its non-contact property from other biometrics that require direct participation of the subject in most cases. Nearly every face recognition algorithm can be divided into four main stages: pre-processing, face detection, feature extraction, and classification. These four stages are usually done in the same order, with few exceptions. In the pre-processing stage, the algorithm deals with the noise. Also, in some instances pre-processing is used to correct the illumination of the image and mostly normalize the input image for the algorithm requirements. Hence, it seems logical to use pre-processing in the preliminary stages. As the next step, face detection helps reduce the amount of computation by eliminating the useless parts of the face image. In creating the feature vector, feature extraction is a crucial step. Features presented in the feature vector must be distinctive enough to enable the recognition of each person for the algorithm. Finding new features, using different features together, and choosing the most distinctive and resilient to change features are some of the actions that can improve the feature extraction stage. Finally, in the last stage, the algorithm implements a classifier to classify the instances of the dataset based on their features and then later uses this to assign an unknown input image to a class.

1.3. Identification vs. Verification

Biometric applications fall into two categories: identification and verification. The application plays a decisive role in selecting the proper face recognition method. In verification, the main objective is whether to approve or deny an anticipated identity of the input against an assertion, whereas, in identification, the algorithm tries to determine the identity of an unknown instance by searching through a database and assigning the face image to a pre-defined group. Figure 1.1 is the comparison of the identification and verification.

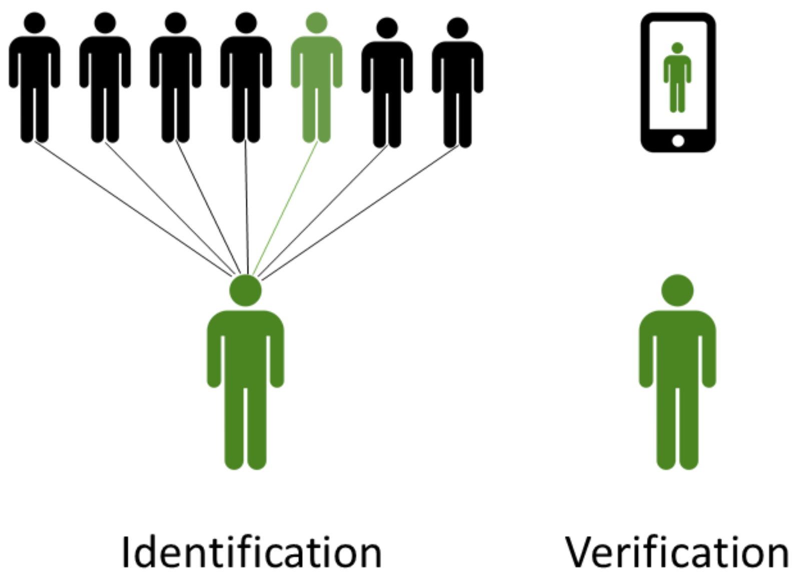

Figure 1.1 Identification vs Verification

1.4. Application

Recent improvements in technology have not left face recognition untouched. These advances make face recognition suitable for many applications, in many areas from commercial to industrial and from security to the government. Some of these applications can be categorized as follows.

- **Security & Surveillance: Airport, ATM Machines, Border Cross Point, Network Security.**
- **Indexing of Videos: Ticket Booking, Criminal Justice System.**
- **Investigation Image Database: Missing Children, Face Reconstruction, Driving License.**
- **Verification for Identity: Banking Field, Electronic Commerce.**

1.5. Challenges

Initially, face recognition algorithms were extremely limited, with comparatively low accuracy and high demanding computational complexity. Advances through the last three decades have gotten the methods to a level that the existing algorithms show an acceptable accuracy percentage. However, this amount of accuracy is almost only achievable under controlled situations, and whenever we are dealing with unpredicted parameters such as head pose, shadows, facial expression, etc., the performance of the system is affected by them. The most influential phenomena that can cause those undesired variations are as follows.

1.5.1. Illumination

The human eye perceives the image of the objects by collecting and analyzing the reflected lights from that object. Images are mainly created the same way. Hence during the process of an image, the lighting of the object plays a determinative role. It has been shown that changes of just illumination in the face images can dramatically affect the face recognition algorithm's results. This is somehow anticipated due to the 3D shape of a human face image. Changing the lighting condition can create shadows on the face, which can eventually weaken or even mask some of the face features. Accordingly, as we discussed before, losing the features reduces the robustness of the system. Variations of illumination is shown in figure 1.2.

Figure 1.2 Different illuminations in face images

1.5.2. Head Pose

Location of the face in a face image, i.e., positioning of the face features relative to each other, is another aspect of the face recognition algorithm. Since an image is a 2D representation of a 3D object, i.e., a face, then obviously, variations in the direction of the face can change the geometry of that representation. Hence, these variations could change the features of the image and decrease the accuracy of the algorithm. In some cases, this phenomenon can be prevented by providing a template in the acquisition stage (e.g., access granting systems or passport validation stations), but we barely have access to a standard form of a face image in most other cases (e.g., surveillance systems.) Few variations of head poses can be seen in figure 1.3.

Figure 1.3 Different head poses of a person

1.5.3. Facial Expression

Facial expressions, same as the head pose and illumination variations, can affect the features of a face. Especially considering that in most cases, some level of normalization is applied on training data sets, which demands the image to be from a natural state of the face. One of the methods that initially seems to be effective against these types of distortion is to choose the most robust features of the face. Figure 1.4 demonstrates few examples of these expressions.

Figure 1.4 Facial Expressions

1.5.4. Occlusion

Any blockage on a face image could result not only in losing some of the facial features but also in replacing the features with some unrelated and unwanted ones (e.g., when the obstacle is a part of another person's face.) Face recognition systems experienced difficulties regarding this challenge, especially in the last couple of years, due to the mask mandate forced by the pandemic. A common technique adopted by many methods is to identify the occluded part of the face and subtract that part from the image. This helps the algorithm to deal with the features that are not affected by the mask.

Figure 1.5 Examples of occlusion in face images

1.5.5. Inter-Class Similarities

Even humans can experience difficulties differentiating between two persons with facial resemblance (e.g., identical twins, actors, and actresses). Given that humans employ many more parameters in addition to facial features to identify a person, it should not be unexpected that machines would have more difficulties dealing with inter-class similar individuals. Simply put, distinguishing between two separate members with somehow similar features is the problem.

1.6. Face Recognition Methods

As discussed before, a face recognition algorithm usually consists of 4 stages. Figure 1.6 shows the flow diagram of a face recognition system. Pre-processing is the step in which we attempt to normalize the image with. The next step would be face recognition, in which we try to outline the parts of the image that contain face features information. This part would help towards minimizing the computations effectively. The core process of a recognition algorithm could be the feature extraction step. In this stage, a desirable algorithm searches for the most distinguishable features of a face image. These features are pre-defined in the algorithm. In the end, an unknown instance is fed to the algorithm to be assigned to a class of features by the algorithm's classifier. Later on, we will dive into the details of each of these steps.

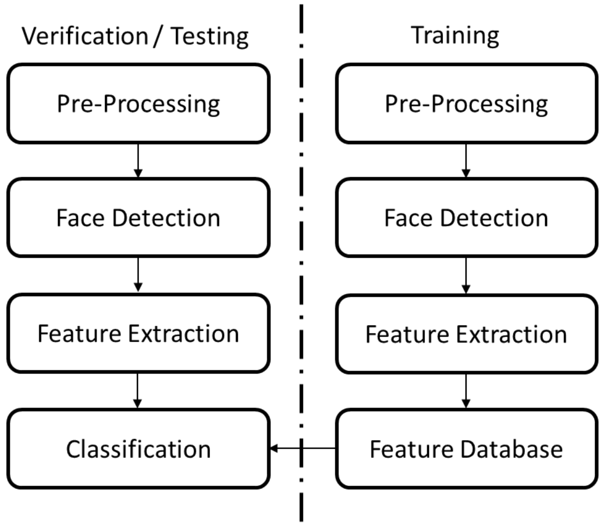

Figure 1.6 Flow diagram of the face recognition process

1.6.1. Pre-Processing

Other than the image acquisition step, the first stage of most face recognition algorithms is pre-processing. In fact, in most image processing applications, this step is considered to be the first one. The face images could contain all or some of the distortions we mentioned before, such as noise, uneven exposure, off-center face position, non-matching coloring method, etc. This highlights the importance of pre-processing as the first stage of the algorithm more and more since we most likely want to deal with these alterations before doing any training or classification processes. One of the most affecting distortions is poor illumination. Uneven lighting and shadows on the face image usually are the first and most common problems which a face recognition system should deal with. These shadows and dark parts of the image lead to the loss of significant parts of the information, reducing the technique's accuracy. Histogram equalization [4] and Gamma intensity correction [5] are two of the most common methods of eliminating lighting problems.

1.6.2. Histogram Equalization

Histogram equalization is one of the pre-processing techniques that does not depend on the information of the distortion of the image. The HE method is a global technique that acts as a global function to improve the unity of the image's contrast based on the intensity distribution of the pixels gathered from the image's histogram. The main reason for its popularity comes from the fact that it is one of the easiest to implement techniques. However, the poor results of the technique in instances with extreme illumination variation [6], its tendency to eliminate the detail in well-lit areas as well as adding unwanted artifacts and noises make it a less effective method.

The main target of the HE is to smooths the histogram of the image by utilizing the cumulative distribution function, which increases the dynamic range of the image. This flattening happens globally throughout all of the parts of the image since HE is a global function.

We define the histogram of an input image of $I(x, y)$ with Q levels of gray based on the probability function of the gray level i as shown by equation (1.1): [6]

$$p(i) = m_i/M \tag{1.1}$$

Where $i \in \{0, 1 \ldots Q - 1\}$ is the grey level, and M is the total number of pixels in the image.

1.6.3. Contrast-Limited Adaptive Histogram Equalization [7]

CLAHE is an improved form of histogram equalization. Compared to the HE global function, CLAHE acts on the image blocks, and the action is particularly limited by the clipping point. That block division and clipping points help the CLAHE to reduce the effect of noise in the image.

CLAHE tends to create a more natural output image. However, the technique is not strong due to its modest contrast enhancement and weakness in creating considerable grayscale differences when needed. In addition, CLAHE faces difficulties whenever it deals with the gamma distortion in the input image.

1.6.4. Gamma Intensity Correction (GIC) [8]

Gamma distortion is a phenomenon that affects the illumination of the images and comes from some limitations and characteristics of the photography hardware. GIC is one of the techniques that we use to handle this type of distortion. This method simultaneously improves the dynamic span of the dark regions and compresses the lighter area of the image. Consequently, GIC changes the image's lighting while maintaining the overall dynamic range. The limitations that we face in this technique are the difficulty of calculating the gamma correction value and shadows, which are not affected by the GIC.

The non-linear equation of the GIC is shown in (1.2): [9]

$$I = I^{\frac{1}{\gamma}} \tag{1.2}$$

Figure 1.7 shows the curve of intensity correction based on the gamma value. As we can see, the method lightens the input for values greater than one, and less than one makes the image darker.

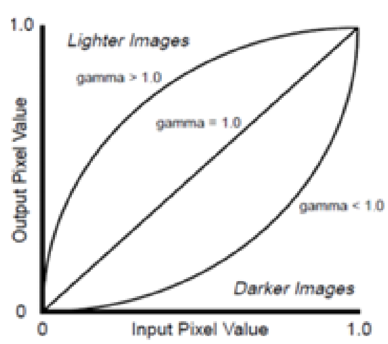

Figure 1.7 Gamma Intensity Correction

1.6.5. Difference of Gaussian (DoG) [10]

The DoG is mainly successful in dealing with the shadows in the image. We can consider the DoG to generally be a filter on the input image to better understand this method. This filter is designed to eliminate the high-frequency parts of the image, which are considered to be noise, as well as low-frequency elements which construct the homogenous area of the input. The output of this specific filter can be imagined as the edge of the image.

A convolution with two distinct Gaussian kernels happens in the DoG, and the contrast between the two convolved pictures is used to create the output image.

The downsides of this technique are that, firstly, since the DoG deals with the image as a whole, a further step is required for the contrast correction after applying the filter. Secondly, before applying the filter, a gamma normalization step is necessary to prevent impacting the contrast in the shadow areas.

If we want to give another definition for the DoG method, we can say it is subtracting two blurred versions of the original from each other. Those two blurred images are produced by a convolution between the gaussian kernels and the input, considering that the kernels have different standard deviations. The equation (1.3) shows this operation: [11]

$$DOG(x,y) = \frac{1}{2\pi{\sigma_1}^2} e^{-\frac{x^2+y^2}{2{\sigma_1}^2}} - \frac{1}{2\pi{\sigma_2}^2} e^{-\frac{x^2+y^2}{2{\sigma_2}^2}} \tag{1.3}$$

σ_1, σ_2 are widths of the gaussian kernel.

1.6.6. Locally Tuned Inverse Sine Nonlinear (LTISN) [12]

The LTISN is an effective solution, especially in the instances of serious illumination problems. The strength of the method is evident in the contrast enhancement as well as the limitation of the dynamic range. The non-linear intensity correction and different operators in the method help in that regard. The LTISN acts on the pixels of the image by improving the pixel's brightness in the dark areas and reducing the illumination value of the pixels in the bright areas. Also, a function provides the technique with the ability to enhance the local contrast.

However, the weakness of this method is the over and under enhancement in the too dark and too bright regions, which can lead to the development of unwanted artifacts.

The following equations (1.4 to 1.8) show the pixel-wise operation of the inverse sine function and the variable parameters (based on the neighboring pixels) that the method uses to correct the illumination of the pixels:

$$I_{enh}(x,y) = \frac{2}{\pi} sin^{-1}(I_n(x,y)^{q/2} \tag{1.4}$$

$$kernel_i(n_1,n_2) = \frac{h_g(n_1,n_2)}{\sum_{n_1}\sum_{n_2} h_g(n_1,n_2)} \tag{1.5}$$

$$h_g(n_1,n_2) = e^{\frac{-(n1^2+n2^2)}{2\sigma^2}} \tag{1.6}$$

$$I_{M,i}(x,y) = \sum_{m=-\frac{M_i}{2}}^{\frac{M_i}{2}} \sum_{n=-\frac{N_i}{2}}^{\frac{N_i}{2}} I(m,n) kernel_i(m+x,n+y) \tag{1.7}$$

$$q = \begin{cases} tan(\frac{\pi}{C_1} I_{M_n}(x,y)) + C_2 & I_{M_n}(x,y) \geq 0.3 \\ \frac{1}{C_3} ln(\frac{1}{0.3} I_{M_n}(x,y)) + C_4 & I_{M_n}(x,y) < 0.3 \end{cases} \tag{1.8}$$

The image intensity is normalized to [0 1].

1.7. Face Detection

Face detection is mainly locating the position of a face in an input image. The algorithm locates a single face according to its scale and orientation [13]. Extracting faces from video sources created a target in image processing to increase detection speed to match the frame rate. Papageorgiou *et al.* [14] proposed a technique for the detection of objects based on wavelet representation and statistical learning methods. Osuna *et al.* employ Vapnik's support vector machine technique for face detection in [15], and Romdhani *et al.* enhanced it by introducing reduced training vector sets for their classifier [16]. Schneiderman and Kanade's used a statistical method for profile detection [17], but their method only acted upon three face orientations, and each orientation is considered a different object. Rowley and Kanade utilized neural network-based filters [18], while Fleuret and Geman attempted a coarse-to-fine approach for face detection [19]. Their main focus was to minimize computation. Viola and Jones [20] used the concept of an integral image, along with a rectangular feature representation and a boosting algorithm, as their learning method. They obtained a good result in detecting faces at 15 frames per second. Their method minimized computation time while achieving a high detection rate.

1.8. Feature Extraction

We already mentioned many possible face recognition applications in real-life situations, from surveillance to security and human-computer communications to data gathering. One of the contributing factors to the effectiveness of these tools is the quality of the input, or better to say, the resilience of the algorithm to the poor quality of the input. A crucial component of the design is deciding on the best and most efficient feature extraction approach. These techniques could be classified into two types. The first form is to extract the geometrical features of the face. This method directly deals with the structure and proportions of the face. The information, such as the size and locations of the eyes, mouth, lips, and nose, as well as the ratio of the distances of face components, are used in this method. The second technique, or the holistic method, consists of two major approaches [21]. Principal component analysis (PCA) [22], linear discriminant analysis (LDA) [23], and singular value decomposition (SVD) [24] are examples of appearance-based subspace methods that focus on dimension reduction. A frequency-domain transformation, such as discrete cosine transform (DCT) [25] or Gabor wavelet [26], is used in the second category of holistic feature extraction approaches to capture the image's distinctive features. In these attempts, some of the frequency domain coefficients are used to represent picture characteristics. The high dimensionality of the spatial frequency features affects the Gabor wavelet negatively. The dimensionality reduction subspace approaches can be used to solve this.

1.8.1. PCA

PCA is a linear holistic-based approach that converts image data to a lower-dimensional subspace termed eigenspace when applied to a complete face image of size $R \times C$. By reducing correlated variables into a smaller number of uncorrelated variables, PCA tends to lower the dimensionality of the feature space. The term "principal components" refers to these unrelated variables. This method implies that the majority of class information is concentrated in the directions where the variations or scatterings are higher. For a training dataset of N images, the first step is to generate the covariance matrix C by finding the mean of the photos and subtracting this mean, μ, from all the images.:

$$C = \frac{1}{N}\sum_{1}^{N}(x_i - \mu)(x_i - \mu)^T \qquad (1.9)$$

For M as the largest eigenvalues of C, we have:

$$CT_i = \lambda_i T_i \qquad i = 1, \ldots, M \qquad (1.10)$$

where, T_i's are the eigenvectors or eigenfaces. The covariance matrix's eigenvalues are sorted in the proper sequence. To reduce the subspace dimension of the test and train data, eigenvectors associated with a few numbers of the biggest eigenvalues are utilized. [27].

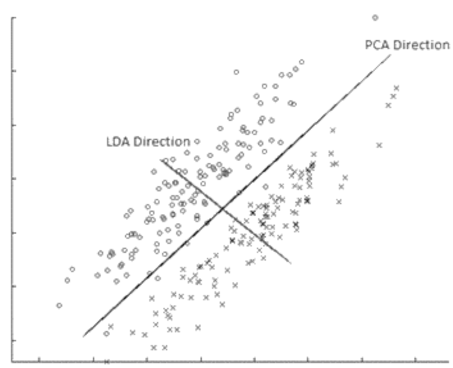

Figure 1.8 PCA maps the samples to the direction with the maximum variation, where LDA finds the direction with maximum between-class variation to within-class variation.

1.8.2. LDA and FDA

In PCA, the main target is the directions related to large eigenvalues, i.e., large variation. Some critical information may be conveyed in directions with minor variations in some cases. The goal of linear discriminant analysis (LDA) is to optimize this objective as shown Figure 1.8. Both classes generated by PCA are distributed in the same direction, as seen in

Figure 2.2. The ratio of between-class scattering tends to be maximized in LDA. As a result, the outcome is better.

Fisher's linear discriminant technique linearly combines PCA and LDA. LDA reduces dimensionality and maximizes the matrix of the ratio of between-class scatter to within-class variation by lowering dimensionality. The following is the formula for calculating the ratio matrix: [22].

$$M_b = \sum_{k=1}^{In_s} In_{sk}.(Avg_k - Avg)(Avg_k - Avg)^T, M_w = \sum_{k=1}^{N_c} \sum_{I_i \epsilon L_i} (I_i - Avg_k).(I_i - M_k)^T, \quad (1.11)$$

$$V_{opt} = argmax\left(\frac{V^T.S_b.V}{V^T.S_w.V}\right), \quad (1.12)$$

Where M_w and M_b are the within-class and between-class scatter matrices, respectively. In_s is the number of individuals, In_{sk} is the number of samples in the k^{th} class and L_i is the i^{th} class. Avg and Avg_k are the mean images of the training set and the k^{th} class, respectively.

$$S_b.V = \lambda.S_w.V \quad (1.13)$$

The 2.5 defines the eigenvector V. [19]

We will have the rank of V as $N_i - N_c$. Compared to the count of the pixels in the input $(P \times Q)$, the rank is considerably smaller, so by applying the PCA, the dimension is reduced, and then the LDA projects the data to the new $N_c - 1$ dimension. [23]

1.8.3. SVD

The Singular Value Decomposition (SVD) has long been used in signal processing and statistical data analysis as a dimensionality reduction method. A data matrix's singular values convey information about the level of noise, energy, matrix rank, and so on. Considering the fact that singular vectors of the matrix are the bases of the matrix span and the orthonormal nature of those vectors, the majority of the signal's features can be found in them. We can use the SVD method to obtain the numerical characteristics of an image.

$$X = USV^T \tag{1.14}$$

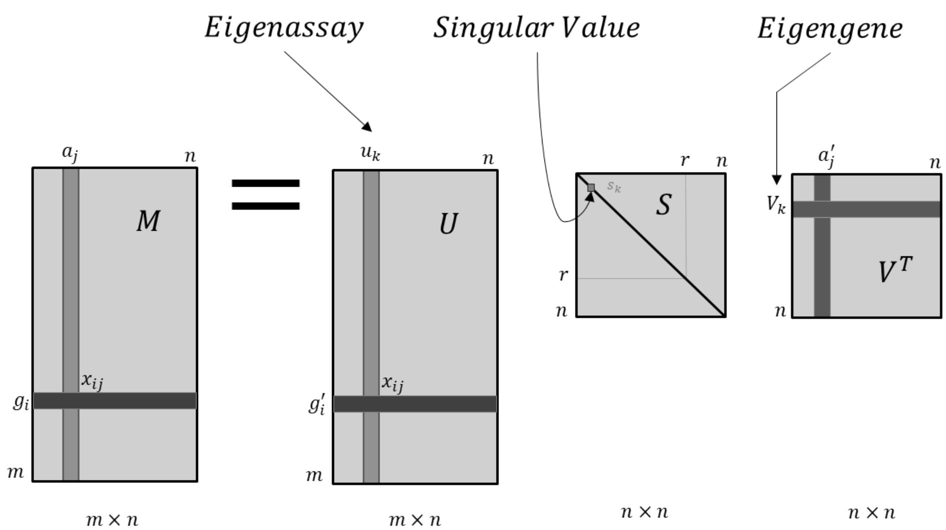

Figure 1.9 Singular Value Decomposition

S is an $n \times n$ diagonal matrix of singular values with components $s_{ij} = 0, i \neq j$ and $s_{ij} > 0$, and $U(m \times n)$ and $V(n \times n)$ are orthogonal matrices. It may also be demonstrated that the non-exclusive matrices of U and V exist such that $s_1 \geq s_2 \geq \ldots \geq 0$. The left and right singular vectors are the columns of the orthogonal matrices U and V, respectively. The fact that U and V are mutually orthogonal is a crucial feature [21]. The primary characteristic of the approach relevant to face image processing is the stability of the SVD on the face image. The stability of the SVD in image processing is the primary characteristic of the method that makes it attractive [24]. Because of these characteristics, many researchers find the SVD a reliable feature extraction tool [27].

1.8.4. Gabor

Gabor's two-dimensional wavelet is a frequency localization filter that functions similarly to the human visual cortex. The Gaborder 2D filter converts image frequency components into time-domain information and can be estimated by multiplying a sine by a Gaussian function.

$$\psi_{s,o}(x,y) = \frac{f_s^2}{\pi\gamma\mu} e^{-[\left(\frac{fs^2\hat{x}^2}{\gamma^2}\right)+(\frac{fs^2\hat{x}^2}{\mu^2})]} . e^{-(j2\pi.f_s\hat{x})}, \quad \begin{cases} s = 0,1,2,\ldots S_{max} - 1 \\ o = 0,1,2,\ldots\ O_{max} - 1 \\ \hat{x} = \ +x\cos(\beta_o) + y\sin(\beta_o) \\ \hat{y} = \ -x\sin(\beta_o) + y\cos(\beta_o) \end{cases}, \quad (1.15)$$

$$\beta_o = (o\pi)/8 \quad (1.16)$$

$$f_s = \ f_{max}/(\sqrt{2})^s \quad (1.17)$$

Where f_{max} is the maximum frequency of the filter bank, and γ and μ are the sharpness of x and y axis, respectively, and equal to $\sqrt{2}$. S_{max} and O_{max} represent the maximum value of scale and orientation.

1.8.5. Discrete Cosine Transform (DCT)

Discrete Cosine Transform is another feature extraction method that is based on reducing the redundancy using a set of trigonometric functions, more precisely cosine functions, in the following arrangement:

$$D(p,q) = \alpha(p)\alpha(q)\sum_{m=0}^{R-1}\sum_{n=0}^{C-1} I(m,n)\cos[\frac{(2m+1)p\pi}{2R}]\cos[\frac{(2n+1)q\pi}{2C}] \quad (1.18)$$

where

$$\begin{cases}\alpha(p) = \frac{1}{\sqrt{R}} & if\ p = 0. \\ \alpha(p) = \sqrt{2/R} & if\ p = 1,2,\dots,R.\end{cases} \quad \begin{cases}\alpha(q) = \frac{1}{\sqrt{C}} & if\ q = 0. \\ \alpha(q) = \sqrt{2/C} & if\ q = 1,2,\dots,C.\end{cases}$$

The output of the DCT on an image with the size of $(P \times Q)$ results in a matrix of coefficients from the frequencies of the image, which preserves the original size $(P \times Q)$ of the input.

1.8.6. Local Binary Pattern

The LBP descriptor is another bit-wise feature extraction technique that extracts the micro-patters hidden in the relations between each pixel of the image and its neighbors. One of the most important stages of the algorithm is assigning a string of binary to each pixel which is called the LBP code. The descriptor utilizes the relations between a specific pixel with its surrounding neighbors. Each pixel is compared to a certain number of pixels around it, based on the setting of the algorithm. This setting also determines the direction of the next step in the process of assessment. In the course of the comparison, whenever the value of the differentiation turns out to be negative, a zero is added to the LBP code. Similarly, a positive value indicates a one in the LBP code. This encoding is basically the main idea of the LBP descriptor. The important factors that change the behavior of the descriptor and need to be tuned are the number of surrounding pixels and the encoder's direction of rotation.

For example, a typical encoder can be defined to consider eight pixels (a 3x3 window with the target pixel in the middle), a clockwise path, and the top left corner pixel as the first element.

The benefits of the LBP are 1) it is relatively easier to implement while showing effective results. 2) The constructed hidden micro-patterns are independent of the conversion.

However, the small size of the encoding window can be noted as one of the downfalls of the LBP.

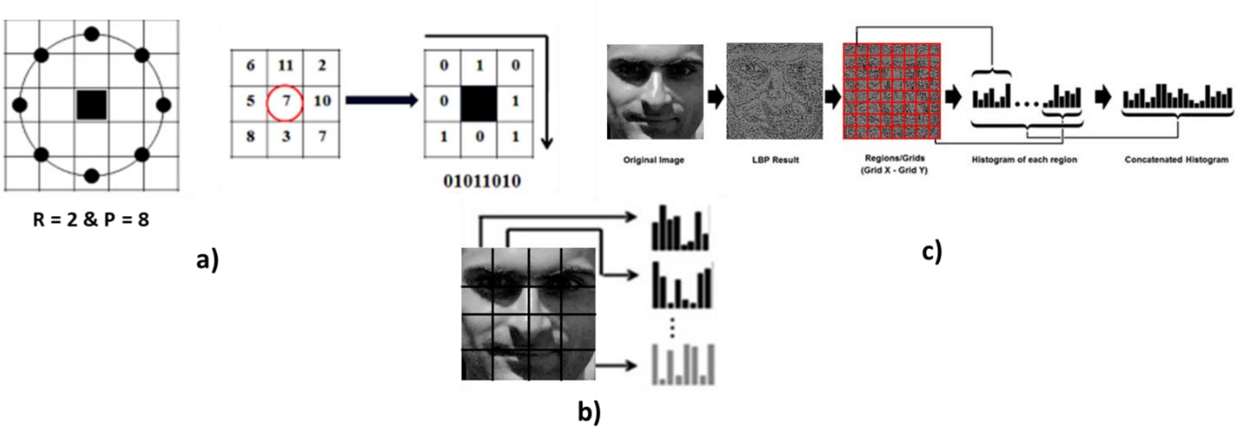

Figure 1.10 Different steps of LBP, a) Radios, number of neighbors, and direction, b) Comparison results, c) Histogram representation of the image

A pixel's intensity is denoted by i. The i_p determines the intensity of the p_{th} neighboring pixel, where $p = 1,2\ldots P$ and P is the number of neighbors. Here, R defines the distance between i_p and i . Considering the mentioned notations, $LBP_{p,R}$ encodes the pixel difference $z_p = i_p - i$ between i_p and i with R as the distance [28].

The following equation shows the encoding concept of the LPB: [28]

$$b_p = \begin{cases} 1 \ if \ D_p \geq 0 \\ 0 \ if \ D_p < 0 \end{cases} \tag{1.19}$$

1.8.7. Convolutional Neural Network

The idea of the convolutional neural network comes from the operation of the brain's visual cortex [29]. The CNNs share similar learning steps of the visual cortex. In the human visual system, each area of the vision is related to a particular group of brain cells. It has been proved that different visual features such as vertical and horizontal edges activate distinct groups of neurons related to that specific feature.

CNNs are a type of Neural Network that has been shown to be particularly successful in image recognition and classification. CNNs are feed-forward neural networks with a large number of layers. They are made up of filters, kernels, or neurons with trainable weights, parameters, and biases. Each filter takes a set of inputs, conducts convolution, and optionally adds non-linearity to the mix [30]. Backpropagation, in the training stage, supports the algorithm in learning by tuning the weights based on the input.

The Architecture. As depicted in Figure 1.11, a typical CNN architecture consists of different types of layers. These layers are convolutional, pooling ReLU, and fully connected.

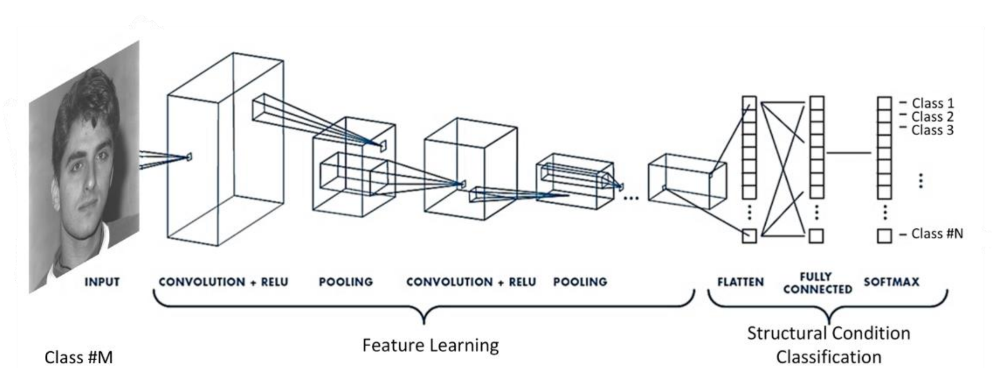

Figure 1.11 The architecture of a CCN

Convolutional Layer. The basic building component of a Convolutional Network, the Convolutional Layer, handles the majority of the computational hard lifting. The basic goal of the Convolution layer is to extract features from the picture input data. By learning picture attributes using tiny squares of the input image, convolution retains the spatial connection between pixels. A collection of learnable neurons is used to obfuscate the input picture. In the output picture, this creates a feature map or activation map. The convolutional layer then receives it as the input on the next level.

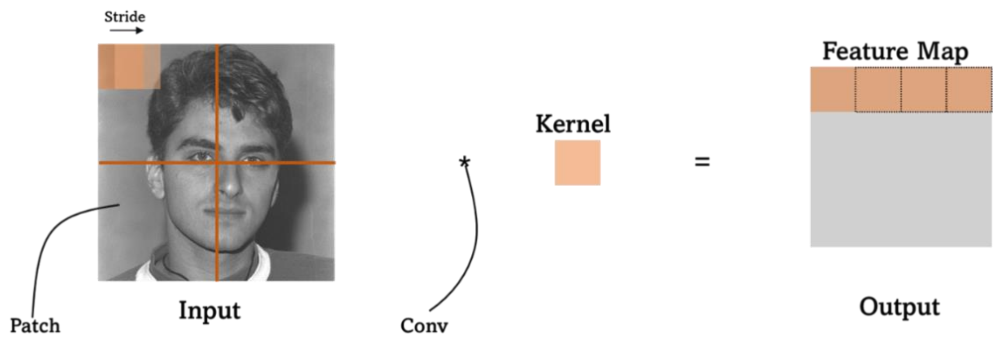

Figure 1.12 The process of convolution

The following equation is the discreet form of the convolution:

$$[W \circledast X](i,j) = \sum_{u=0}^{H-1} \sum_{v=0}^{W-1} \omega_{u,v} x_{i+u,j+v} \tag{1.20}$$

The main structure of the convolutional layer consists of neurons connecting to a limited number of neurons in the next layer. The target neurons form a square region, which its size can is controlled by the Receptive Field parameter. However, the depth of the network is not controlled, and the convolution operates on the entire depth of the network.

A crucial step for the network to distinguish the hidden features of each layer is employing filters. The sets of neurons or the filters able the system to locate unique features in various image regions.

Pooling Layer. The pooling layer reduces each activation map's dimensionality, but the most crucial information is retained. The input images are separated into a collection of rectangles that do not overlap. A non-linear technique such as average or maximum is used to down sample each region. This layer, commonly positioned between convolutional layers, delivers improved generalization and faster convergence and is resistant to translation and distortion.

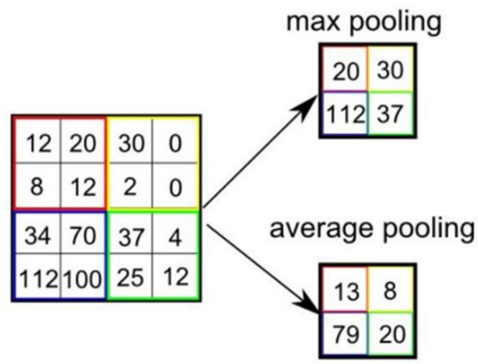

Figure 1.13 Two different types of down sampling

Fully Connected Layer. Every filter in the previous layer is connected to every filter in the following layer, which is referred to as a fully connected layer (FCL). The convolutional, pooling, and ReLU layers' outputs are representations of the input's high-level features.

ReLU is a non-linear operation that comprises rectifier-based units. It's an element-by-element operation, which means it's applied per pixel and replaces all negative values in the feature map with zero. To understand how the ReLU works, we suppose that the neuron input is x, and then the rectifier is specified for neural networks as $f(x) = max(0, x)$.

The simplicity and linear behavior of the ReLU make it more desirable than the other activation functions, such as Sigmoid and Tanh.

Figure 1.14 depicts the ReLU function as well as an example of the function.

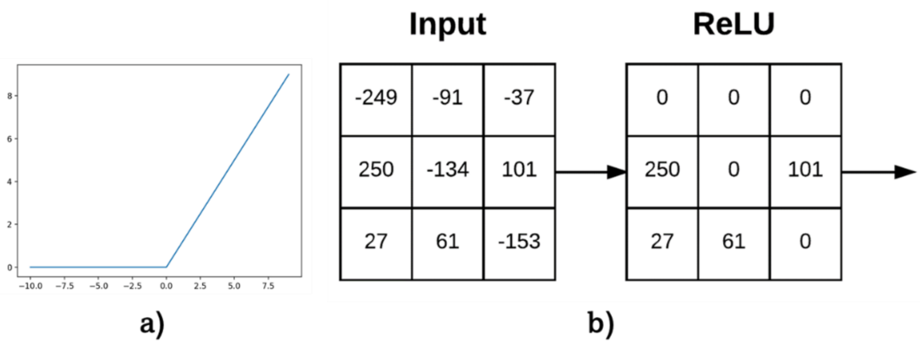

Figure 1.14 ReLU Function

1.9. Classification

After acquiring suitable vectors for all the training samples, a classifier is required to create classes based on the features. Among all the existing classifiers, the followings are some of the most popular ones.

1.9.1. Neural Networks

Artificial neural networks (ANN) have been widely utilized as the classifier in many pattern recognition applications and are extremely frequent in face recognition systems, as the structure of biological neural networks inspires them. A set of artificial neurons is joined together to build a network of neurons that mimics biological brain networks in an artificial neural network. Researchers have developed a variety of neural network topologies, including multilayer perceptron, backpropagation, and radial basis function (RBF) neural networks, among others. Dynamic-link architecture (DLA) is a neural network-based approach that uses information collected from face photos to dynamically arrange neurons into higher orders, resulting in more robust face identification despite environmental fluctuations [31].

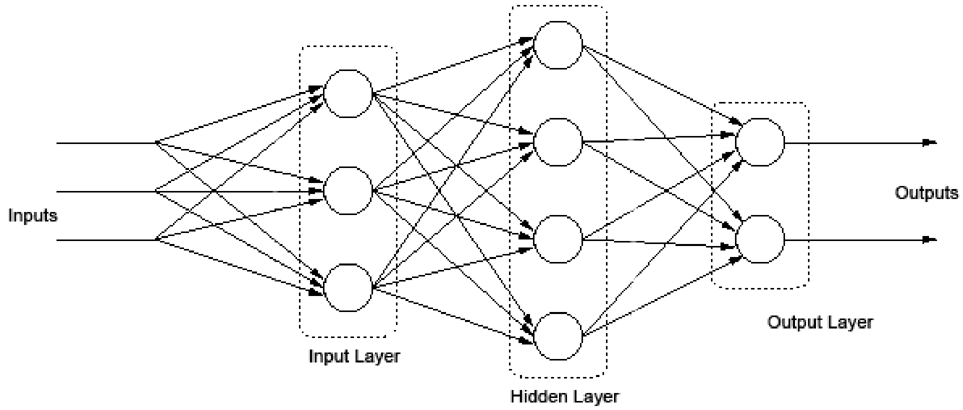

Figure 1.15 Three-layer perceptron

1.9.2. Support Vector Machines

Support vector machines (SVM) are characterized as maximum margin classifiers since they reduce classification error while also increasing geometric margin [32]. An SVM creates a separating hyperplane in the feature space that maximizes the margin between the data sets. To determine the margin, two parallel hyperplanes are built, one on either side of the separating one. These hyperplanes are pushed up against the two data sets, allowing the hyperplane with the farthest distance to the adjoining support vectors of both classes to achieve a decent separation as depicted in figure 1.16. The greater the margin or distance between these parallel hyperplanes, the more likely unknown samples will be accurately identified [33].

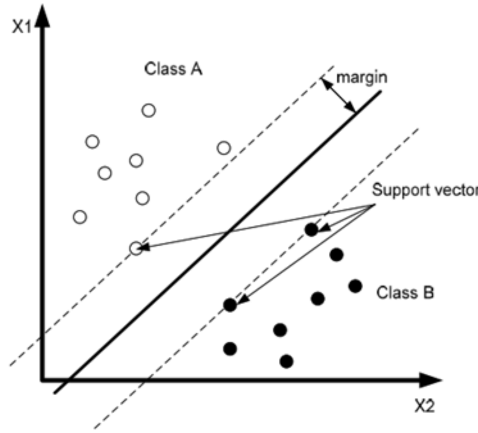

Figure 1.16 Maximum-margin hyperplane for an SVM classifier. Samples on the margin are called the support vectors.

1.9.3. Hidden Markov Model

As a double stochastic process, the hidden Markov model can effectively represent the production of sequential data. Speech, facial, and handwriting recognition have all been effectively implemented using HMM. When it comes to employing HMM as a classifier, there are two main techniques. In the first method, all classes are modelled by a single HMM, which are discriminated against based on distinct paths. A distinct HMM is employed for each class in the second method, known as the model discriminant approach. Nefian's [34] 2D embedded HMM has a collection of super-states, each of which is linked to a set of embedded states. Super states represent the primary face areas, whereas embedded states inside each super-state explain the associated facial region in greater depth. All of the trained HMMs in each class are fed the observation sequence of an unknown image. Each HMM generates a conditional probability, and the HMM with the highest probability determines the person's class.

1.10. Objectives and Contributions

This research aims to improve the accuracy and operation of the face recognition system by introducing a new method at different stages of the system.

In pre-processing, a new method based on the automatic contrast-limited adaptive histogram equalization with dual gamma correction has been introduced, which improves the accuracy of the system by eliminating the under and over enhancements of the histogram and gamma correction while maintaining the texture details in too dark and too bright areas.

A new technique has been introduced for feature extraction. This new method is based on the adaptive local binary pattern features and improves the correct recognition rate of the system. In addition to that, the proposed method improves the system's resistance to unwanted noise.

We also propose a new system based on the convolutional neural network with a new layer arrangement plus data augmentation. The new system shows better performance in accuracy compared to some of the existing superior systems, especially for the datasets with a relatively small number of samples per class.

Finally, a hierarchical method of face recognition is proposed. This system has two stages of feature extraction and classification. The first stage uses the active appearance graph model to extract the geographical features of the face and also employs SVM as the classifier, whereas the second stage utilizes the introduced CNN arrangement and data augmentation as the feature extraction with SVM and SoftMax as classifiers. The system proves to be more accurate compared to other existing methods, particularly for the databases with a high number of classes and a low number of training samples. Additionally, this system reduces the required augmented data to achieve a high accuracy percentage.

References

1. Voulodimos, Athanasios, et al. "Deep learning for computer vision: A brief review." Computational intelligence and neuroscience 2018 (2018).
2. Rajnoha, Martin, et al. "Towards identification of hypomimia in Parkinson's disease based on face recognition methods." 2018 10th International Congress on Ultra Modern Telecommunications and Control Systems and Workshops (ICUMT). IEEE, 2018.
3. Pujol, Francisco A., et al. "Entropy-based face recognition and spoof detection for security applications." Sustainability 12.1 (2019): 85.
4. Yue, Gaili, and Lei Lu. "Face recognition based on histogram equalization and convolution neural network." 2018 10th International Conference on Intelligent Human-Machine Systems and Cybernetics (IHMSC). Vol. 1. IEEE, 2018.
5. Rahman, Shanto, et al. "An adaptive gamma correction for image enhancement." EURASIP Journal on Image and Video Processing 2016.1 (2016): 1-13.
6. Yoon, Hyunsup, Youngjoon Han, and Hernsoo Hahn. "Image contrast enhancement based sub-histogram equalization technique without over-equalization noise." International Journal of Electrical and Computer Engineering 3.2 (2009): 189-195.
7. Reza, Ali M. "Realization of the contrast limited adaptive histogram equalization (CLAHE) for real-time image enhancement." Journal of VLSI signal processing systems for signal, image and video technology 38.1 (2004): 35-44.
8. Anila, Satish, and Nanjundappan Devarajan. "Preprocessing technique for face recognition applications under varying illumination conditions." Global Journal of Computer Science and Technology (2012).
9. Ayyavoo, Thamizharasi, and Jayasudha John Suseela. "Illumination pre-processing method for face recognition using 2D DWT and CLAHE." Iet Biometrics 7.4 (2018): 380-390.

10. Tran, Chi-Kien, et al. "SVM-based face recognition through difference of Gaussians and local phase quantization." International Journal of Computer Theory and Engineering 13.1 (2021): 1-8.

11. Alex, Ann Theja, Vijayan K. Asari, and Alex Mathew. "Local difference of gaussian binary pattern: Robust features for face sketch recognition." 2013 IEEE International Conference on Systems, Man, and Cybernetics. IEEE, 2013.

12. Arigela, Saibabu, and Vijayan K. Asari. "Locally tuned inverse sine nonlinear technique for color image enhancement." Image Processing: Algorithms and Systems XI. Vol. 8655. SPIE, 2013.

13. Kumar, Ashu, Amandeep Kaur, and Munish Kumar. "Face detection techniques: a review." Artificial Intelligence Review 52.2 (2019): 927-948.

14. Papageorgiou, Constantine P., Michael Oren, and Tomaso Poggio. "A general framework for object detection." Sixth International Conference on Computer Vision (IEEE Cat. No. 98CH36271). IEEE, 1998.

15. Osuna, Edgar, Robert Freund, and Federico Girosit. "Training support vector machines: an application to face detection." Proceedings of IEEE computer society conference on computer vision and pattern recognition. IEEE, 1997.

16. Rätsch, Matthias, Sami Romdhani, and Thomas Vetter. "Efficient face detection by a cascaded support vector machine using haar-like features." Joint Pattern Recognition Symposium. Springer, Berlin, Heidelberg, 2004.

17. Schneiderman, Henry, and Takeo Kanade. "A statistical method for 3D object detection applied to faces and cars." Proceedings IEEE Conference on Computer Vision and Pattern Recognition. CVPR 2000 (Cat. No. PR00662). Vol. 1. IEEE, 2000.

18. Rowley, Henry A., Shumeet Baluja, and Takeo Kanade. "Neural network-based face detection." IEEE Transactions on pattern analysis and machine intelligence 20.1 (1998): 23-38.

19. Fleuret, Francois, and Donald Geman. "Coarse-to-fine face detection." International Journal of computer vision 41.1 (2001): 85-107.

20. Wang, Yi-Qing. "An analysis of the Viola-Jones face detection algorithm." Image Processing On Line 4 (2014): 128-148.

21. Karamizadeh, Sasan, Shahidan M. Abdullah, and Mazdak Zamani. "An overview of holistic face recognition." IJRCCT 2.9 (2013): 738-741.

22. Zhu, Yani, Chaoyang Zhu, and Xiaoxin Li. "Improved principal component analysis and linear regression classification for face recognition." Signal Processing 145 (2018): 175-182.

23. Zhou, Changjun, et al. "Face recognition based on PCA image reconstruction and LDA." Optik 124.22 (2013): 5599-5603.

24. Hu, Changhui, et al. "Singular value decomposition and local near neighbors for face recognition under varying illumination." Pattern Recognition 64 (2017): 60-83.

25. Sharif, Muhammad, et al. "Face recognition using edge information and DCT." Sindh University Research Journal-SURJ (Science Series) 43.2 (2015).

26. Li, Chaorong, et al. "Learning features from covariance matrix of gabor wavelet for face recognition under adverse conditions." Pattern Recognition 119 (2021): 108085.

27. Liu, Jun, Songcan Chen, and Xiaoyang Tan. "Fractional order singular value decomposition representation for face recognition." Pattern Recognition 41.1 (2008): 378-395.

28. Hadid, Abdenour. "The local binary pattern approach and its applications to face analysis." 2008 First Workshops on Image Processing Theory, Tools and Applications. IEEE, 2008.

29. O'Shea, Keiron, and Ryan Nash. "An introduction to convolutional neural networks." arXiv preprint arXiv:1511.08458 (2015).

30. Albawi, Saad, Tareq Abed Mohammed, and Saad Al-Zawi. "Understanding of a convolutional neural network." 2017 international conference on engineering and technology (ICET). Ieee, 2017.

31. Wiskott, Laurenz, and Christoph Von Der Malsburg. "Recognizing faces by dynamic link matching." Neuroimage 4.3 (1996): S14-S18.

32. Suthaharan, Shan. "Support vector machine." Machine learning models and algorithms for big data classification. Springer, Boston, MA, 2016. 207-235.

33. Pisner, Derek A., and David M. Schnyer. "Support vector machine." Machine learning. Academic Press, 2020. 101-121.

34. Nefian, Ara V., and Monson H. Hayes. "An embedded HMM-based approach for face detection and recognition." 1999 IEEE International Conference on Acoustics, Speech, and Signal Processing. Proceedings. ICASSP99 (Cat. No. 99CH36258). Vol. 6. IEEE, 1999.

2 Employing New Automatic Contrast-Limited Adaptive Histogram Equalization with Adaptive Average Dual Gamma Correction in A Face Recognition System

2.1. Introduction

Face recognition has become one of the most popular methods in biometrics because of its properties, such as being a non-intrusive method, not requiring participation, and feasible technology. Although recent improvements in technology have created an acceptable level of accuracy, there is still room for improvement, especially when facing problems caused by uncontrolled situations. One of those problems that can affect the accuracy of the system is the variation of illumination in the image. This variation is usually caused by factors such as poor camera sensors, improper light sources, shutter cycle, and poor ambient lighting. Casted shadows, overexposed regions, and too dark or too bright images can reduce the recognition rate of the system by degrading the information and features of the image. One way to overcome this variation and distortion is to adjust the input image prior to feeding it to the system. This step is called pre-processing, and it can be found in almost every image processing application as one of the most important parts. Pre-processing techniques are classified into three main groups [1]: 1) systems based on the non-linear transfer function. 2) Histogram correction techniques. 3) and techniques in the frequency domain. Non-linear methods mostly modify the individual pixel values of the image [2]. The main characteristics of these techniques are ease of use and setting, making them ideal for contrast correction application. A widely used example of non-linear methods is gamma intensity correction. Gamma correction works on the discrete values of the image to enhance the lighting in dark images.

Histogram-correction techniques, on the other hand, tend to correct the contrast of the image by unifying the gray levels throughout the whole image [3],[4]. The remarkable property of the histogram equalization techniques is that they require a relatively low computational cost. A histogram of an image can be generally defined as the grayscale levels and frequency correlation. Equation 1 shows the histogram of a grayscale image.

$$H(i) = \frac{p_i}{N}, \quad i = 0,1, \dots, L-1 \tag{2.1}$$

where i determines the gray level of the image, p_i indicates the number of pixels in that gray level, and N is the total number of pixels in the image. Histogram equalization for the image, based on the definition of the $H(i)$ is:

$$HE_m = M(j) = (L-1)\sum_{i=0}^{j} H(i) \tag{2.2}$$

here $M(j)$ identifies the concept of the mapping function, and L is the dynamic range of the image.

The histogram equalization technique faces its major drawback in the images with noticeably and relatively large uniform dark regions, where the HE method tends to show over-enhancement in the output. Figure 2.1 demonstrates this behavior as an example.

Figure 2.1 a) Original image vs. b) HE applied

As figure 2.1 shows, there is over-enhancement in the bright area, which causes the texture to be lost. Moreover, the under enhancement is visible in the dark areas, especially if they are adjacent to the well-lit parts of the face. This particularly affects the recognition rate of the face recognition system by eliminating some of the valuable information hidden in the terminated textures of the face image.

Some efforts have been made to overcome these problems. Some of these enhancements introduce different distributions of the histogram for the image, such as brightness preserving bi-histogram equalization (BBHE) [5], equal area dualistic sub-image histogram equalization (DSIHE) [6], minimum mean brightness error bi-histogram equalization (MMBEBHE) [7], and Gaussian mixture model (GMM). In contrast, some other techniques try to modify the other aspects of the conventional HE, such as changing the type of the explored information to spatial, employing dual probability, and so on. Examples of these techniques are the bilateral Bezier curve (BBC) [8] and Contextual contrast enhancement [9].

The other methods for improving the conventional HE are related to the segmentation of the input image. The most noticeable disadvantage of the HE is the degradation of the features in the image with relatively large dark or bright areas. Dividing the image into different segments and regions helps to prevent that issue by breaking down those areas. Also, dealing with the contrast related to that specific sub-section gives better control to the algorithm over the image compared to considering a uniform contrast for the input. Some examples of these techniques are exposure-based sub-image histogram equalization (ESIHE) [10], median-mean-based sub-image clipped histogram equalization (MMSICHE) [11], segment selective dynamic histogram equalization (SSDHE) [12], and segment-dependent dynamic multi-histogram equalization (SDDMHE) [13]. The main difference between these methods is how they address the contrast enhancement distribution and correction.

Another considerable method of contrast enhancement is contrast limited adaptive histogram equalization (CLAHE) [14], which effectively solves the over-enhancement issue of histogram equalization. The main idea of the CLAHE is to divide the image into equal sub-section and perform the HE on each of those sections. Many studies applied CLAHE to improve the contrast of the image while preserving the information and avoiding over-enhancement [14][15]. However, in instances where the dark areas contain the majority portion of the image, CLAHE is incapable of improving the overall brightness of the image drastically. This flaw, however, could be solved if a non-linear technique such as gamma correction is applied to the output of the CLAHE. Combining these two techniques can provide a method to enhance the overall illuminance of the image while preserving the texture information in dark and bright regions. Techniques such as adaptive gamma correction with weighting distribution (AGCWD) [16], adaptive gamma correction (AGC) [17], and combination AGC and range limited bi-histogram equalization (RLBBHE) [18] take advantage of the principle of mixing two methods. Still, in the images with various illumination regions, the effectiveness of the adaptive gamma correction techniques is in doubt. Chang *et al.* [14] proposed a CLAHE combination with the dual gamma correction, but since the proposed technique is for improving the image for visual purposes, the subtlety of the overall function prevents the technique from being ideal for improving the illumination of the image for face recognition purposes.

This research proposes a new form of automatic contrast-limited adaptive histogram equalization. The adaptation happens in the clip point of the histogram based on the blocks' information. The proposed algorithm first divides the image into equal-sized blocks. The adaptive CLAHE is then applied to each block. Contrast enhancement is done adaptively based on the dynamic range of the whole image and each block for the clip point of that specific segment. After this stage, the first gamma correction is utilized to enhance the brightness of the block. Finally, the value of the second gamma correction is calculated based on an adaptive threshold. The threshold in this stage determines the mapping curve value to be either γ_1 or the average of γ_1 and γ_2.

The contributions of the proposed method are 1) The adaptive CLAHE based on the dynamic range of the block improves the texture details of the image without over enhancement in bright regions, 2) The firm gamma correction improves the overall illumination of the image while considering the multi-toned property of the image, 3) The second gamma correction eliminates the weakness of the general gamma correction only if necessary.

The rest of this chapter is organized as follows: Section 2.2 briefly explains the concept of CLAHE. The structure of the proposed algorithm is discussed in detail in section 2.3. Section 2.4 contains the results of the evaluation of the proposed technique. Finally, the conclusion is presented in section 2.5.

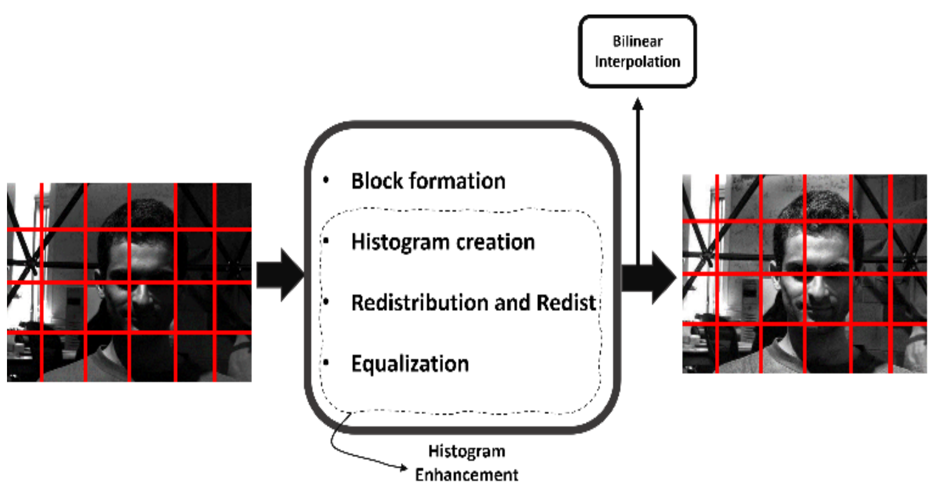

Figure 2.2 Block diagram of the CLAHE

2.2. CLAHE

The main difference between the contrast-limited adaptive histogram equalization and conventional HE is in the method that CLAHE uses to limit the clipping point. Figure 2.2 depicts the general stages of the CLAHE. As illustrated, there are five main stages to CLAHE. The algorithm divides the image into equal blocks in the first step. Then, the histogram enhancement is performed on each block separately. A histogram enhancement consists of three stages of histogram creation, histogram trimming, and redistribution of the clipped part. The final stage in CLAHE is the bilinear interpolation, which helps to reduce the artifacts between blocks and smoothen the output. The pick clipping point in CLAHE is as follows:

$$\beta = \frac{N}{D}\left(1 + \frac{\alpha}{100} G_m\right) \tag{2.3}$$

where N is the pixel count in the block, D is the dynamic range of that block, α is the clipping coefficient, and G_m is the maximum gradient. Needless to say, that α determines the level of contrast enhancement here. The bigger the α (closer to 100), the higher the contrast boost level will be. The same goes for the CDF and the remap function.

$$cdf(g) = \sum_{l=0}^{k} pdf(g) \tag{2.4}$$

$$R(g) = cdf(g) \times g_{max} \tag{2.5}$$

where $R(g)$ defines the function that remaps the clipped part to the histogram curve and g_{max} is the maximum gray value of the pixels in the block.

The next step in the process is the bilinear interpolation, which is defined by the following equations:

$$R\big(p(i)\big) = m.\big(n.R_a.p(i) + (1-n).R_b.p(i)\big) + (1-m).\big(n.R_c.p(i) + (1-n).R_d.p(i)\big) \quad (2.6)$$

$$\begin{cases} n = (x_b - x_p)/(x_b - x_a) \\ m = (y_c - y_p)/(y_c - y_a) \end{cases} \quad (2.7)$$

The advantage of the CLAHE is that it has a relatively lower computational cost due to its implication on the small sub-blocks of the image rather than acting on the whole image. Figure 2.3 demonstrates a) the clip point and b) the concept of bilinear interpolation.

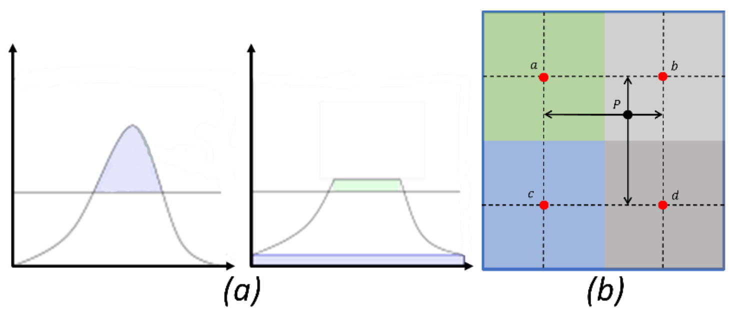

Figure 2.3 a) Clipping point and redistribution in CLAHE b) Bilinear Interpolation

2.3. Proposed Method

Although contrast enhancement methods provide some acceptable results in many cases, they still face shortcomings when the input image suffers from heavy dark regions and deep shadows. This is mainly due to the fact that β or clipping point is a fixed value, and therefore it cannot change by the nature of the poor illumination. In other words, a fixed global clipping point could give satisfactory results for some parts of the database, while there could be instances that suffer from contrast over-enhancement as well as cases that caries artifacts. In addition to that, improving the image's contrast alone is not enough to enhance the illumination to a sufficient level. This, especially, is visible in cases with very dark non-uniform areas.

Thus, these points justify the definition of an adaptive clipping point and gamma correction as the proposed method. The proposed technique uses an adaptive clipping point based on the parameters of each block. The algorithm also sets the gamma correction for g_{max} to enhance the dynamic range of the block.

2.3.1. Adaptive Clip Point Based on the Block Parameters

In conventional, fixed-point CLAHE, α and maximum gradient (G_m) are the parameters that determine the clipping point. To improve the performance of CLAHE, the algorithm arranges the clip point in a way to have lower values for smooth and harmonized areas and higher values for the blocks containing more illumination variation.

Two valuable parameters which reveal useful information about the block are median gray level and standard deviation. Using these parameters to set the adaptive clip point is presented in the form of the following equation:

$$\delta = \frac{g_{mod}}{D_b} \tag{2.8}$$

$$\beta = \frac{N}{D_b} \left(1 + Q\frac{\delta}{D} + \frac{\alpha}{100}\left(\frac{\sigma}{Avg+c}\right)\right) \tag{2.9}$$

where g_{mod} is the most repeated gray value of the block, δ is the dynamic factor, σ is the standard deviation, Avg is the mean value, and c is a small value. Q and α control dynamic range and entropy weight, respectively.

The new clip point has a higher value for the blocks with more texture information since those blocks have a bigger σ/Ave. Figure 2.4 shows an example of the new clipping point for different image blocks.

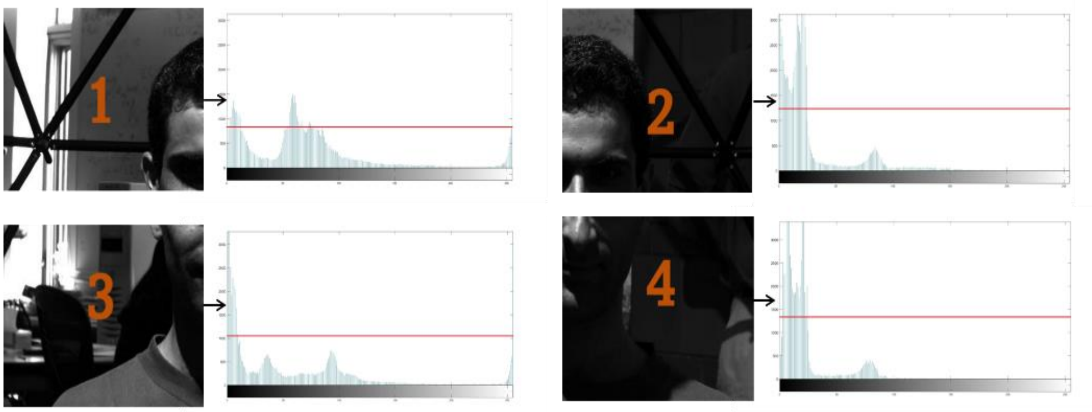

Figure 2.4 Content Controlled Clip Point

2.3.2. Gamma Correction

Although the algorithm improves the contrast based on the global and local information and properties of the image, the histogram equalization alone cannot provide enough enhancement to the image, especially in the uniform dark areas of the image. This shows its effects in the regions with extreme cast shadows or in the situation where a spotlight is used as the source of lighting as the face image contains shadows of the face parts blocking the light (around the nose or eyes area). To solve this issue, we propose gamma correction. The gamma correction process is technically increasing the gray level variation in shaded areas. This improves the illumination by increasing the features in the dark regions. The mapping function of gamma correction is defined as follows:

$$R(g) = g_{max}\left(\frac{g}{g_{max}}\right)^{\gamma} \tag{2.10}$$

This relation between g_{max}, $R(g)$, and α determines the behaviour of the gamma correction. The lower values of γ, would deliver a more intense increase in the value of the corresponding pixel. Still, because of the pre-defined value of the gamma correction, this change happens through all the pixels of the image block and can result in contrast degradation. Even though the gamma value acts locally, the contrast degradation cancels the effect of the CLAHE. The proposed algorithm implements the gamma correction into the introduced CLAHE to prevent this. To gain this, we first need to introduce a weighted gray level correction based on the γ value.

$$w = \left(\frac{GL_{mod}}{GL_{med}}\right)^{1-\gamma_1} \tag{2.11}$$

where GL_{mod} is the most repeated gray level and GL_{med} represents the median gray level of the image.

We then calculate the new g'_{max} and after that, get the new remapping function by implementing that new maximum gray level into the equation (2.5).

$$g'_{max} = w \times g_{max} \tag{2.12}$$

$$R_1(g) = cdf(g) \times g'_{max} \tag{2.13}$$

So far, the proposed algorithm enhances the image by improving the contrast of the blocks of the image with the introduced adaptive clipping function while preventing the halo artifacts, and, at the same time, the implemented gamma correction enhances the illumination by creating the gray level differences and perform this without under enhancement in the shadowed areas. However, in cases where both dark parts and lit parts exist, the system encounters a problem. This problem comes from the fact that in those cases, our method tends to under-enhance the image in dark regions because of the overall average of the image's luminance. In other words, because the image's dark and bright areas cancel each other, the overall global grayscale difference of the image does not improve properly by applying just one gamma correction. Our solution to that problem is adding another gamma correction stage to the algorithm to just be applied to those specific cases. To better soften the transition of the mapping curve from the CLAHE-gamma correction to the second gamma correction, we introduce an average mapping function as R_γ. The overall mapping function, the threshold factor, and second gamma correction are shown in the following equations:

$$R_\gamma(g) = \frac{\left[w_2 \times g_{max}\left(\frac{g}{g_{max}}\right)^2\right] + [cdf(g) \times g'_{max}]}{2 \times w_2} \qquad (2.14)$$

$$\text{If} \quad d > D_{th}, \quad R(g) = \max(R_1, R_g)$$

$$\text{else}, \; R(g) = R_\gamma$$

where R_1 is the remapping function of CLAHE-gamma enhancement, w_2 is the second weight correction, d is the dynamic range of the block, D_{th} is the dynamic range threshold, and R is the final remapping function of the algorithm. Figure 2.5 demonstrates an example of the mapping curve correction.

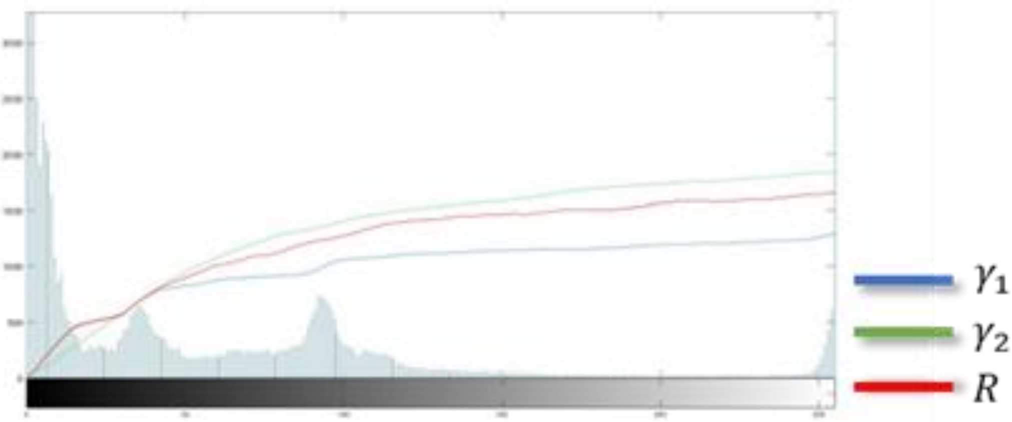

Figure 2.5 Mapping curve adjustment example

For final adjustments, as suggested in [14], we set the γ_1 and γ_2 as follows to be increasing functions:

$$\gamma_1 = \frac{\ln(e+cdf(g))}{8} \tag{2.15}$$

$$\gamma_2 = \frac{1+cdf_{\omega}(g)}{2} \tag{2.16}$$

where cdf_{ω} is the weighted distribution function [19], and e is a constant. Figure (2.6) shows the result of performing a) CLAHE and b) the proposed method on a face image.

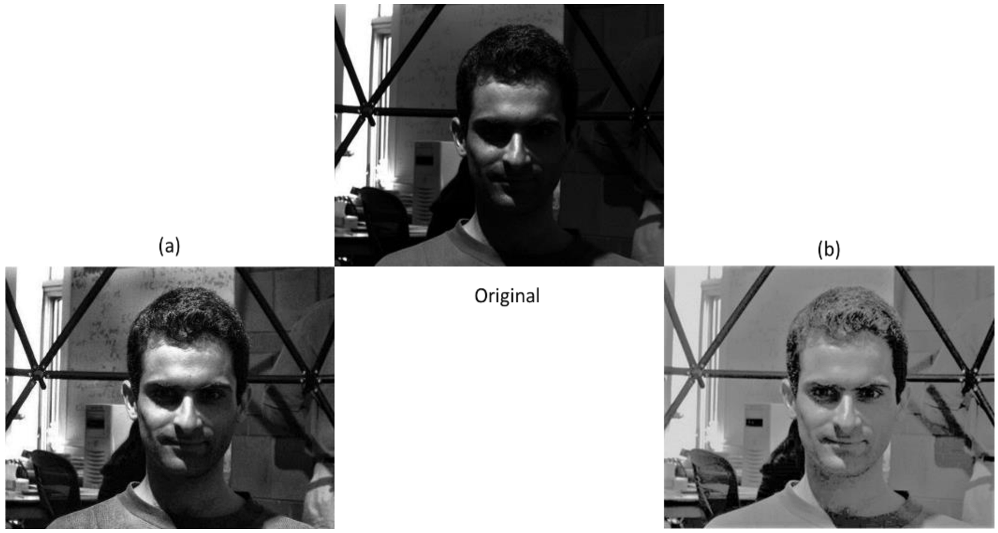

Figure 2.6 The outputs from a) CLAHE b) Proposed method

2.4. Experimental Setup and Results

In this research, the Extended Yale Dataset is used to evaluate the quality of the proposed method by comparing the accuracy rate of the different face recognition methods with and without the suggested technique as the pre-processing stage, as well as assessing the performance of various pre-processing techniques including the proposed system applied to a specific face recognition algorithm.

2.4.1. Local Binary Pattern [20]

LBP is considered to be one of the most popular feature extracting methods in local appearance-based systems. The principle behind LPB is to extract the texture features of the image by comparing a pixel with a certain number of its neighbouring pixels. However, there are some parameters to set the behaviour of the LBP. Some of these variables are the number of the neighbouring pixels, the direction of rotation in LBP code formation, radius, and encoding function. The following equations are the mathematical representation of LBP code formation:

$$b_p = \begin{cases} 1 \; if \; D_p \geq 0 \\ 0 \; if \; D_p < 0 \end{cases} \tag{2.17}$$

where D_p is the pixel difference. Figure (2.7) shows a general diagram of the LBP.

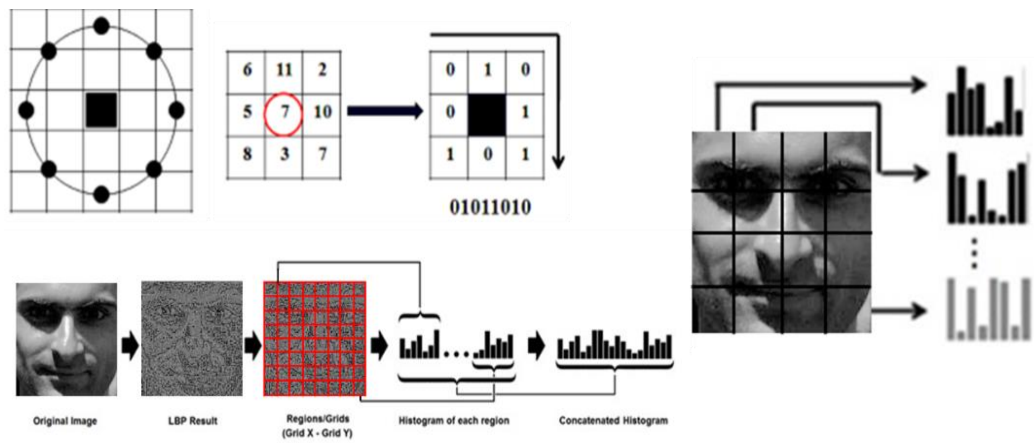

Figure 2.7 Different steps of local binary pattern feature extractor

2.4.2. Extended Yale Dataset [21]

The Extended Yale Dataset of face images contains a 28-subject expansion of the original Yale B Dataset. The dataset includes 21888 photos from a mono lighting setup. Thirty-eight individuals participated in the experiment, with a total number of 576 viewing conditions. For each member, a sample with background illumination was captured, accompanied by a specific head pose of that person. Extended Yale B Dataset is divided into 5 distinct subgroups based on the angle between the light source, direction and the camera axis, and lighting situation. Subgroup 5, based on our experience, has the most challenging lighting condition and, in most cases, shows the lowest recognition rate.

2.5. Experimental Setup

First Experiment. In the first setup, we evaluate the effectiveness of our method by implementing the proposed technique in several existing face recognition methods and investigate the changes that our approach brings to those systems. As for the arrangement of the experiment, we used 40 images of the Yale dataset samples for training and 24 for testing. We also resize the images to 64 by 64 pixels as the final stage of pre-processing after the illumination enhancement. Table 2.1 compares the recognition rate of different methods for two situations of pre-processing, 1) without any and 2) The proposed technique. It is worth mentioning that the results were acquired from an average of 20 runs of the algorithms on the dataset for each technique.

Table 2.1 Effect of presence and absence of the proposed method on the %recognition rate of the different face recognition methods

Method	No pre-processing	**Proposed Method**
MRF [22]	78.11	**84.51**
PCA [23]	66.10	**78.19**
HMM [24]	84.38	**89.46**
LDA+IPMML [25]	81.74	**88.63**

As we can see from table 2.1, adding the proposed pre-processing technique increases the recognition rate of each method significantly.

Second Experiment. In this test, we try to evaluate the performance of the proposed method compared to other pre-processing techniques. To do so, we need to apply different pre-processing techniques to one specific face recognition method to be able to assess the results of each technique appropriately. In other words, the recognition rate of the combinations of different pre-processing techniques with the selected face recognition system is utilized to determine the quality and effectiveness of that pre-processing. Ultimately, we chose the local binary pattern (LBP) as the feature extractor of the face recognition algorithm as well as the support vector machine (SVM) [26] as the classifier of that system. Again, we resize the images to 64 by 64 before feeding them into the feature extractor. We also use the same 40 and 24 numbers as the training set and test set for this test. Table 2.2 contains the results of this experiment.

Table 2.2 %Recognition rate of a LBP + SVM face recognition system with different preprocessing methods

Pre-process Method	Recognition Rate
None	82.13
HE [27]	82.44
SQI [28]	89.70
GIC [29]	84.24
DoG [30]	90.15
CLAHE [14]	87.36
Proposed Method	**93.74**

2.6. Conclusion

This research proposes a new pre-processing method for application in face recognition. The proposed method is based on contrast-limited adaptive histogram equalization and gamma correction. In the process of illumination enhancement, the algorithm first employs a new adaptive clipping point for histogram equalization based on the texture information of the image blocks. Then in order to prevent the problems such as halo artifacts, under-enhancement and over-enhancement, we implemented a first gamma correction function into the structure of the CLAHE. This first gamma correction helps improve the overall image's grayscale variation. Finally, since some of the samples could have dark and bright regions together, we introduce a second gamma correction function to prevent the negative effect of the existing those two types of lighting in a face image. To evaluate the success of our method, we designed two experiments. In the first setup, we chose some existing face recognition systems and compared the results of adding our technique to those methods by measuring the recognition rate before and after our method. The results showed that the proposed pre-processing technique considerably improves the recognition rate in every one of those systems. For the second part of the experiment, the LBP + SVM technique is selected as the reference face recognition system. In this arrangement, the efficacy of the proposed method is evaluated by comparing the recognition rate of the system above mated to different pre-processing methods. This experiment also shows that the proposed method has a better performance in improving the recognition rate of that specific face recognition method compared to others.

References

1. Łoza, Artur, et al. "Automatic contrast enhancement of low-light images based on local statistics of wavelet coefficients." Digital Signal Processing 23.6 (2013): 1856-1866.

2. Gonzalez, Rafael C. Digital image processing. Pearson Education India, 2009.

3. Kaur, Manpreet, Jasdeep Kaur, and Jappreet Kaur. "Survey of contrast enhancement techniques based on histogram equalization." International Journal of Advanced Computer Science and Applications 2.7 (2011).

4. Arici, Tarik, Salih Dikbas, and Yucel Altunbasak. "A histogram modification framework and its application for image contrast enhancement." IEEE Transactions on image processing 18.9 (2009): 1921-1935.

5. Moniruzzaman, Md, Md Shafuzzaman, and Md Foisal Hossain. "Brightness preserving Bi-histogram equalization using edge pixels information." 2013 International Conference on Electrical Information and Communication Technology (EICT). IEEE, 2014.

6. Dhamodharan, S., and P. Shanmugavadivu. "Brightness preserving contrast enhancement of digital mammogram using modified-dualistic sub-image histogram equalization." Int. J. Comput. Sci. Eng 6.4 (2018): 257-261.

7. Chen, Soong-Der, and Abd Rahman Ramli. "Minimum mean brightness error bi-histogram equalization in contrast enhancement." IEEE transactions on Consumer Electronics 49.4 (2003): 1310-1319.

8. Lai, Yu-Ren, et al. "Improved local histogram equalization with gradient-based weighting process for edge preservation." Multimedia Tools and Applications 76.1 (2017): 1585-1613.

9. Parihar, Anil Singh, Om Prakash Verma, and Chintan Khanna. "Fuzzy-contextual contrast enhancement." IEEE Transactions on Image Processing 26.4 (2017): 1810-1819.

10. Acharya, Upendra Kumar, and Sandeep Kumar. "Directed searching optimized mean-exposure based sub-image histogram equalization for grayscale image enhancement." Multimedia Tools and Applications 80.16 (2021): 24005-24025.

11. Srikanth, Rangu, et al. "Improved Image Enhancement of Natural Images with Median Mean-Based Sub-Image Clipped Histogram Equalization." Soft Computing for Security Applications. Springer, Singapore, 2022. 873-887.

12. Khan, Mohammad Farhan, Ekram Khan, and Z. A. Abbasi. "Segment selective dynamic histogram equalization for brightness preserving contrast enhancement of images." Optik 125.3 (2014): 1385-1389.

13. Khan, Mohammad Farhan, Ekram Khan, and Z. A. Abbasi. "Segment dependent dynamic multi-histogram equalization for image contrast enhancement." Digital Signal Processing 25 (2014): 198-223.

14. Chang, Yakun, et al. "Automatic contrast-limited adaptive histogram equalization with dual gamma correction." Ieee Access 6 (2018): 11782-11792.

15. Stimper, Vincent, et al. "Multidimensional contrast limited adaptive histogram equalization." IEEE Access 7 (2019): 165437-165447.

16. Huang, Shih-Chia, Fan-Chieh Cheng, and Yi-Sheng Chiu. "Efficient contrast enhancement using adaptive gamma correction with weighting distribution." IEEE transactions on image processing 22.3 (2012): 1032-1041.

17. Rahman, Shanto, et al. "An adaptive gamma correction for image enhancement." EURASIP Journal on Image and Video Processing 2016.1 (2016): 1-13.

18. Zuo, Chao, Qian Chen, and Xiubao Sui. "Range limited bi-histogram equalization for image contrast enhancement." Optik 124.5 (2013): 425-431.

19. Huang, Shih-Chia, Fan-Chieh Cheng, and Yi-Sheng Chiu. "Efficient contrast enhancement using adaptive gamma correction with weighting distribution." IEEE transactions on image processing 22.3 (2012): 1032-1041.

20. Deeba, Farah, et al. "LBPH-based enhanced real-time face recognition." International Journal of Advanced Computer Science and Applications 10.5 (2019).

21. Lee, Kuang-Chih, Jeffrey Ho, and David J. Kriegman. "Acquiring linear subspaces for face recognition under variable lighting." IEEE Transactions on pattern analysis and machine intelligence 27.5 (2005): 684-698.

22. Arashloo, Shervin Rahimzadeh. "A comparison of deep multilayer networks and Markov random field matching models for face recognition in the wild." IET Computer Vision 10.6 (2016): 466-474.

23. Zhi, Hui, and Sanyang Liu. "Face recognition based on genetic algorithm." Journal of Visual Communication and Image Representation 58 (2019): 495-502.

24. Aggarwal, Akarsh, et al. "Principal component analysis, hidden Markov model, and artificial neural network inspired techniques to recognize faces." Concurrency and Computation: Practice and Experience 33.9 (2021): e6157.

25. Zhou, Lijian, et al. "Face recognition based on LDA and improved pairwise-constrained multiple metric learning method." Journal of Information Hiding and Multimedia Signal Processing 7.5 (2016): 1092.

26. Guo, Shanshan, Shiyu Chen, and Yanjie Li. "Face recognition based on convolutional neural network and support vector machine." 2016 IEEE International conference on Information and Automation (ICIA). IEEE, 2016.

27. Yue, Gaili, and Lei Lu. "Face recognition based on histogram equalization and convolution neural network." 2018 10th International Conference on Intelligent Human-Machine Systems and Cybernetics (IHMSC). Vol. 1. IEEE, 2018.

28. Jung, Cheolkon, and Jianpeng Yin. "SQI-based illumination normalization for face recognition based on discrete wavelet transform." 2016 IEEE International Conference on Image Processing (ICIP). IEEE, 2016.

29. Han, Hu, et al. "A comparative study on illumination pre-processing in face recognition." Pattern Recognition 46.6 (2013): 1691-1699.

30. Tran, Chi-Kien, et al. "SVM-based face recognition through difference of Gaussians and local phase quantization." International Journal of Computer Theory and Engineering 13.1 (2021): 1-8.

3 New Local Binary Pattern Feature Extractor with Adaptive Threshold for Face Recognition Applications

3.1. Introduction

One of the fastest-growing topics among image processing applications is face recognition. Two factors contributing to that expansion are improvements in infrastructure and demand increase. Face recognition as a non-intrusive biometric system shows a promising future in terms of the scope of application. All these together make face recognition a desirable research topic. Over the past few decades, significant improvements have been brought to face recognition systems. The early methods were slow, complicated, and relatively unreliable, with a low accuracy rate.

In comparison, recent advances improved the algorithms in all aspects. These days, face recognition systems are far more trustworthy, consistence, easier to implement, and comparatively fast. These systems can be found vastly in any number of practical applications, from security to marketing and customer service to safety. However, there is still plenty of room available to make enhancements. For instance, one of the most important parameters of a face recognition system is that system’s accuracy. In most applications, accuracy is preferred over speed. Recently, methods and approaches have been introduced to the field that provide a promising recognition. However, a challenge current systems face is unpredicted parameters such as noise and poor illumination. These uncontrolled situations drastically affect the system's reliability to the extent that they render the system unfit for some applications. It is worth mentioning that the unwanted factors are more likely to happen in real-life situations and applications, which simulated experiments lack. Therefore, any effort to improve the system's accuracy is still appreciated. Multiple approaches have been developed to deal with the mentioned issues. However, to better understand the pathway, it is sensible to first break down the stages of a typical face recognition system. Then the enhancements can be directed towards improving each of those levels. Figure 3.1 demonstrates the block diagram of a general face recognition algorithm.

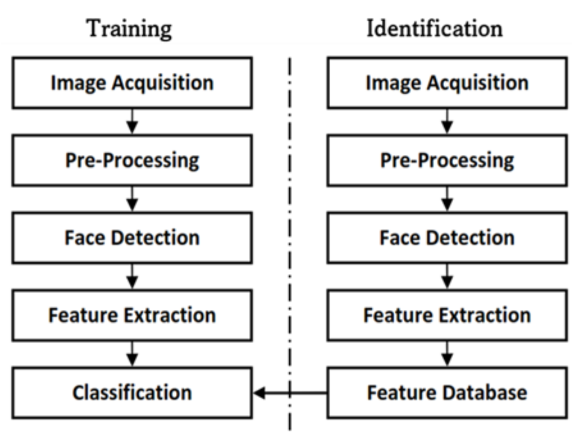

Figure 3.1 Flow diagram of the face recognition process

As figure 3.1 shows, the four main stages of the algorithm that are in reality in-volved in the computation are 1) pre-processing, 2) face detection, 3) feature extraction, and 4) Classification. Improving the operation of any of these stages could result in a system with a higher recognition rate. For the pre-processing stage, [1] and [2] suggested methods to enhance the illumination of the image while preserving the texture information and by creating a more uniform global bright-ness on the image and therefore increasing the features, boosting the recognition rate. Techniques employed in the face detection section [3] and [4] promote the accuracy of the system by removing the unnecessary and excess information from the face image, such as background, hair, and sometimes even forehead, to reduce the computational load of the system as well as decrease the chance of error by narrowing down the choices for the system. Other approaches [5] and [6] try to utilize the most suitable classifier based on the type of dataset that they are working with to optimize the system's operation and improve the rate by that.

However, one of the most crucial sections of any face recognition system or, in general, an image processing system is the feature extraction part. The feature extractor has the highest weight in defining the complexity of the system. In addition to that, some of the system's properties directly come from this stage, features such as robustness against noise, speed, and most of the computational complexity. Feature extraction, mainly, can be divided into two major categories: 1) holistic-based methods, which tend to reduce the dimensionality of the face image while maintaining the essential required information. Principle component analysis (PCA) [7], linear discriminant analysis (LDA) [8], and local binary pattern [9] are examples of this approach. 2) Methods that are based on the features of the frequency space of the image, such as discrete cosine transform (DCT) [10] and Gabor wavelet [11].

Computational complexity is a parameter that plays a determining role in selecting an extraction technique. Due to its computational simplicity, strength against illumination variation, and ease of implementation, the local binary pattern (LBP) is one of the most popular techniques for face recognition. Though the conventional local binary pattern has its downsides, nevertheless, the advantages outweigh the shortcomings. This paper proposes a new approach to LBP for face recognition using a new adaptive function in the operator's structure. The new operator is modified to adapt based on a variation of the gaussian distribution function. The idea for this improvement comes from a conventional LPB's weaknesses: having a fixed threshold.

The contributions of the proposed method are 1) The new adaptive threshold by replacing the traditional way of thresholding at zero, maintains the information that is hidden in the amplitude of the pixel difference, 2) The proposed method improves the accuracy of the system by introducing the adaptive thresh-old function and a new error function to replace the fixed zero thresholding in conventional LBP, 3) In addition to that, the proposed method improves the robustness of the system against the noise.

The rest of this chapter is structured as follows: Section 3.2 briefly explains the theory of LBP. The shape of the proposed system is explained in detail in section 3.3. Section 3.4 includes the results of the experiment with the proposed method. Finally, the conclusion is stated in section 3.5.

3.2. Local Binary Pattern

Among different existing methods of feature extraction, LBP is one the most widely held techniques, especially in face recognition applications, due to its ease of implementation and successful outcomes. The LBP is a local appearance-based method that extracts the image's texture features by comparing each pixel with its neighbors. There is no training requirement which makes it fast and easily integrable into the new data sets, and also, it is robust to rotation, scaling and illumination variation.

Local Binary Pattern feature extraction is based on the fact that a double corresponding parameter can be defined to describe a 2-dimensional interpretation of an image. Local pattern and grayscale contrast are those two paired elements that are used to form the histogram. The idea was first introduced by Ojala et al. in 1996 [12]. The initial LBP operator considers a 3 by 3 window of neighboring pixels to assign a tag to a pixel. This is called LBP coding. The three neighboring pixels provide $2^8 = 256$ levels to form the histogram as the feature map. Later on, the system was modified to adopt a different window size than the original 3x3. Furthermore, a circular area, as well as bilinear interpolation, were introduced to improve the freedom of choice for selecting the neighbors’ number. The figure 3.2 shows the concept of the LBP neighbor selection process.

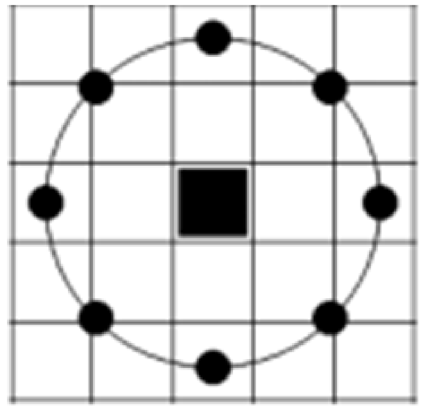

Figure 3.2 Radius and number of neighboring pixels in LBP

P and R are the number of pixels and radius of the circle, respectively. These two parameters determine the pixels that are used to be compared with the center to create the LBP code. Now that the algorithm specified the pixels involved in the process, the next step is finding the difference between the reference pixel in the center and other pixels. The calculated pixel difference ultimately denotes the LBP code based on the equation (3.1). Figure 3.3 depicts this step for a sample.

$$b_p = \begin{cases} 1 \ if \ D_p \geq 0 \\ 0 \ if \ D_p \leq 0 \end{cases} \quad p = 1,2,\ldots,N \tag{3.1}$$

where b_p is the LBP code of the pixel p, N is the number of pixels, and D_p is pixel difference.

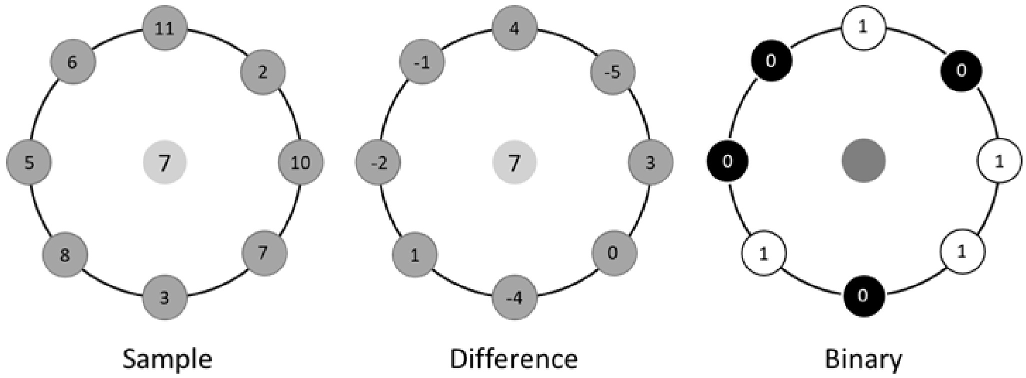

Figure 3.3 An example of LBP coding execution

Finally, a decimal number is calculated based on the binary code created by the algorithm. The following equation shows the process of computing this value.

$$LPB_i = \sum_{p=1}^{N} b_p(D_p) \times 2^{p-1} \qquad (3.2)$$

The other variable in this setting is the direction of rotation which defines the sequence of the pixels and gets two values, clockwise and counterclockwise. We also can set the starting point in the process of decimal production, which should be the same for all the pixels of the image.

Now that the image is converted into a series of decimal values created based on the relationship between each pixel and its neighbors, the next step is generating the image's histogram. Albeit the histogram development is done on the blocks of the image rather than the whole image. The diagram of these steps can be seen in figure 3.4.

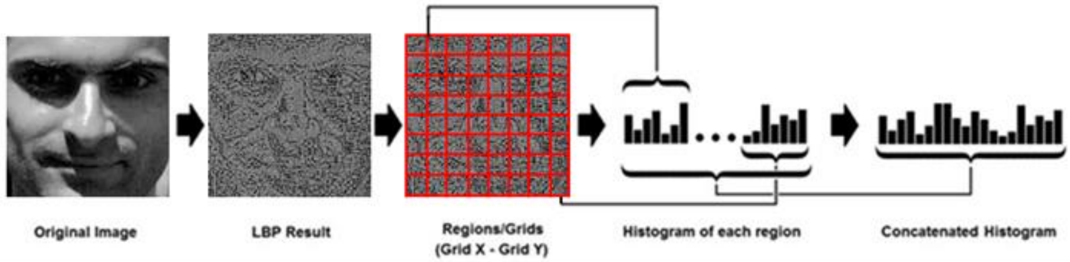

Figure 3.4 Histogram of an image based on the LBP code

3.2.1. Pros and Cons of The Conventional LBP

As mentioned before, LBP is considered one of the most reliable texture feature extraction methods. Many researchers in the field of face recognition employed LBP as the feature extractor for their system. The features like simplicity and robustness against monotonic changes in the image, such as illumination variation, rotation, and scaling, help in that regard. However, when it comes to maximum potential, one major flaw is that the conventional LBP employs a fixed zero threshold for the pixel difference, only considers the sign of the D_p, and dismisses the amplitude. A fixed zero threshold increases the sensitivity of the system to noise, and important information in the amplitude is missed, which decreases the accuracy of the method.

3.3. Proposed Method

Monotonic changes in the face image are the most affecting occurrences that alter the final results of the face recognition algorithm. Illumination variation, as one of these changes, shifts the pixel intensity of the image and thus changes the pixel difference D_p. However, it can be determined from (3.1) that this variation is not completely reflected since the algorithm only considers the sign of the difference and omits amplitude. Needless to say, that important information about the image can be extracted from the dismissed amplitude. In addition to that, the fixed zero threshold in conventional LBP adds sensitivity to noise and makes the system vulnerable. This comes from the fact that an added noise can change the binary bits from zero to one or vice versa by bringing a slight alteration. To overcome these issues, we propose to replace the zero thresholding for the pixel difference with an adaptive function, as is shown follows:

$$b_p^i = \begin{cases} 1 & if \;\; D_p \geq T_i \\ 0 & if \;\; D_p < T_i \end{cases} \qquad p = 1,2,\ldots,N \tag{3.3}$$

Here i refers to the block number, N is the number of pixels, and T_i is the adaptive threshold function.

The block size and numbers are fixed in our proposed method. But the threshold adapts to the global and local information of the image and blocks. We based the threshold function on the cumulative density function (CDF) of the Gaussian distribution function anticipating having both strong and responsive LBP features, as below:

$$f_i = {}^1\!/_2 \left[1 + \operatorname{erf}\left(\frac{\sigma_i}{\sigma\sqrt{2}\sqrt{|\mu-\mu_i|}}\right)\right], \qquad i = 1,\ldots,M \tag{3.4}$$

Where M is the number of image blocks, σ and μ are standard deviation and global mean, and μ_i and σ_i are the mean and standard deviation in a frame of the image. The error function is defined by erf (x) as below:

$$\operatorname{erf}(x) = \frac{2}{\sqrt{\pi}}\int_0^x e^{-t^2}\,dt \tag{3.5}$$

Figure 3.5 illustrates the proposed adaptive threshold for different values of μ and σ.

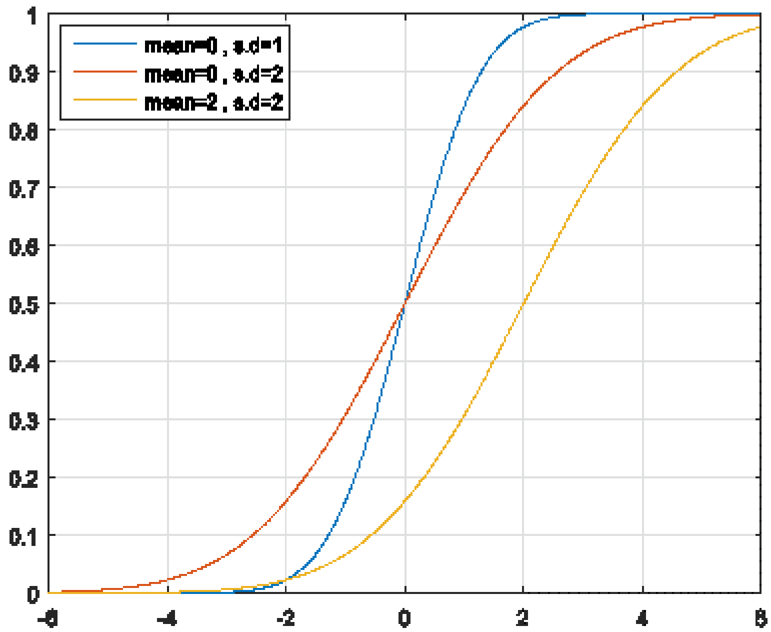

Figure 3.5 The proposed adaptive thresholding function with three different mean and standard deviation values

Finally, substituting (3.4) into the main framework of the LBP in (3.3) provides the ultimate LBP coding of the proposed method (A$_d$T-LBP):

$$b_p = \begin{cases} 1 & if \;\; D_p \geq f_i \\ 0 & if \;\; D_p < f_i \end{cases} \tag{3.6}$$

Also, worth mentioning that noise properties have a governing role in the development of the threshold function. Increases in the noise will result in a rise in the value of the threshold function and vice versa, which can be translated to more robust LBP features.

3.4. Experimental Results

In this section, we conduct two experiments and simulations and compare the results with other existing approaches to evaluate the effectiveness of the proposed method over the others. We first need to define a face recognition system based on the diagram in figure 3.1. For the pre-processing, our system employs a simple image resize and normalization. The feature extraction obviously is the method proposed in this paper (A_dT-LBP), and finally, for the classifier, we decided to go with the support vector machine (SVM). The selected dataset for the experiments is Extended Yale B. The first experiment compares the recognition rates of several existing face recognition systems with different feature extraction methods with the results of the proposed technique. In the second simulation, two different types of random noise (Salt & Pepper and Gaussian) are added to the database, and the effect of these noises on the accuracy is investigated for the previous approaches as well as the proposed method (A_dT-LBP). In the following, a brief introduction to the SVM and dataset is presented before we get into the results.

3.4.1. Support Vector Machine

Support vector machines (SVM) are characterized as maximum margin classifiers since they reduce classification error while also increasing geometric margin. An SVM creates a separating hyperplane in the feature space that maximizes the margin between the data sets. To determine the margin, two parallel hyperplanes are built, one on either side of the separating one. These hyperplanes are pushed up against the two data sets, allowing the hyperplane with the farthest distance to the adjoining support vectors of both classes to achieve a decent separation. The greater the margin or distance between these parallel hyperplanes, the more likely unknown samples will be accurately identified [13]. Figure 3.6 demonstrates the typical arrangement of the SVM.

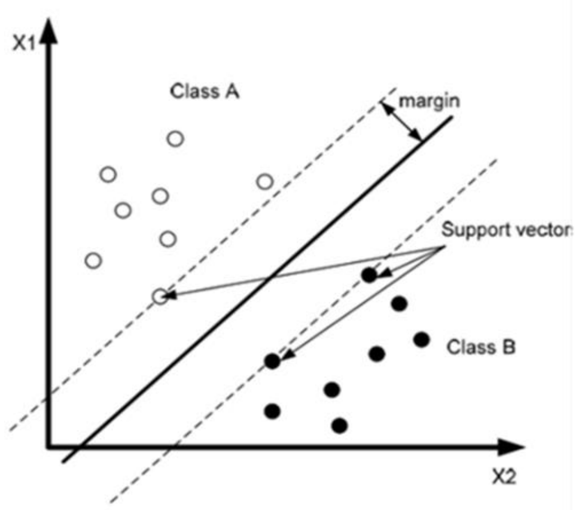

Figure 3.6 SVM and its maximum-margin hyperplane.

3.4.2. Extended Yale Dataset [14]

The Extended Yale Dataset of face images contains a 28-subject expansion of the original Yale B Dataset. The dataset includes 21888 photos from a mono lighting setup. Thirty-eight individuals participated in the experiment, with a total number of 576 viewing conditions. For each member, a sample with background illumination was captured, accompanied by a specific head pose of that person. Extended Yale B Dataset is divided into 5 distinct subgroups based on the angle between the light source, direction and the camera axis, and lighting situation. Subgroup 5, based on our experience, has the most challenging lighting condition and, in most cases, shows the lowest recognition rate.

3.4.3. Experimental Setup

First Experiment

In the first experiment, we compare the accuracy of the proposed system with other face recognition methods. The Extended Yale B dataset is selected for the experiment. Our proposed system uses SVM as the classifier. All the images of the database are resized to 64x64 prior to being fed to the algorithm. We used 40 random samples for training and 40 random for the test. Table 3.1 contains the results of the simulation. These results are gathered from the average of 20 runs of the algorithm on the dataset.

Table 3.1 %Recognition rate of different methods

Method	Recognition Rate
LDA+IPMML [15]	81.74
MRF [16]	78.11
LBP + SVM [17]	84.38
SVD+HMM [18]	95.56
Proposed Method	**97.75**

As we can see, the proposed method has a higher recognition rate than other techniques.

Second Experiment

In this simulation, the algorithm is tested against image noise. Two different types of noise, salt and pepper and Gaussian noise, are added to the samples, and their effects on the recognition rate are investigated. These two noises are categorized into two groups. Noise #1 is Gaussian noise with mean = 0 and variance = 0.05. Noise #2 is Salt and Pepper noise with noise density = 0.06. Note that these noises are also added to the inputs for other methods. The proposed system in this experiment has the LBP as the feature extractor and SVM as the classifier. We used the Extended Yale B database for training and testing. All the images of the database are resized to 64x64 prior to being fed to the algorithm. We used 40 random samples for training and 40 random for the test. Table 3.2 contains the results of the simulation.

Table 3.2 %Recognition rate of several face recognition systems in the presence of two different types of noise compared to the proposed method

Method	Recognition Rate	Noise #1	Noise #2
LDA+IPMML [15]	81.74	73.53 ↓8.21	66.2 ↓15.54
MRF [16]	78.11	66.93 ↓11.18	60.75 ↓17.36
LBP + SVM [17]	84.38	74.56 ↓9.82	68.79 ↓15.59
SVD+HMM [18]	95.56	87.94 ↓7.62	83.35 ↓12.21
Proposed Method	**97.75**	**96.38 ↓1.37**	**93.2 ↓4.55**

The second experiment shows that the proposed method has the lowest decrease in the recognition rate with the added noise.

3.5. Conclusion

This research proposed a new adaptive thresholding function for the LBP feature extractor in face recognition applications. The new function is based on a variation of the Gaussian distribution function. The proposed approach eliminates the shortcomings of employing fixed thresholding at zero in conventional LBP. The proposed technique amplifies the resistance of the LBP against noise by changing the threshold from zero to an adaptive function. Also, by considering the amplitude of the pixel difference and utilizing it in the computations of the LBP code, the proposed system preserves a significant part of the image information previously neglected in conventional LBP. In the course of this research, the new adaptive thresholding approach is evaluated by comparing the accuracy of a face recognition system that makes use of the introduced feature extraction technique and SVM classifier with other existing methods. The Extended Yale B dataset is used as the test set. The proposed method shows a higher recognition rate than other existing techniques. Also, to evaluate the system's behavior against noise, we added two types of Gaussian and Salt and Pepper noise to the dataset and measured the new recognition rate of the system. Again, the proposed approach demonstrates the lowest decrease among the presented techniques and proves to have the highest robustness to the unwanted noise.

References

1. Ayyavoo, Thamizharasi, and Jayasudha John Suseela. "Illumination pre-processing method for face recognition using 2D DWT and CLAHE." Iet Biometrics 7.4 (2018): 380-390.
2. Starovoitov, V. V., et al. "Image enhancement for face recognition." International Conference on Iconics. 2003.
3. Fu, Limei, and Xinxin Shao. "Reseach and Implementation of Face Detection, Tracking and Recognition Based on Video." 2020 International Conference on Intelligent Transportation, Big Data & Smart City (ICITBS). IEEE, 2020.
4. Li, Chong, et al. "Face detection based on YOLOv3." Recent Trends in Intelligent Computing, Communication and Devices. Springer, Singapore, 2020. 277-284.
5. Gong, Sixue, Xiaoming Liu, and Anil K. Jain. "Mitigating face recognition bias via group adaptive classifier." Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2021.
6. Ibrahim, Abdelhameed, et al. "Optimized superpixel and AdaBoost classifier for human thermal face recognition." Signal, Image and Video Processing 12.4 (2018): 711-719.
7. Gottumukkal, Rajkiran, and Vijayan K. Asari. "An improved face recognition technique based on modular PCA approach." Pattern Recognition Letters 25.4 (2004): 429-436.
8. Lu, Juwei, Konstantinos N. Plataniotis, and Anastasios N. Venetsanopoulos. "Face recognition using LDA-based algorithms." IEEE Transactions on Neural networks 14.1 (2003): 195-200.
9. Rahim, Md Abdur, et al. "Face recognition using local binary patterns (LBP)." Global Journal of Computer Science and Technology (2013).
10. Kohir, Vinayadatt V., and Uday B. Desai. "Face recognition using a DCT-HMM approach." Proceedings Fourth IEEE Workshop on Applications of Computer Vision. WACV'98 (Cat. No. 98EX201). IEEE, 1998.
11. Choi, Wing-Pong, et al. "Simplified Gabor wavelets for human face recognition." Pattern Recognition 41.3 (2008): 1186-1199.

12. Ojala, Timo, Matti Pietikäinen, and David Harwood. "A comparative study of texture measures with classification based on featured distributions." Pattern recognition 29.1 (1996): 51-59.
13. Suthaharan, Shan. "Support vector machine." Machine learning models and algorithms for big data classification. Springer, Boston, MA, 2016. 207-235.
14. Georghiades, Athinodoros S., Peter N. Belhumeur, and David J. Kriegman. "From few to many: Illumination cone models for face recognition under variable lighting and pose." IEEE transactions on pattern analysis and machine intelligence 23.6 (2001): 643-660.
15. Zhou, Lijian, et al. "Face recognition based on local binary pattern and improved pairwise-constrained multiple metric learning." Multimedia Tools and Applications 79.1 (2020): 675-691.
16. Huang, Rui, Vladimir Pavlovic, and Dimitris N. Metaxas. "A hybrid face recognition method using markov random fields." Proceedings of the 17th International Conference on Pattern Recognition, 2004. ICPR 2004.. Vol. 3. IEEE, 2004.
17. Luo, Yuan, Cai-ming Wu, and Yi Zhang. "Facial expression recognition based on fusion feature of PCA and LBP with SVM." Optik-International Journal for Light and Electron Optics 124.17 (2013): 2767-2770.
18. Kiani, Kourosh, Sepideh Rezaeirad, and Razieh Rastgoo. "HMM-based face recognition using SVD and half of the face image." Modeling and Simulation in Electrical and Electronics Engineering 1.2 (2021): 45-50.

4 A Low Error Face Recognition System Based on A New Arrangement of Convolutional Neural Network and Data Augmentation

4.1. Introduction

Face recognition, without a doubt, has become one of the most popular biometric measures in the last two decades, especially in verification applications. Many reasons can be counted for this popularity, to mention a few, 1) face recognition is a non-intrusive approach, which means that the system does not require an active interaction with the subject to perform the biometric verification. 2) recent advances in technology and equipment have provided the required infrastructure for face recognition systems to be implemented fairly effortlessly in places that are needed. 3) Advances in image processing and object recognition have greatly improved the capabilities of these systems. Tasks that seemed like the stuff of science fiction a few years ago are now widely used commercially.

These enhancements can be seen in the face recognition systems as well. In most cases, the existing systems show a promising accuracy rate, acceptable speed and also adequate low complexity. However, those acceptable performances are only achievable for the under-control situation. Adding an uncontrolled parameter such as noise, illumination variation, or occlusion significantly affects the system's performance to the extent that it sometimes renders the system useless.

Many pieces of research are presented to cover these problems and improve the existing systems to be more robust against these undesirable parameters. Figure 4.1 shows the main stages of a face recognition system. Improving each of these steps enhances the system’s accuracy. [1] and [2] introduced new approaches to improve the illumination of the face images with preserving the information in the pre-processing. In feature extraction, [3], [4], and [5] proposed new methods that are faster and more accurate. Also, enhancements to the classification are provided by [6] and [7], which ultimately increases the system's accuracy.

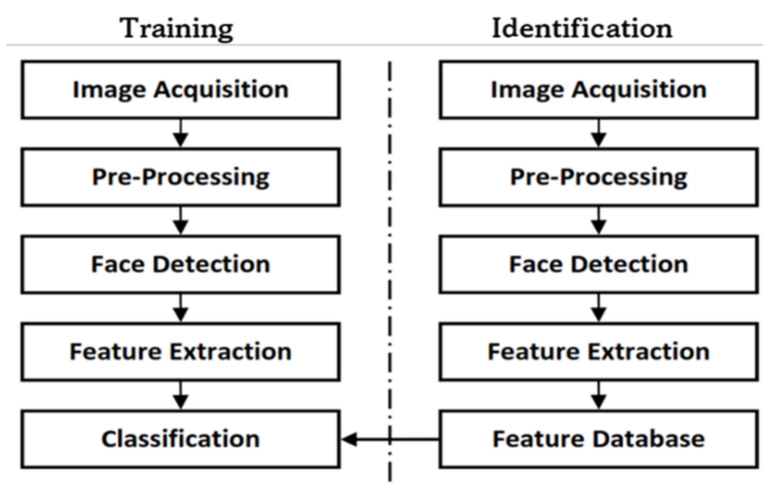

Figure 4.1 Stages of a face recognition system

In this research, however, our focus is on the feature extraction part of the face recognition system. There are different approaches to extracting the features of a face. Linear approaches map the linear features of the face image into a lower-dimensionality feature map. Examples of these methods are Principal Component Analysis (PCA) [8], Linear Discriminant Analysis (LDA) [9], and Fisher's Linear Discriminant (FLD) [10]. These methods are also called appearance-based subspace methods and focus on dimension reduction. A frequency-domain transformation, such as discrete cosine transform (DCT) [11] or Gabor wavelet [12], is used in the second category of feature extraction approaches to capture the image's distinctive features. Some of the frequency domain coefficients are used in these approaches to create picture characteristics. The high dimensionality of the spatial frequency features affects the Gabor wavelet negatively. The dimensionality reduction subspace approaches can be used to solve this. The next approach is the one that we are addressing in this paper. Deep learning is a method that has solved many restrictions, shortcomings and problems that machine learning techniques have been facing for many years. Deep learning techniques have been proven to be the most reliable and effective method in image processing, especially in face recognition applications. The methods based on the neural networks extract the deep and complex features of the face that have more unique-ness to them, hence providing more distinguishing features to be used in the process. Stacked Autoencoder [13], and Deep Belief Network (DBN) [14], [15], are examples of these types of networks. Another successful neural network that is mostly used in face recognition applications and has been confirmed to be the best in image processing applications is Convolutional Neural Network (CNN) [16]. The method is based on the convolution of the pixels of the image with a kernel to bold and extract the hid-den features of the image. The method employs other adjusting parameters and layers to reduce the dimensionality and extract the deeper features of the image. It was first introduced in [17] and was used in a handwriting recognition system. But has been implemented into countless face recognition algorithms since then. Nonetheless, CNNs have some weaknesses in some applications. For instance, when dealing with databases with a relatively low number of samples per class, systems based on CNN show low accuracy. This condition is more likely to hap-pen in real-life situations where we do not have access to many samples for each individual.

In this research, we are introducing a new arrangement to the layers of the CNN, as well as adding three normalization layers to the structure. In addition, we pro-pose data augmentation to cover the vulnerability of CNN in handling the database with a low number of samples.

The rest of this chapter is structured as follows: Section 4.2 briefly describes the theory and structure of the convolutional neural networks. In section 4.3, a concise explanation of the data augmentation is provided. The details of the proposed system are expressed in section 4.4. Section 4.5 contains the results of the experiment with the proposed method. Finally, the conclusion is presented in section 4.6.

4.2. Convolutional Neural Network

CNN is a type of feed-forward neural network that is based on the convolution of the image pixels as input and the kernel, which is an important part of the network. Because of their nature, they are extremely effective in extracting the deep and hidden features of the image, hence very practical in face recognition applications. CNNs, in general, are categorized as unsupervised methods and use features that are not always visually interpretable by human eyes. These features are weighted by the algorithm based on their importance later on in the process. The other benefit of the CNN is that the kernels and filters of these types of networks do not need to be designed or set manually, but the system, based on the training datasets and through time, learns the values related to the features for those parameters and sets them automatically. It is worth mentioning that the inspiration behind the design of the CNN and the pattern of the connection in this network is based on the connection of the brain neural connection, specifically, the visual cortex of the human brain. A typical architecture of a CNN is illustrated in figure 4.2. Convolutional neural networks consist of different types of layers. These layers that more or less can be found in almost every CNN architecture are the convolution layer, pooling layer, ReLU layer, and fully connected layer.

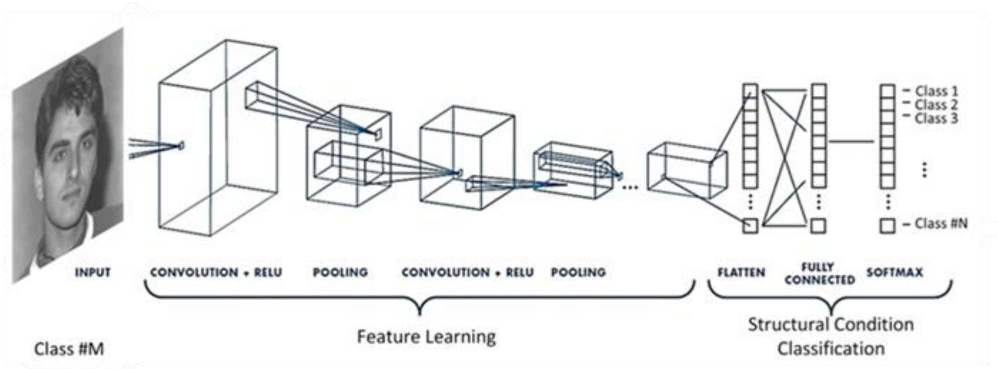

Figure 4.2 The overview of a CNN

4.2.1. Convolution Layer

The convolution operation happens in this layer. Also, kernels are part of this layer. The heaviest computation of the network occurs in the convolution layer. Before any convolution, the algorithm divides the image into equal-sized sections called image blocks. The main operation here is the convolution of the pixels of these blocks and the kernel, as mentioned earlier. Convolving those parts maintains the spatial relations of the pixels of the block. The final stage of this layer is carried by a set of learnable neurons, which is to form the feature map. Note that this map as the output of the convolution layer can be fed to the next layer as the input.

4.2.2. Pooling Layer

The purpose of this layer is to reduce the dimensionality of the data and decrease the complexity of the system. The pooling function is a non-linear operation that reduces dimensionality while preserving the necessary information. Here again, the first step is to divide the image into smaller non-overlapping blocks of pixels. In the next step, a non-linear process down samples the pixels of each block. The type of down sampling operation defines the type of pooling in the algorithm. Figure 4.3 represents two popular types of pooling operation, max pooling and average pooling. These types of layers typically can be found after convolution layers.

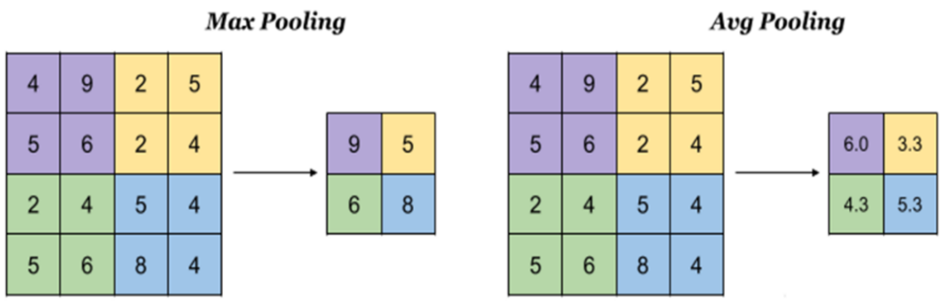

Figure 4.3 Max pooling and average pooling operations

4.2.3. ReLU Layer

Rectifying is another non-linear operation in the structure of a conventional CNN. The ReLU is a pixel-wise operation that eliminates the pixels with a negative value by substituting that pixel with a zero. The mathematical visualization of this operation is shown in equation 1, and figure 4.4 shows the function diagram.

$$f(i) = max\ (0, i) \tag{4.1}$$

where i is the value of the image's pixel.

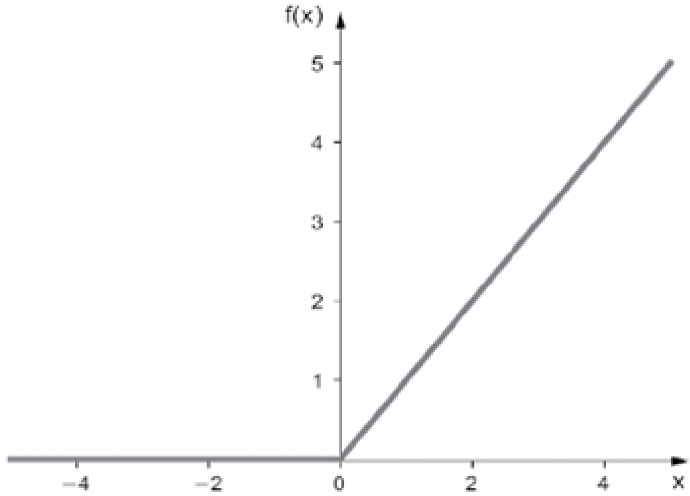

Figure 4.4 ReLU function

4.2.4. Fully Connected Layer (FCL)

FCL refers to a specific property of the network: every kernel in each layer is connected to the other kernels in other layers. The layers of the network together shape a complex relation of the image’s features. FCL allows the system to use those features and creates the feature map to be forwarded to the classifier. The last pooling layer forwards the extracted features into the classifier and performs as the FLC in the algorithm.

4.2.5. Kernels and other parameters

Another concept of the network is kernels. Kernels are the aspect of the network that extracts the features of the image. They perform this action by amplifying that specific feature of the image. The operation involves moving the kernel, which is a matrix over the image and convolves the image's relative pixels to the kernel's weights. In CNN, we do not set the values of the kernel manually, but the system determines them through the course of learning. Figure 4.5 show the way that a kernel functions. The term stride here refers to the number of pixels that each movement of the kernel covers. Zero-padding is a technique used to correct the image size when it is not dividable by the kernel size by adding zero pixels to the image.

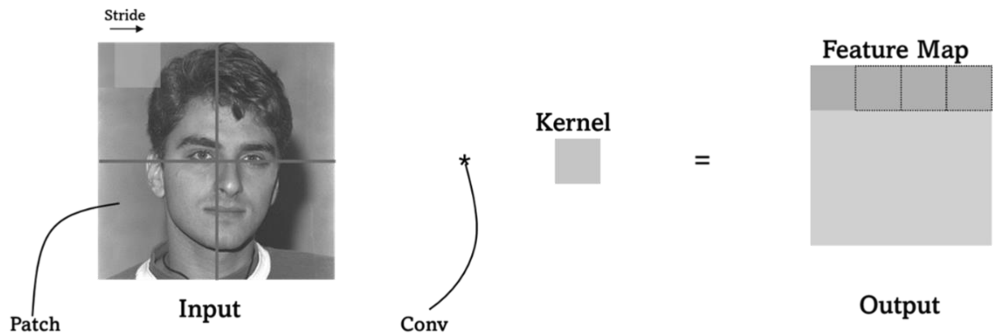

Figure 4.5 Kernel in CNN

4.3. Data Augmentation

Data augmentation is a way to create more samples from the existing samples. This method can increase the number of samples in cases where the number of samples per class is low. As mentioned before, CNN could face difficulties in those situations because of the lack of enough data for proper training. Some of the popular augmentation techniques for data augmentation are horizontal flip, shift, scaling, and rotation. These techniques, while being simple to implement, are proved to be effective. Figure 4.6 displays a visual presentation of the mentioned methods.

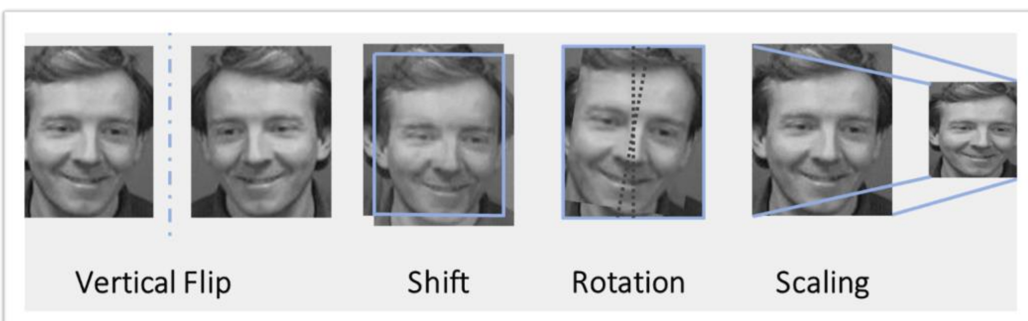

Figure 4.6 Data augmentation

4.4. Proposed Method

The convolutional neural network as the feature extractor is one of the most successful methods in face recognition applications. However, because of this network's different types of layers, various arrangements can be imagined for it; also, it is not without flaws. In this research, 1) we propose a face recognition system based on a new arrangement of the CNN layers. This new arrangement adds three separate batch normalization layers to the conventional CNN architecture. In addition to that, to improve the system's accuracy and enable the proposed algorithm to be used in real-life identification applications, which most of the time provide a low number of samples for each individual, we propose a data augmentation to the system. The data augmentation algorithm aims to increase the number of samples to be used for the CNN's training for the existing samples. Figure 4.7 illustrates the steps of the proposed face recognition system.

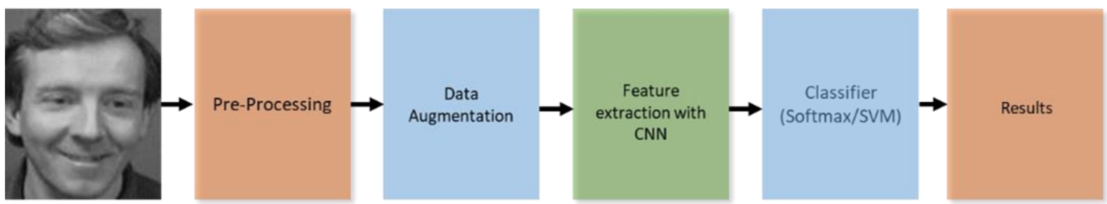

Figure 4.7 Steps of the proposed face recognition system

Pre-processing. In this step, the proposed system resizes the input images to 64 by 64.

Data Augmentation. This step involves creating new samples for each person from the existing models with the help of the introduced augmentation methods. The type of augmentation is selected randomly and normalized over all the samples.

Feature Extraction with CNN. The proposed architecture of the CNN is used in this step for the feature extraction. The architecture and the parameter setting of this CNN are demonstrated in figure 4.8 and table 4.1.

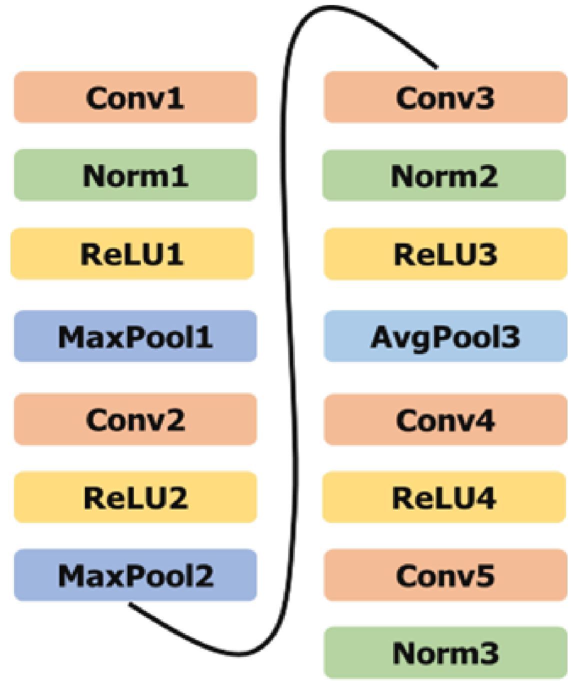

Figure 4.8 The proposed arrangement of CNN's layers

Table 4.1 Parameter of the CNN's layers

Type	Patch Size	Stride
Conv1	5X5	2
Conv2	3X3	2
Conv3	3X3	1
Conv4	3X3	2
Conv5	3X3	1
Pool1	3X3	1
Pool2	3X3	1
Pool3	8X8	-

Classifier. We consider two types of classifiers for our proposed method. Support Vector Machine (SVM) [18] and Softmax [19] are employed to classify the inputs and assign a tag to each.

Support Vector Machine. Support vector machines (SVM) are described as maximum margin classifiers. They decrease classification miscalculation as well as raise the geometric margin. An SVM generates a dividing hyperplane in the feature space, maximizing the boundary between the data. Two parallel hyperplanes are constructed to establish the margin on either side of the splitting one. These hyperplanes move toward the two data sets, letting the hyperplane with the maximum gap to the neighboring support vectors of both groups gain a proper separation. The larger the margin or space between these parallel hyperplanes, the higher the chance of the unknown samples being precisely discovered [18]. Figure 4.9 shows a standard display of the SVM.

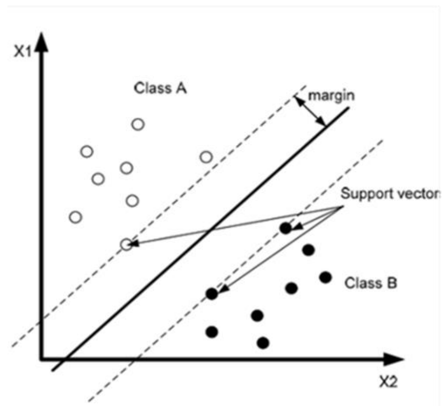

Figure 4.9 SVM and its maximum-margin hyperplane

Softmax. The Softmax function calculates the probability of a sample fitting a tag. As a classifier, it allows the algorithm to foresee the odds of the label in a multi-tag database. Figure 4.10 depicts the Softmax function and its position in a typical CNN.

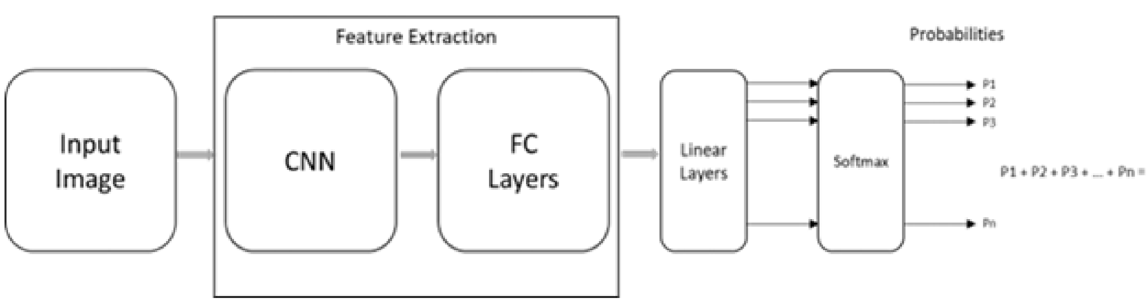

Figure 4.10 Location and relation of the Softmax to the CNN

4.4.1. ORL Dataset [20]

The ORL dataset is a collection of 400 face images gathered from 40 individuals. The faces contain various facial expressions such as smiling, open eyes, and closed eyes. Facial details such as glasses also can be found among the samples. The face images are in the size of 92 by 112. They are formatted as grayscale with 256 levels of grey for each pixel.

4.5. Experimental Setup and Results

In this section, we validate the effectiveness of our proposed system by putting it to the test. To do so, we used the ORL dataset. %50 of the samples of the dataset are used for training and the rest for testing. The images are first resized to 64 by 64 in the pre-processing stage. Then the data augmentation is used to increase the number of samples to 40000, 60000, 80000, and 120000. The augmented dataset is fed to the CNN in the next step. We use two classifiers, SVM and Softmax, and run the test for each separately. The recognition rate is calculated for each of those augmentations. Again, data augmentation is applied to the samples that we use in the test stage, and the number of samples increased to 4000, 6000, 8000, and 12000 this time. Note that the calculated accuracy rates in our experiment are gathered from averaging the results of 20 runs of each test.

4.5.1. Results

The results of the tests are presented in the following figures and charts. Figures 4.11 and 4.12 show the accuracy rates of the system for different numbers of data augmentation in both training sets and test sets and different epochs.

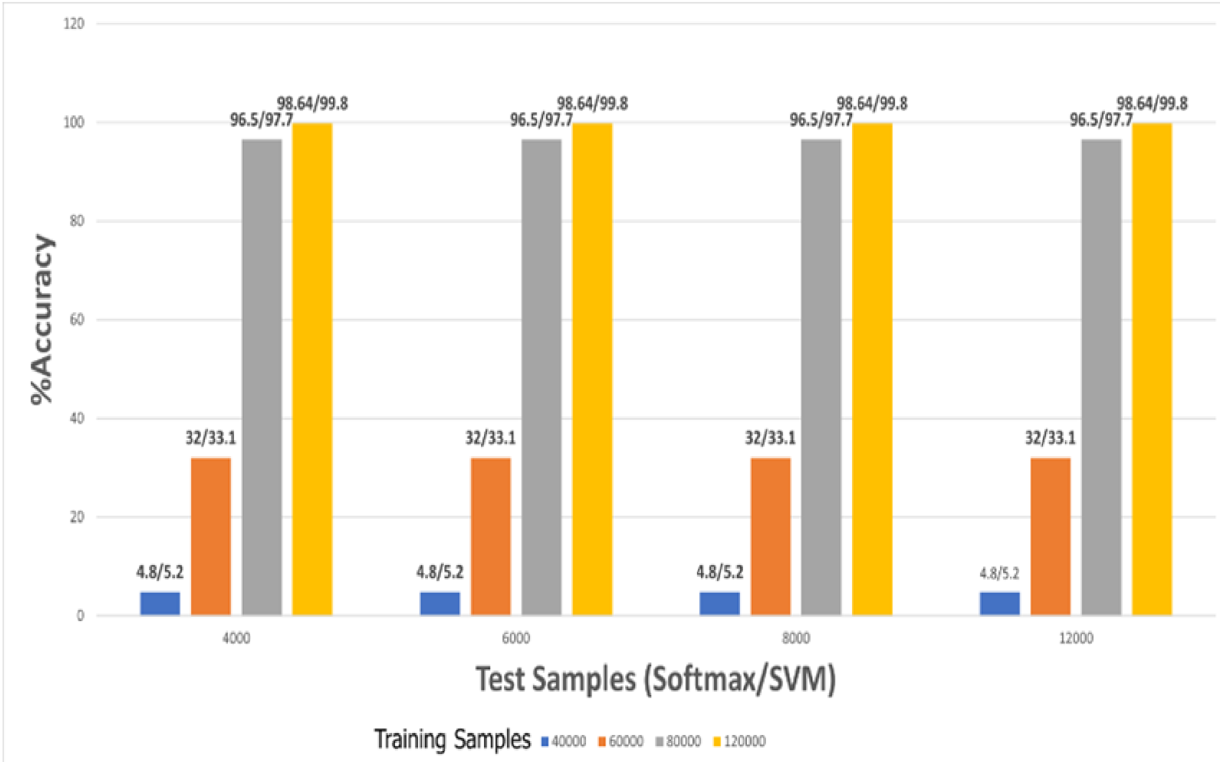

Figure 4.11 System's accuracy for different numbers of data augmentation

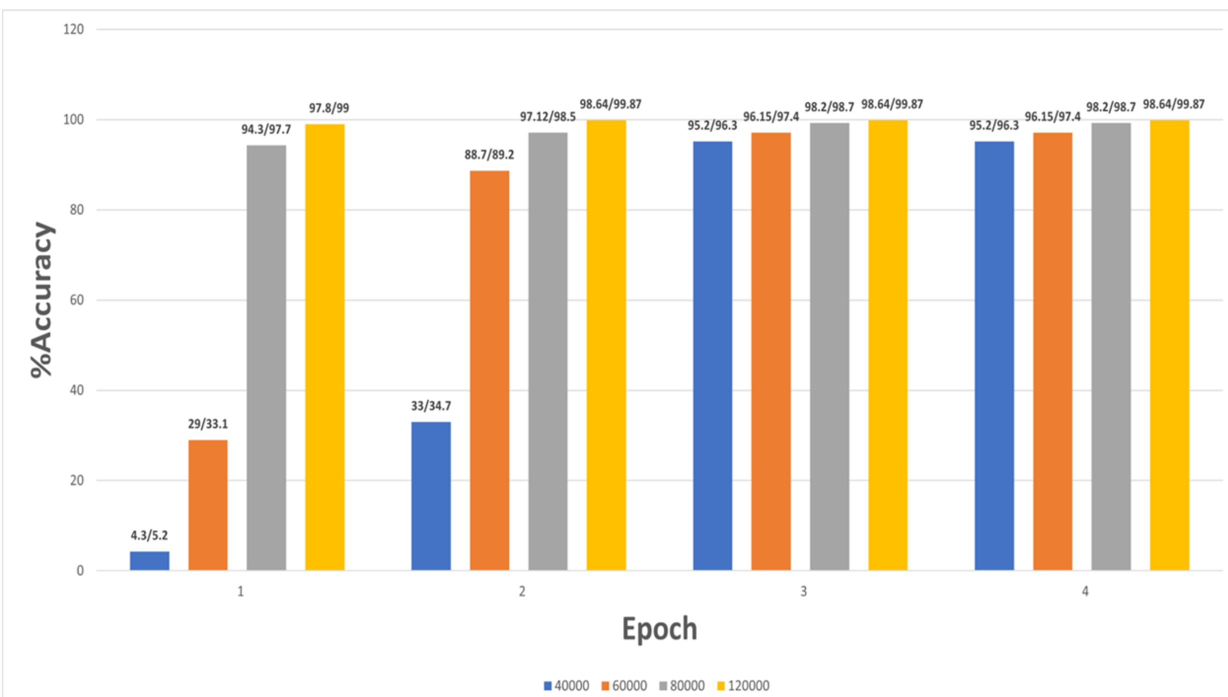

Figure 4.12 System's accuracy for different epochs

Table 4.2 contains a comparison between the recognition rate of existing approaches and the proposed system for a similar dataset.

Table 4.2 % Recognition rate of different methods

Method	Recognition Rate
ANN [21]	80.3
PCA + ANN [21]	91.0
PCA +SVM [22]	97.4
Wavelet + SVM [23]	98.1
Wavelet + PCA + SVM [24]	98
Proposed Method (Softmax)	**98.64**
Proposed Method (SVM)	**99.7**

4.6. Conclusion

This research proposed a low error face recognition system based on the proposed CNN arrangement. The proposed arrangement adds three batch normalization layers to the conventional CNN arrangement. Also, to improve the performance of the CNN for databases that have a relatively low number of samples per class, which could lead to a lack of enough samples for training, we introduced a data augmentation to increase the number of samples from the ones on hand and enhance the accuracy of the system. Moreover, in our research, the proposed algorithm is tested for two different types of classifiers. The SVM shows a higher accuracy rate in our test compared to the Softmax. The effects of various amounts of data augmentation are investigated as well. Finally, assessing the proposed technique against other existing methods shows that our system performs better than the others.

References

1. Ayyavoo, Thamizharasi, and Jayasudha John Suseela. "Illumination pre-processing method for face recognition using 2D DWT and CLAHE." Iet Biometrics 7.4 (2018): 380-390.
2. Tarawneh, Ahmad S., et al. "Automatic Gamma Correction Based on Root-Mean-Square-Error Maximization." 2020 International Conference on Computing and Information Technology (ICCIT-1441). IEEE, 2020.
3. Karanwal, Shekhar, and Manoj Diwakar. "Neighborhood and center difference-based-LBP for face recognition." Pattern Analysis and Applications 24.2 (2021): 741-761.
4. Tripathi, Rajesh Kumar, and Anand Singh Jalal. "Novel local feature extraction for age invariant face recognition." Expert Systems with Applications 175 (2021): 114786.
5. Ning, Xin, et al. "BULDP: biomimetic uncorrelated locality discriminant projection for feature extraction in face recognition." IEEE Transactions on Image Processing 27.5 (2018): 2575-2586.
6. Orrù, Giulia, Gian Luca Marcialis, and Fabio Roli. "A novel classification-selection approach for the self updating of template-based face recognition systems." Pattern Recognition 100 (2020): 107121.
7. Peng, Yali, et al. "An improvement to linear regression classification for face recognition. " International Journal of Machine Learning and Cybernetics 10.9 (2019): 2229-2243.
8. Gottumukkal, Rajkiran, and Vijayan K. Asari. "An improved face recognition technique based on modular PCA approach." Pattern Recognition Letters 25.4 (2004): 429-436.
9. Lu, Juwei, Konstantinos N. Plataniotis, and Anastasios N. Venetsanopoulos. "Face recognition using LDA-based algorithms." IEEE Transactions on Neural networks 14.1 (2003): 195-200.
10. Xiong, Huilin, M. N. S. Swamy, and M. Omair Ahmad. "Two-dimensional FLD for face recognition." Pattern Recognition 38.7 (2005): 1121-1124.

11. Sharif, Muhammad, et al. "Face recognition using edge information and DCT." Sindh University Research Journal-SURJ (Science Series) 43.2 (2015).
12. Choi, Wing-Pong, et al. "Simplified Gabor wavelets for human face recognition. " Pattern Recognition 41.3 (2008): 1186-1199.
13. Görgel, Pelin, and Ahmet Simsek. "Face recognition via deep stacked denoising sparse autoencoders (DSDSA)." Applied Mathematics and Computation 355 (2019): 325-342.
14. Benouis, Mohamed. "Face Recognition Based on Fractal Code and Deep Belief Networks." Journal of Information Technology Research (JITR) 14.4 (2021): 82-93.
15. Fatahi, Mazdak, et al. "Towards an spiking deep belief network for face recognition application." 2016 6th International conference on computer and knowledge engineering (ICCKE). IEEE, 2016.
16. Albawi, Saad, Tareq Abed Mohammed, and Saad Al-Zawi. "Understanding of a convolutional neural network." 2017 international conference on engineering and technology (ICET). Ieee, 2017.
17. LeCun, Yann, et al. "Gradient-based learning applied to document recognition." Proceedings of the IEEE 86.11 (1998): 2278-2324.
18. Pisner, Derek A., and David M. Schnyer. "Support vector machine." Machine learning. Academic Press, 2020. 101-121.
19. Qi, Xingqun, Tianhui Wang, and Jiaming Liu. "Comparison of support vector machine and softmax classifiers in computer vision." 2017 Second International Conference on Mechanical, Control and Computer Engineering (ICMCCE). IEEE, 2017.
20. https://cam-orl.co.uk/facedatabase.html
21. Nguyen, Hung K., and Xuan-Tu Tran. "An Implementation of PCA and ANN-based Face Recognition System on Coarse-grained Reconfigurable Computing Platform." VNU Journal of Science: Computer Science and Communication Engineering 36.2 (2020).
22. Lahaw, Zied Bannour, Dhekra Essaidani, and Hassene Seddik. "Robust face recognition approaches using PCA, ICA, LDA based on DWT, and SVM

algorithms." 2018 41st International Conference on Telecommunications and Signal Processing (TSP). IEEE, 2018.

23. Luo, Bing, Yun Zhang, and Yun-Hong Pan. "Face recognition based on wavelet transform and SVM." 2005 IEEE international conference on information acquisition. IEEE, 2005.

24. Gheni, Eman A., and Zahraa M. Algelal. "Human face recognition methods based on principle component analysis (PCA), wavelet and support vector machine (SVM): a comparative study." Indones. J. Electr. Eng. Comput. Sci 20.2 (2020): 991-999.

5 A Novel Hierarchical Face Recognition Method Based on the Geometrical Face Features and Convolutional Neural Network with a New Layer Arrangement

5.1. Introduction

Biometric measurement systems have become an inseparable part of our life. Recent improvements in technology, alongside the changes in the requirements of the security systems, have forced us to implement those parameters into almost every security system that we use today. Fingerprint, IRIS scan, voice recognition, and face recognition are mainly the techniques we employ. There are two main applications for these techniques: Verification and Identification. In verification, we are dealing with a system based on the information from an independent individual. Whenever a new input is put through the system, the algorithm determines if the input belongs to the group of data that exists in the system. This eventually will lead to accepting the input as an instance of the in questionindividual or rejecting it. On the other hand, identification is to set the system on multiple individuals, whom we refer to as classes, and the purpose of the system is to identify the class of each input based on the relations between the input and the information from the database of classes. The proposed method in this chapter is to be employed for identification, and the biometric measure that we consider is face recognition.

We can confidently say that face recognition is the only technique among biometric measures that we use daily, and of course, we do it naturally. This recognition is mainly done by distinguishing people based on the unique combination of their facial features, albeit unconsciously. Hence, it makes sense to base the design of an automated face recognition system on the features that can be extracted from our source. However, these features could be completely different from those we use in our daily interactions with others. For feature extraction, two different techniques exist among the current methods in use. Geometrical facial features [1]; Shape, location, and distances between the face parts such as the nose, eyes, eyebrows, and lips. This approach eliminates the effect of irrelevant information, such as background, different coloring methods, and mainly the data unrelated to face geometry. However, this method makes the system very sensitive to the uncontrolled variation of the image, such as face rotation, illumination, and sometimes facial expressions. The other approach, on the other hand, uses the global elements of the whole image, regardless of their relevance to the face. It is needless to say that the downside of this holistic method is its sensitivity to the unrelated information of the image. Active Appearance graph Models (AAM) [2] and Convolutional Neural Networks (CNN) [3] are examples of the first and second methods, respectively. Also, a combination of some methods exists as the hierarchical technique.

CNN has been proven to be one of the most efficient techniques in image processing [4]. Because of the nature of the convolution operation and the existence of different types of filters, CNNs give desirable results in most feature extraction and classification, especially in image processing [5]. However, they are not free of shortcomings. One of the obstacles in employing CNNs for face recognition is that these networks need a relatively large number of samples per person to show acceptable performance, and evidently, the high number of samples is not the case in most artificial and real-life situations. One of the ways to overcome this problem is to use data augmentation. The main goal of data augmentation is to create enough samples from existing data to satisfy the system's prerequisites.

The other disadvantage of these systems is their computational complexity [6]. This issue is more critical in the instance with a relatively high number of individuals in the database, which is more likely to happen in real-life situations. The introduction of the GPU to the process and utilizing the computational power of new Graphical Processing Units helped to some extent [7].

The hierarchical approaches try to utilize various techniques simultaneously, to address multiple obstacles and deficiencies of the system [8]. In these techniques, usually, several methods are employed for different stages of a face recognition system in a hierarchical arrangement [9], [10]. This approach helps to improve the accuracy by adding multiple layers to the feature extraction. However, in most cases, because the hierarchy is implemented in the structure of the feature extractor rather than the structure of the entire system, the impact is only on the system's accuracy.

The rest of the chapter is organized as follows: Section 5.2 briefly explains the proposed method. The structure of the proposed algorithm is discussed in detail in section 5.3. Section 5.4 presents the experimental setup, and section 5.5 is dedicated to the results of the experiments. Finally, in section 5.6, the conclusion is represented.

5.2. The proposed hierarchical face recognition method

Dealing with large databases with a large number of individuals is one of the obstacles of face recognition systems. This issue is more noticeable when a new face is needed to be added to the database. The other weakness of conventional face recognition systems is that the system's accuracy is affected by the shortcoming of the single extractor and classifier. On the other hand, the proposed method uses a simple and fast feature extractor based on the geometrical features of the face in the first stage to create new classes of individuals, however, with more members from the initial individual classes in each new class. In the process of recognition, the first stage short-lists the probable classes that the input image might belong to.

Meanwhile, the CNN in the second stage is employed to search through the short-listed classes determined by the first stage to precisely establish the class of the input image. The advantage of this method is that the CNN, which is the most resource-demanding part of the system, works on a comparatively smaller database. Also, reducing the number of individuals decreases the number of data augmentations that are required to enhance the individuals to samples ratio.

5.2.1. Geometrical Feature Extraction [11]

Face landmarking, or the localizing of the face parts is the main task in geometrical feature extraction. Figure 5.1 shows some of the landmark features of a human face.

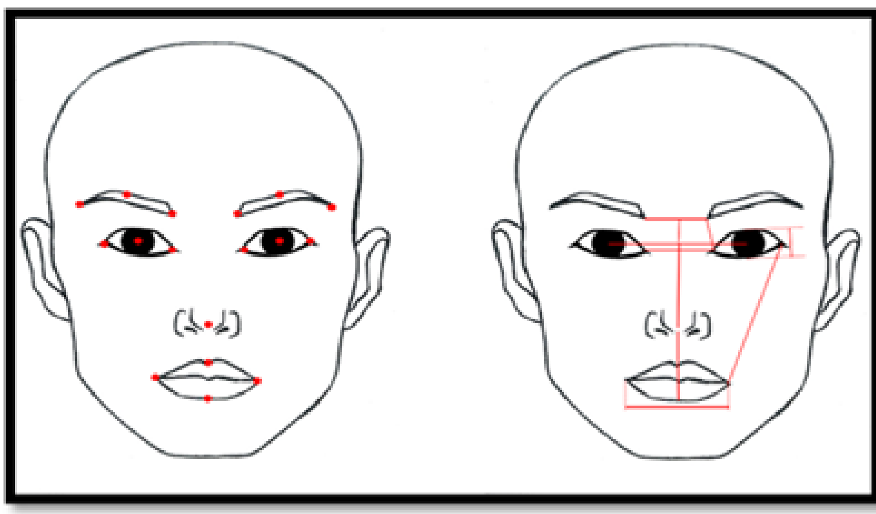

Figure 5.1 Left: Face Landmarks, Right: Feature Axes

AMM is one of the most effective techniques for extracting geometrical features [12]. Points and locations extracted by the algorithm can be used to form the feature axes, which are the distances between certain locations of the face image.

5.2.2. Convolutional Neural Network

CNNs are feed-forward types of neural networks that have been used especially in image processing and face recognition. They are made up of different layers. The most important part of these networks is the kernel or filter.

The dot product of the convolution is between the pixels of the input image and the kernel. The layer that contains the convolution operation is called the convolution layer. Other layers, such as the Pooling layer, ReLU layer, and fully connected layer, can be found in the structure of a CNN. Figure 5.2 shows a general overview of a CNN.

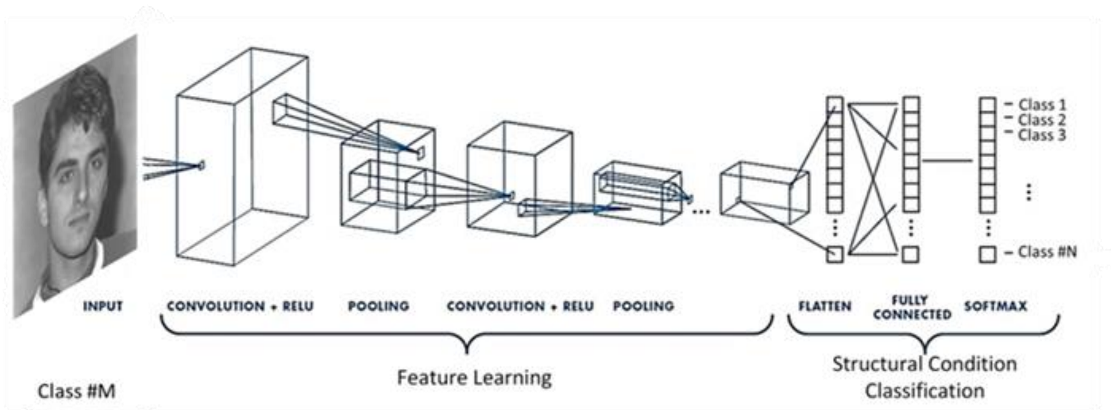

Figure 5.2 Overview of the Convolutional Neural Network

Convolution Layer. The main process of convolution happens in this layer. Because of that, we can consider this layer to be the carrier of the main computational load of the neural network. The main idea here is to extract hidden features of the face image. This idea is achieved by dividing the input image into smaller sections and convolving those parts with the kernel. The convolution operation preserves the spatial relations of the pixels in that small section. A set of learnable neurons are responsible for creating the convolution and constructing the activation or feature map. Note that the output of each convolution layer could be the input of another convolution or another layer.

Pooling Layer. A dimensionality reduction step with the nature of pooling is applied to the input to reduce the system's complexity. In the pooling layer, the target is to reduce the dimension of the data without losing any important parts. The procedure starts with dividing the input into smaller non-overlapping blocks. Then a process of a non-linear down-sampling is applied to each region. The most famous non-linear operations in the pooling layer are max pooling and average pooling. This layer specifically helps with faster convergence, more suitable generalization, and robustness against distortion, and this layer's usual placement is between convolution layers.

ReLU Layer. Rectifying happens in the non-linear ReLU layer. The layer operates on each pixel and replaces all the negative values of the feature map with zero. To better understand the operation of the ReLU function, we can look at the mathematical expression of the layer as $f(i) = \max(0, i)$, in which i is the input value representing a pixel of the input image.

Fully Connected Layer. The term FCL is basically a property of the network rather than being a separate layer. It refers to the fact that every kernel in the layer is connected to the kernel of the next layer. The layers that we mentioned before are the interpretation of the complex features of the input. FCL enables the system to put those complex features into use. We consider the FCL to be the last pooling layer that forwards the extracted features into our classifier.

The main steps of the CNN as a face recognition system are shown in the block diagram in figure 5.3.

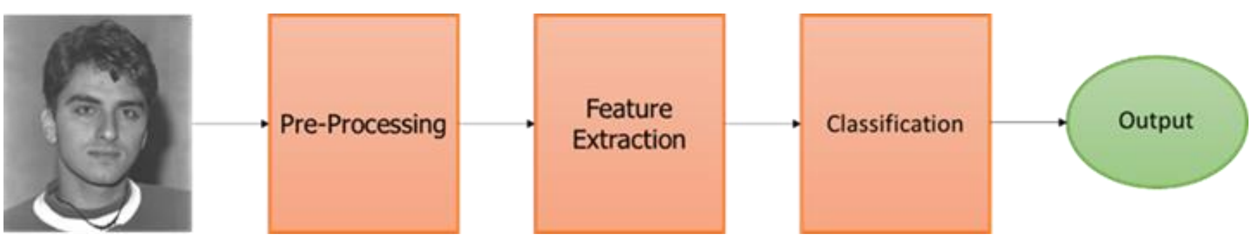

Figure 5.3 Block diagram of the CNN

5.3. The proposed algorithm

The block diagram of the proposed method is illustrated in figure 5.4. The technique mainly consists of two major sections; Geometrical and CNN feature extraction and classification, and a few other subsections, as shown in figure 5.4.

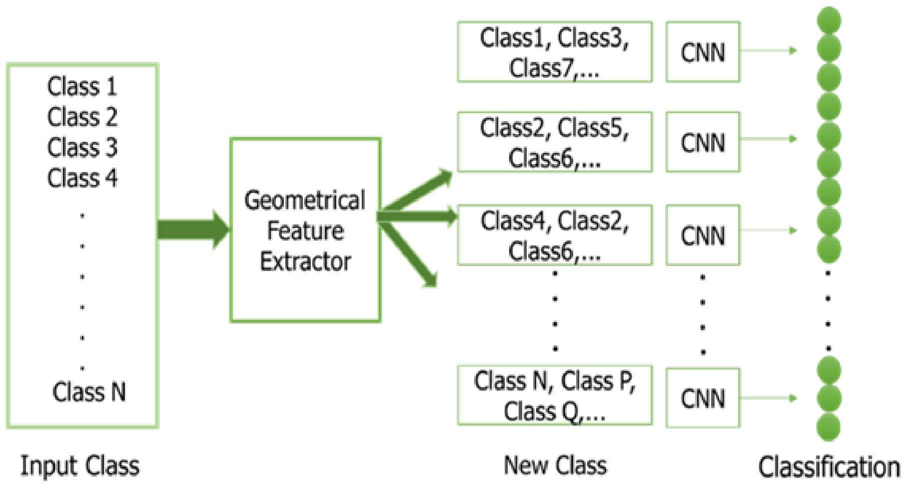

Figure 5.4 The block diagram of the proposed algorithm

5.3.1. AMM

In the geometrical feature extractor, the algorithm employs AMM to extract the geometrical features of the face based on the location of some of the face parts. Six main points and three axillary points are considered to calculate five axes or distances as a feature map for each input. The tags of those five distances are WA, EA, WEA, LA, and ELA, as shown in figure 5.5. Each distance creates two axes, the left and the right side of the distance. This is mainly done to prevent the effect of variation in face orientation.

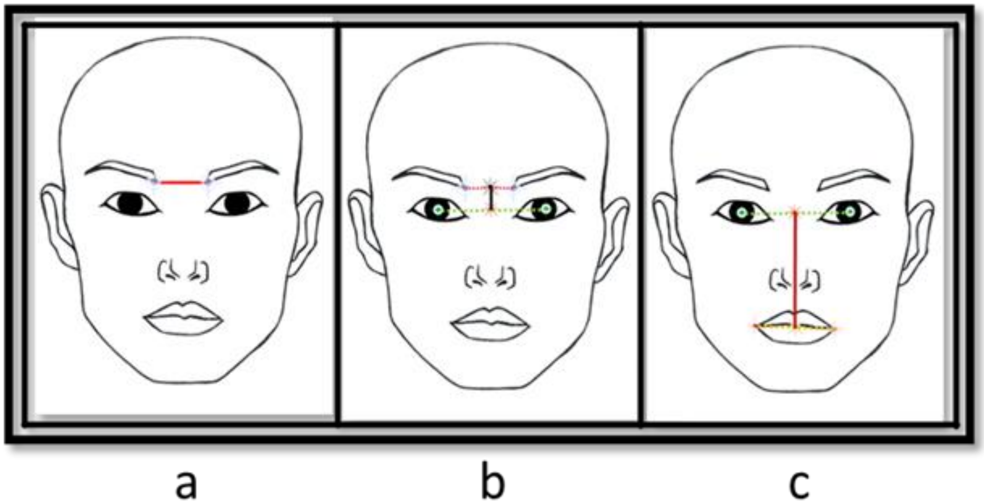

Figure 5.5 a) WA, b) EA and WEA, c) LA and ELA

In the next stage, the algorithm feeds the feature map into an SVM classifier. The SVM ranks the dependency of the input to the classes. Specifically, the first ten most probable classes are assigned to each input. This process leads to the creation of new classes containing multiple members of the original input classes.

5.3.2. Data Augmentation

Before putting the data into the CNN for training, a process of augmentation is done on the classes of the database. The augmentation step helps to overcome the negative effect of the low number of input samples. Some of the techniques we used for data augmentation are horizontal flip, shift, scaling, and rotation. These methods are shown in figure 5.6.

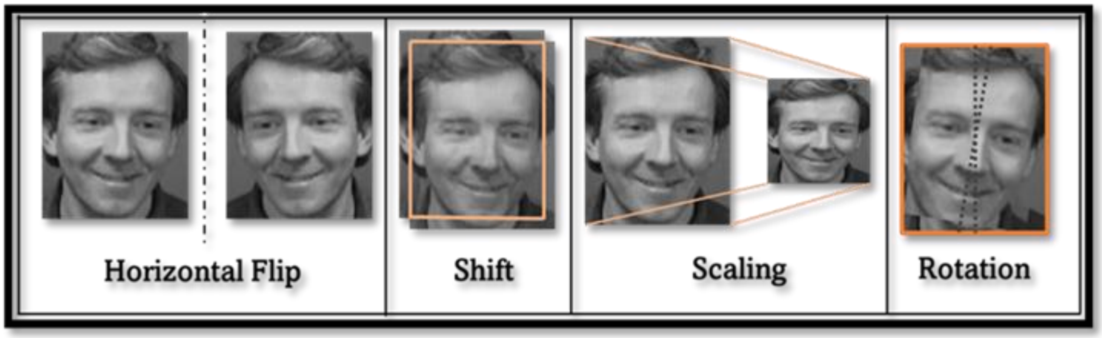

Figure 5.6 Data Augmentation

5.3.3. CNN

The convolutional neural network stage employs a new proposed arrangement of layers. Multiple layers of convolution, ReLU, and pooling exist in the arrangement. Other than those layers, we proposed the addition of normalization layers to the arrangement. Figure 5.7 shows the proposed arrangement of layers of the feature extraction section of the CNN, and table 5.1 displays the parameters of those layers.

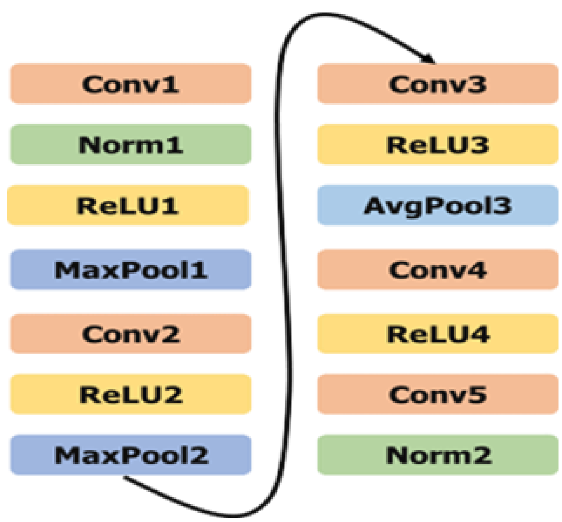

Figure 5.7 The proposed arrangement of layers

Table 5.1 Parameter of the CNN's layers

Type	Patch Size	Stride
Conv1	5X5	2
Conv2	3X3	2
Conv3	3X3	1
Conv4	3X3	2
Conv5	3X3	1
Pool1	3X3	1
Pool2	3X3	1
Pool3	8X8	-

SVM. Support vector machines (SVM) are characterized as maximum margin classifiers since they reduce classification error while also increasing geometric margin [13]. An SVM creates a separating hyperplane in the feature space that maximizes the margin between the data sets. To determine the margin, two parallel hyperplanes are built, one on either side of the separating one. These hyperplanes are pushed up against the two data sets, allowing the hyperplane with the farthest distance to the adjoining support vectors of both classes to achieve a decent separation, as shown in figure 8. The greater the margin or distance between these parallel hyperplanes, the more likely unknown samples will be accurately identified [14].

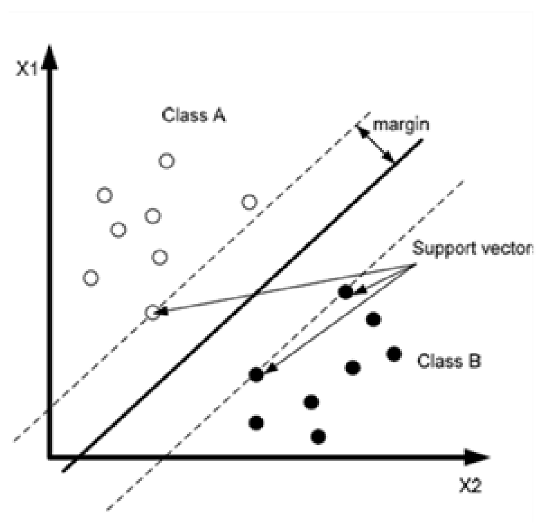

Figure 5.8 Maximum-margin hyperplane for an SVM classifier. Samples on the margin are called the support vectors.

Softmax. Softmax measures the probability of input belonging to a class. As a classifier, it enables the algorithm to predict the probability of the label in a multi-label database. Figure 5.9 illustrates the Softmax operation and its location in a conventional CNN.

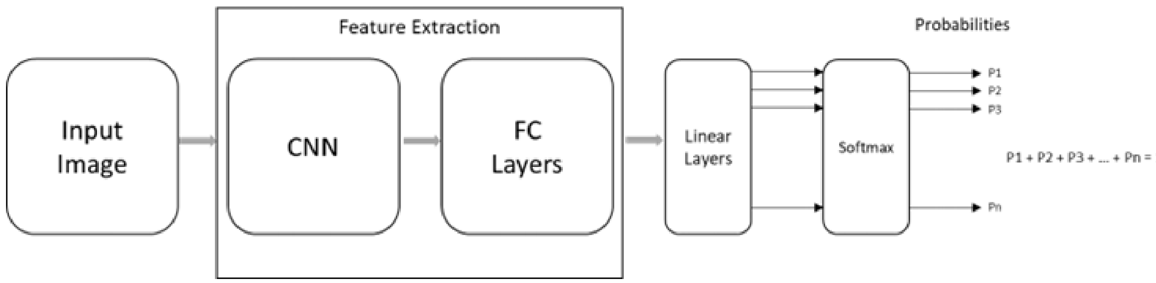

Figure 5.9 Location and relation of the Softmax to the CNN

5.3.4. Database

The color FERET [15][16] database is a face recognition dataset. It comprises 11,338 color photos at a resolution of 512x768 pixels taken in a semi-controlled setting with 1208 individuals in multiple distinct stances. The database contains different poses for each individual ranging from frontal to half-face images. Also, different facial expressions and lighting conditions exist for each individual. In our experiment, we only used frontal images of each person, and the images were resized to 64x64 whenever it was fed to CNN. A few samples of the dataset are shown in figure 5.10.

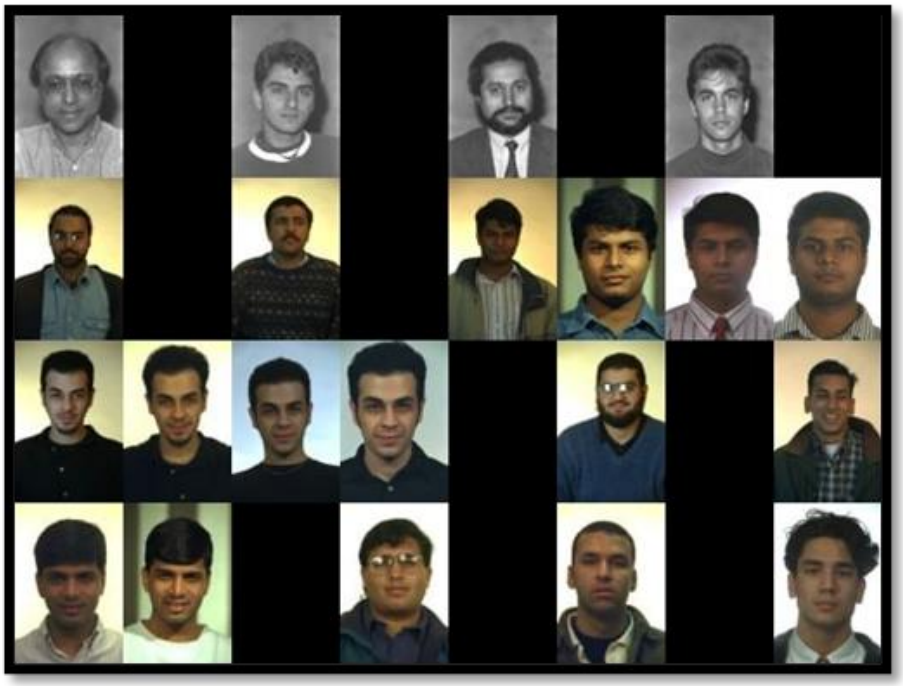

Figure 5.10 Samples of the color FERET database

5.4. Experimental Setup

In the experiment, all the classes of the color FERET database are used. AMM is the algorithm used to extract the geometrical features, and the classifier in this stage is SVM. The combination of AMM and SVM creates 64 new classes out of 1208 original classes in the database. The average number of individuals in new classes is around 21 persons. The total number of individuals the algorithm deals with after AAM is 1384 groups. The biggest new class in terms of the number of individuals has 42 members, and the smallest has 5.

Before putting the data into the CNN, the samples are resized to 64x64. The data augmentation is carried out in the new classes to increase the number of samples to 10000 in each new class. The recognition results are gathered for both SVM and Softmax as the classifiers of the convolutional neural network.

Half of the data are used for training, and the remaining half for the test. If CNN does not provide an output, i.e., does not merge, a return to the previous stage will happen, and the algorithm will consider the next tag in rank as the new class for AMM output.

5.5. Experimental Results

The results of running the algorithm on all 1208 classes of the color FERET database (%50 training %50 test) and a comparison between the proposed method and other existing techniques are shown in table 5.2. Table 5.3 illustrates some of the parameters of the test and results for different geometrical feature extraction methods, and table 5.4 shows the results of different parameters such as input size and results for CNN. As we can see, the highest recognition rate of around %97 happens when the size of the input image is 64x64 and also, using the SVM classifier increases the recognition rate compared to the Softmax. (%97.68 over %96.35).

Table 5.2 The recognition rate on the color FERET database and comparison of the results

Method	%Accuracy
SVM-Linear [17]	81
Eigenfaces and DoG filter [18]	84
SVM-Polynomial [17]	87
SVM-RBF [17]	91
HMM [17]	90
AMM + CNN + Softmax (Proposed)	**96.35**
AMM + CNN + SVM (Proposed)	**97.68**

Table 5.3 The parameters and outputs of the different geometrical feature extractors

Method	Image Size	Training Time per Sample	Classes (Highest)	Number of Returns	Number of Sub-groups
Tzimiro [19]	Original	1.012	1208	89	64
Cootes [20]	Original	0.88	1208	145	115
FACS [21]	Original	0.98	1208	107	71

Table 5.4 The parameters and output results of different sizes of input and classifiers in CNN

Method	Image Size	Time per Epoch Sec (Avg)	#Classes (Highest)	%Accuracy	Incorrect Prediction
Proposed + SVM	16X16	14	42	71.85	340
Proposed + SVM	32X32	19	42	91.14	107
Proposed + SVM	64X64	33	42	97.68	28
Proposed + Softmax	16X16	18	42	69.86	364
Proposed + Softmax	32X32	24	42	90.39	116
Proposed + Softmax	64X64	41	42	96.35	44

5.6. Conclusion

This research proposes a new system for face recognition applications. The new method has a hierarchical structure and consists of two stages of feature extraction and classification. The first stage extracts the geometrical features of the face image by employing the AMM technique and uses SVM as the classifier. The output of the first classifier forms the new classes of the data from the original input classes. We propose a new layer arrangement for the CNN and add two normalization layers to the structure. The data augmentation is also employed to compensate for the low number of input samples. The second stage utilizes the proposed CNN to extract the global features of the samples. We compared the recognition rate for two different classifiers, Softmax and SVM, as the second classifier. This paper uses the color FERET database for evaluation purposes. The results show a better performance of the proposed method by providing a higher recognition rate.

References

1. Marcolin, Federica, and Enrico Vezzetti. "Novel descriptors for geometrical 3D face analysis." Multimedia Tools and Applications 76, no. 12 (2017): 13805-13834.

2. Tzimiropoulos, Georgios, and Maja Pantic. "Fast algorithms for fitting active appearance models to unconstrained images." International journal of computer vision 122.1 (2017): 17-33.

3. Coşkun, Musab, et al. "Face recognition based on convolutional neural network." 2017 International Conference on Modern Electrical and Energy Systems (MEES). IEEE, 2017.

4. Yan, Kewen, et al. "Face recognition based on convolution neural network." 2017 36th Chinese Control Conference (CCC). IEEE, 2017.

5. Albawi, Saad, Tareq Abed Mohammed, and Saad Al-Zawi. "Understanding of a convolutional neural network." 2017 international conference on engineering and technology (ICET). Ieee, 2017.

6. Alzubaidi, Laith, et al. "Review of deep learning: Concepts, CNN architectures, challenges, applications, future directions." Journal of big Data 8.1 (2021): 1-74.

7. Sati, Vishwani, et al. "Face Detection and Recognition, Face Emotion Recognition Through NVIDIA Jetson Nano." International Symposium on Ambient Intelligence. Springer, Cham, 2020.

8. Wang, Mei, and Weihong Deng. "Deep face recognition: A survey." Neurocomputing 429 (2021): 215-244.

9. Li, Haoxiang, and Gang Hua. "Hierarchical-pep model for real-world face recognition." Proceedings of the IEEE conference on computer vision and pattern recognition. 2015.

10. Gao, Guangwei, et al. "Hierarchical deep cnn feature set-based representation learning for robust cross-resolution face recognition." IEEE Transactions on Circuits and Systems for Video Technology (2020).

11. Amarapur, Basavaraj, and Nagaraj Patil. "The facial features extraction for face recognition based on geometrical approach." 2006 Canadian Conference on Electrical and Computer Engineering. IEEE, 2006.

12. Tzimiropoulos, Georgios, et al. "Generic active appearance models revisited." Asian Conference on Computer Vision. Springer, Berlin, Heidelberg, 2012.

13. Suthaharan, Shan. "Support vector machine." Machine learning models and algorithms for big data classification. Springer, Boston, MA, 2016. 207-235.

14. Kecman, Vojislav. Learning and soft computing: support vector machines, neural networks, and fuzzy logic models. MIT press, 2001.

15. Phillips, P. Jonathon, et al. "The FERET database and evaluation procedure for face-recognition algorithms." Image and vision computing 16.5 (1998): 295-306.

16. Phillips, P. Jonathon, et al. "The FERET evaluation methodology for face-recognition algorithms." IEEE Transactions on pattern analysis and machine intelligence 22.10 (2000): 1090-1104.

17. Nabatchian, Amirhosein. "Human face recognition." (2011).

18. Kortli, Yassin, et al. "Face recognition systems: A survey." Sensors 20.2 (2020): 342.

19. Tzimiropoulos, Georgios, and Maja Pantic. "Fast algorithms for fitting active appearance models to unconstrained images." International journal of computer vision 122.1 (2017): 1733.

20. Cootes, Timothy F., Gareth J. Edwards, and Christopher J. Taylor. "Active appearance models." European conference on computer vision. Springer, Berlin, Heidelberg, 1998.

21. Lien, James J., et al. "Automated facial expression recognition based on FACS action units." Proceedings Third IEEE International Conference on Automatic Face and Gesture Recognition. IEEE, 1998.

6 Summary, Conclusion and Future Works

6.1. Conducted Research

The research in this dissertation targets the enhancement of the accuracy of the face recognition systems. Face recognition is dealt with as a four-step process in this work, and the improvements are gained through improving each of those four steps. From this research point of view, the three main steps of a face recognition system are 1) pre-processing, 2) feature extraction, and 3) classification.

In this work, a pre-processing method is developed to overcome the challenge of poor illumination in face images which can affect the system's accuracy by spoiling the crucial parts of the image's texture information. The introduced method improves the illumination and contrast of the image while maintaining the texture information and avoiding over and under enhancements in dark and bright regions. The technique provides that by applying global and local enhancements and also by considering the lighting situation of the image's blocks. The proposed method uses a specific type of histogram equalization for contrast correction and employs two levels of gamma correction to enhance the brightness by averaging gamma levels when necessary. As a result, the proposed system shows a considerable improvement to various face recognition systems and also when compared to other techniques in increasing the accuracy of a particular face recognition system.

The other contribution of this research is a feature extraction technique based on the texture information of the image. The proposed method utilizes the LBP feature extractor approach to create feature maps from an input image. An adaptive threshold function based on the Gaussian distribution function is introduced to help maintain the amplitude information and also improve the system’s resistance to noise. The method employs an error function in the Gaussian distribution and replaces this new function with the fixed zero thresholds in the conventional LBP. Assessing the new technique demonstrates a higher recognition rate compared to conventional LBL and other existing methods. Further testing against two types of image noise confirms the improved performance of the proposed approach in handling the noise.

The next achievement is a low error face recognition system. This system utilizes a convolutional neural network to extract the hidden features of the face images. This research proposes a new arrangement for CNN by introducing three batch normalization layers to the architecture. These layers particularly decrease the possible internal covariant shift. The suggested system applies data augmentation to cover the low number of samples per class, which can drastically impact CNN's effectiveness. In this part of the research, the suitability of two different classifiers, namely SVM and Softmax, is investigated. The results show that the proposed system with SVM as the classifier outperforms other existing methods in recognition rate and demonstrates a lower error.

The final part of this research is dedicated to introducing a new face recognition system. This new system has a hierarchical structure, containing two levels of feature extraction and classification. The system's architecture is developed to first rearrange the members of the database into bigger groups based on their sharing features. This helps the algorithm shortlist the possible tag or group of the input in the recognition phase faster with lower resources and computational complexity and then use a more precise and complex recognition method to further narrow down and assign the final tag to the input image. Since the second recognition level works with fewer classes, the system can perform more accurately with the same amount of resources. In the first recognition phase, the system uses the geometrical facial features to categorize the classes of the individuals from the database into new groups. The AAM technique is employed to perform this task, and SVM acts as the classifier for this level. Six main and three auxiliary locations are considered the face features and are utilized to form the feature axes that create the feature vector. In the second stage, the input image goes through a CNN, which is trained on the determined groups of individuals. The performance of the system is measured for both SVM and the Softmax classifiers in this stage. To address the issue of the low samples per class for the CNN, a data augmentation step is provided for the system before the second stage recognition. The proposed system demonstrates better performance regarding the recognition rate compared to other existing methods. Also, since the first recognition level reduces the number of classes for which the CNN is required to be trained on, the amount of required data augmentation is considerably decreased.

6.2. Future Works

Although the existing methods and techniques provide acceptable accuracy for most applications, there are still situations and usages existing that require lower errors or different approaches to deal with unpredicted challenges. One potential future topic of research can be developing face recognition systems and methods targeting those specific obstacles. Barriers like the low quality of the database samples. The bottleneck of the image quality can be investigated and improved as an interesting topic.

Furthermore, other types of neural networks, such as correlation neural networks, combined with other methods can be used to develop a new face recognition system with even a higher recognition rate.

Also, improving the other parameters of a face recognition system, such as speed and computational complexity, can be an attractive research topic. This especially helps to improve the systems that are using the videos as the source of input. Improving the speed of a face recognition system allows for using higher frame rate videos, hence giving the system more data to work with.

VITA AUCTORIS

NAME:	Soroosh Parsai
PLACE OF BIRTH:	Shiraz, Iran
YEAR OF BIRTH:	1984
EDUCATION:	NODET, Abadan, KHZ, IR, 2002, High School.
	Sadjad University, Mashhad, KHR, IR, 2009, B.Sc.
	IAU of Kermanshah, Kermanshah, KRS, IR, 2014, M.Sc.